Aquaculture International (2026) 34:10
https://doi.org/10.1007/s10499-025-02401-1

RESEARCH

Uncovering patterns in skin and gut microbiota of rainbow trout (*Oncorhynchus mykiss*): insights from machine learning and feeding trials with sustainable aquafeeds based on yellow mealworm (*Tenebrio molitor*)

Silvio Rizzi[1] · Giulio Saroglia[2] · Violeta Kalemi[1] · Simona Rimoldi[1] · Genciana Terova[1]

Received: 22 October 2025 / Accepted: 5 December 2025

Abstract
The aquaculture sector has been progressively transitioning toward environmentally sustainable feed production, with insect meals emerging as viable alternatives to fish meal (FM); however, their effects on the microbiota of fish still remain insufficiently characterized. This study examined gut and skin microbiota in rainbow trout (*Oncorhynchus mykiss*) following complete FM substitution with yellow mealworm (*Tenebrio molitor*, TM), utilizing machine learning (ML) to investigate diet-microbiota relationships. To this end, microbial abundance data from a prior in vivo trial were analyzed by means of a structured ML pipeline. On the one hand, classification models assessed the association between microbial profiles and dietary regimens, while, on the other hand, regression models evaluated the predictive capacity of feed ingredient variations on microbial abundance shifts. Within this processing framework, feature selection identified informative taxa across taxonomic levels, enhancing model generalizability and reducing overfitting. Several classification algorithms attained optimal accuracy, whereas regression models showed moderate performance, with error values decreasing from higher to lower taxonomic ranks. In particular, feature selection and explainability analyses identified both diet- and tissue-associated indicators: *Cutibacterium*, *Enhydrobacter*, and *Lactobacillus* in the gut; *Chryseobacterium*, *Flectobacillus*, and *Sphingopyxis* in the skin. The occurrence of *Deefgea* in both tissue types suggested potential water-fish microbial exchange. In conclusion, despite conventional analyses showing only limited dietary modulation, ML models effectively detected diet- and tissue-specific indicators in rainbow trout following FM substitution with TM, ultimately underscoring the potential of integrating AI-driven techniques with next-generation sequencing to uncover ecological patterns across fish tissue types and taxonomic levels.

Keywords Aquaculture · Machine learning · Fish microbiota · *Tenebrio molitor* · *Clostridium* · *Deefgea*

Handling Editor: Ercument Genc

Extended author information available on the last page of the article

Introduction

The longstanding reliance of aquafeed production on marine-derived resources, particularly fish meal (FM) and fish oil (FO), has contributed to the depletion of oceanic resources, resulting in increasing disruption of food webs and heightened pressure on marine stocks, with 64.5% being fished within biologically sustainable levels and 35.5% being classified as severely overfished (Sharma et al. 2025). Such resource degradation has imposed considerable pressures on other intensive production systems, particularly livestock sectors, aggravating existing sustainability issues and intensifying competition with human food supply chains (Hua et al. 2019; Olsen and Hasan 2012). As one of the most rapidly expanding sectors within global food production, aquaculture must confront both environmental (e.g., climate change) and anthropogenic (e.g., population growth, ethical production standards, responsible consumption) challenges to secure long-term sustainable production (D'Abramo 2021). To this end, ongoing research initiatives are investigating viable alternative ingredients to reduce feed reliance on FM and FO while maintaining optimal growth performance and health outcomes in cultured species. Indeed, sustainable feed formulation is gradually redefining the notion of essential ingredient in terms of complementarity rather than substitution to meet nutritional requirements, guarantee food security, promote economic growth, and safeguard natural ecosystems (Turchini et al. 2019).

When designing alternative aquafeeds, manufacturers should adhere to holistic principles prioritizing ingredient sustainability. Indeed, according to aquafeed stage definitions proposed by Eroldoğan et al. (2023), the transition from Aquafeed 1.0 (marine-based) to Aquafeed 2.0 (animal- and plant-based) has introduced several challenges related to feed, including contamination risks of animal-derived meals as well as the environmental footprint of crop-based ingredients. To address these issues and further reduce reliance on fishmeal (FM) and fish oil (FO), researchers have increasingly evaluated microorganisms and macroorganisms as promising alternative resources (Glencross et al. 2020), even though formulating effective feeds remains more difficult for carnivorous species (Turchini et al. 2019). In the case of macroorganisms, in particular, insects have exhibited significant potential due to advantageous traits, including minimal environmental footprint, favorable nutritional composition, elevated feed conversion efficiency, commercial-level production scalability, and non-competition with human food production (Colombo and Turchini 2021). In an effort to complete the transition to Aquafeed 3.0 (biocircular-based), manufacturers have thus progressively shifted toward alternative resources, enhancing production efficiency and economic resilience while simultaneously minimizing environmental impact and progressing sustainability objectives (Colombo and Turchini 2021). Studies have demonstrated the sustainable and regenerative capacity of biocircular ingredients, encompassing the use of more environmentally sustainable organisms, the valorization of cross-sector by-products, and the remediation of nutrient discharges (Eroldoğan et al. 2023). Nevertheless, future feeds are expected to consist of multi-ingredient formulations, with biocircular ingredients as the core and other sustainable sources as blend to achieve nutritional completeness and economic viability.

Alongside the nutritional and environmental advancements characterizing the transition to Aquafeed 3.0, the aquaculture production system is undergoing a profound technological revolution, as marked by the paradigm shift from Aquaculture 3.0 to Aquaculture 4.0. This transformation is hallmarked by the seamless integration of smart technologies and cyber-physical systems, although ensuring system reliability remains a critical issue (Biazi and Marques 2023). Therefore, these interconnected network architectures are expected to

facilitate the implementation of intelligent automated systems capable of real-time monitoring and evidence-based decision-making across the entire production chain, from water quality control and feeding frequency management to microbiota analysis, thereby minimizing external biases linked to human intervention. Within this emerging technological landscape, precision fish farming offers a robust framework to implement predictive models through artificial intelligence (AI) for the investigation of microbiota dynamics.

Currently recognized as an emerging protein source for aquafeed formulation, the yellow mealworm (*Tenebrio molitor*, TM) represents a promising candidate for FM replacement thanks to nutritionally advantageous properties marking both developmental stages (larva, pupa, and adult) and by-products (exuviae and excreta). Notably, the larval form exhibits a composition especially well suited for feed formulation (Noyens et al. 2024; Gkinali et al. 2022). As demonstrated by rearing trials conducted under controlled laboratory conditions (Ravzanaadii et al. 2012), TM larvae possess both high protein content (46%), with a balanced amino acid profile, and high lipid content (33%), with a predominance of oleic acid, linoleic acid, and palmitic acid. Moreover, as documented by Jankauskienė et al. (2024), TM larvae showed comparatively higher energy values when fed with plant-based feed combinations using wheat bran (W), brewer's yeast (Y), potatoes (P), and carrots (C). While substrate combinations WYP and WYC contained 179.58 kcal and 168.71 kcal respectively, TM larvae achieved 701.64 kcal and 708.26 kcal when fed with WYP and WYC, compared to 689.27 kcal in the WY-based agar–agar control group. Nevertheless, elevated protein and lipid levels have been found in other developmental stages and by-products, indicating potential for broader utilization in subsequent recycling processes. In light of the possibility of modulating nutritional composition through substrate manipulation, TM meal has been incorporated into fish diets at different FM replacement levels, showing no significant reduction in growth performance in both freshwater and marine species. In this regard, noteworthy trials include rainbow trout (*Oncorhynchus mykiss*) at 20%, 30%, 60%, and 100% replacement (Rema et al. 2019); red sea bream (*Pagrus major*) at 38%, 60%, and 100% replacement (Ido et al. 2019); and African sharptooth catfish (*Clarias gariepinus*) at 20%, 40%, 60%, and 80% replacement (Ng et al. 2001). However, in rainbow trout, despite good tolerance of FM replacement levels with respect to growth performance, the fatty acid profile has been shown to change markedly, with C16:0, C18:1n9, and C18:2n6 levels increasing, and eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) levels decreasing in response to higher levels of TM inclusion (Iaconisi et al. 2018).

Compared to traditional sources that have characterized Aquafeed 1.0 and Aquafeed 2.0, biocircular ingredients, particularly insects, have introduced novel biochemical components, which are generally absent in other meals (e.g., chitin and unique fatty acids), and greater composition variability, which is influenced by rearing substrate dependence and meal processing procedures. Given the influence of such factors on microbial communities, insect incorporation into aquaculture feeds has necessitated further analysis to assess fish microbiota responses, host health implications, and nutritional outcome consistency. The need for such characterization breadth has resulted in the generation of substantial data volumes, primarily through metabarcoding and metagenomics sequencing. However, in contrast to human and livestock microbiome research, in which both machine learning (ML) and deep learning (DL) have been extensively leveraged as analytical tools, these computational approaches have seen limited use in elucidating the intricate dynamics of fish microbiota. Indeed, existing literature on ML applications in fish microbiome studies is restricted to the investigations of microbiota composition in relation to various environmental variables (Turner et al. 2022; Zhang et al. 2024) and to network interaction modeling between microbial taxa and biotic/abiotic parameters associated with rearing system conditions

(Soriano et al. 2023). Additionally, research efforts have been disproportionately skewed toward intestinal microbial communities, often overlooking other anatomical regions such as the skin, whose microbiota remains substantially unmapped (Gomez and Primm 2021). Despite fish microbiome analysis lagging behind human and livestock research, AI models have been successfully applied in other aquaculture contexts, encompassing both image-based (Barbedo 2022; Gladju et al. 2022; Yang et al. 2021) and non-image-based (Gladju et al. 2022) implementations, with a particular emphasis on feeding frequency optimization (Budaev et al. 2025; Huang et al. 2025; Saad et al. 2024). Given the significant growth in computational power and the increasing accessibility of programming interfaces, the present juncture offers a compelling opportunity to advance our understanding of fish microbiota through AI methodologies.

To the best of our knowledge, at the time this study was conducted, no papers had been published on the application of artificial intelligence, especially machine learning, to investigate the relationship between fish microbiota and aquafeed composition, despite growing scientific interest in ML-driven design of environmentally efficient aquafeeds (Cooney et al. 2021), in light of the recent successes in livestock feeds (Garcia-Launay et al. 2018). In the companion article to the present study (Terova et al. 2021), FM substitution with insect meal derived from partially defatted yellow mealworm did not produce significant alterations in bacterial richness and diversity in skin and gut microbiota from rainbow trout, aside from minor dietary modulation effects on bacterial communities. Nevertheless, we hypothesized the existence of underlying correlation patterns between microbial abundance and dietary regime (TM0, no FM-TM substitution; TM100, full FM-TM substitution). Accordingly, we used ML techniques to examine the relationship between microbial abundance and feed composition, and thus address the following research questions (RQs): *RQ1*, is there a minimal subset of microbial taxa that could act as biomarkers for FM-TM substitution?; *RQ2*, can a given microbial composition, expressed as abundance counts, be reliably associated with a particular dietary regime?; *RQ3*, can shifts in feed ingredient quantities result in statistically significant variations in microbial abundance? Addressing these research questions is anticipated to generate fundamental insights into fish biology, with particular emphasis on the integrated functioning of fish and their microbial communities as a metaorganism. Moreover, the resulting findings could guide the development of microbiome-driven strategies to enhance productivity in the aquaculture industry at both regional and global scales, while, at the same time, fostering environmentally responsible practices to mitigate the escalating challenges posed by climate change. However, the machine-learning outputs in this study should be considered preliminary observations rather than generalizable biological patterns, due to the limited sample size.

Materials and methods

In Terova et al. (2021), two extruded diets were formulated: a control diet (TM0, 0% FM replacement) and an experimental diet (TM100, 100% FM replacement). The experimental protocol complied with European Directive 86/609/EEC and Italian law (D.L. 116/92) and was approved by DISAFA Ethical Committee (Protocol 143811). Rainbow trout (mean initial weight 78.3 ± 6.24 g) were reared for 22 weeks in 6 400-L flow-through tanks (3 tanks per diet, 21 fish per tank) under stable water conditions (water temperature 13 ± 1 °C; water inflow 8 L/min; dissolved oxygen 7.6–8.7 mg/L; pH 7.5–7.6). Fish were fed twice daily, 6 days per week, at 1.6% biomass for 8 weeks, then reduced to 1.4%; they were weighed

biweekly and mortality was monitored daily. At trial end, 6 fish per diet were sacrificed (MS-222, 500 mg/L) for microbiota sampling from gut mucosa (whole intestine, excluding pyloric caeca) and skin mucus. DNA was extracted with the DNeasy PowerSoil Kit, quantified, and stored at − 20 °C. 16S rRNA gene libraries targeting the V3–V4 variable region (primers Pro341F and Pro805R) were prepared with Platinum Taq High Fidelity, indexed with Nextera XT adapters, quality-checked by qPCR, pooled equimolarly, and sequenced on the Illumina MiSeq platform. Sequencing data were deposited in the European Nucleotide Archive (ENA) under accession code PRJEB38845. Raw reads were processed in QIIME2 2018.8 (Bolyen et al. 2019) by an external service: barcode sequences and primers were removed with Cutadapt; sequences were quality-filtered ($Q > 30$), trimmed, and merged with DADA2; high-quality reads were dereplicated, chimeras removed, and sequences clustered into operational taxonomic units (OTUs) at 99% similarity, filtered at 0.005% frequency; rarefaction was performed to normalize samples. Taxonomic assignment utilized Greengenes 13.8 (DeSantis et al. 2006), excluding chloroplast and mitochondrial reads. The use of the Greengenes 13.8 database may limit taxonomic resolution and therefore requires cautious interpretation of some model outputs. Although Greengenes 13.8 is no longer maintained, it was the only fully integrated database available at the time the companion study (Terova et al. 2021) was conducted. Alpha diversity (observed OTUs, Shannon, Pielou's evenness, Faith's PD) and beta diversity (weighted/unweighted UniFrac) were computed, together with statistical tests including Kruskal–Wallis, Adonis, and analysis of similarity (ANOSIM).

Building on the obtained results, we analyzed skin and gut microbiota of rainbow trout following FM replacement with TM, using machine learning on microbial abundance data derived from 16S rRNA gene sequencing. In particular, the reference study targeted the V3–V4 region for metabarcoding, as it provides an optimal balance between taxonomic resolution and community-level diversity comparisons relative to other single- or paired-region targets (Abellan-Schneyder et al. 2021). OTU taxonomic classifications were used without further modification or subset pooling. While microbial counts were available at various taxonomic resolutions (phylum, class, order, family, genus, and species), we focused on lower levels (family, genus, and species) so as to enhance specificity as well as biological relevance (Wakita et al. 2018). Given the limited resolution achieved at the species level, however, ML analyses carried out at this resolution were considered primarily indicative, and manual amplicon-to-species assignment verification was deemed unnecessary in light of the known limitations of the Illumina MiSeq platform in resolving closely related microbial species from metabarcoding-derived reads due to insufficient unique sequence variation (Biada et al. 2025; Buetas et al. 2024).

To implement our analytical pipeline, we selected Python 3.13.3 as reference programming language due to its extensive library support for data analysis. Our approach leveraged the following libraries: Pandas 2.2.3 (McKinney 2010), SciPy 1.15.2 (Virtanen et al. 2020), Matplotlib 3.10.1 (Hunter 2007), Seaborn 0.13.2 (Waskom 2021), scikit-learn 1.6.1 (Pedregosa et al. 2011), XGBoost 3.0.0 (Chen and Guestrin 2016), CatBoost 1.2.8 (Prokhorenkova et al. 2018), SHAP 0.46.0 (Lundberg and Lee 2017). Given the limited sample size (12 fish, 6 fish per diet), we made use of leave-one-out cross-validation (LOOCV) to maximize information utilization as well as minimize prediction biases, despite the high dimensionality of microbial variables across taxonomic levels (46 families, 63 genera, 80 species for gut microbiota; 112 families, 206 genera, 279 species for skin microbiota). Because LOOCV evaluates models using only one sample per iteration, it can increase estimate variance in small datasets; therefore, the resulting performance metrics should be interpreted with caution. For model performance assessment throughout the pipeline, we applied evaluation metrics appropriate

for balanced datasets: for classification, accuracy (ACC) and Matthews correlation coefficient (MCC); for regression, the coefficient of determination (R^2), the mean absolute error (MAE), which assigns equal weight to all errors, and the root mean squared error (RMSE), which assigns greater weight to larger discrepancies. In light of the exploratory character of this study, we opted to determine the highest performing model through a general comparison of performance metrics rather than inferential statistics. In particular, accuracy was prioritized for classification tasks, while R^2 was prioritized for regression tasks, as both are widely recognized indicators of predictive reliability. This strategy allowed for the identification of promising modeling approaches while avoiding additional statistical analyses, acknowledging that more rigorous statistical validation would be required in future confirmatory studies. Finally, prior to performing predictive modeling, we preprocessed the dataset to address missing values, null columns, and other data inconsistences, thus ensuring data integrity and reliability. Within this methodological framework, we addressed the research questions as outlined below.

RQ1

As a fundamental stage in data preprocessing, we performed feature selection to determine the most influential microbial features, enhancing generalizability and mitigating overfitting. For *RQ2*, we used recursive feature elimination (RFE) with tree-based, ensemble-based, and generalized linear models as estimators. To ensure balanced selection between permissive and conservative algorithms, we retained microbial taxa with the highest overlap across estimators, under the constraint that, given n as the number of features and m as the number of samples, $1 \leq n < m$. For *RQ3*, we identified statistically significant microbial taxa exhibiting moderate-to-high correlation with the dietary regime through point-biserial correlation by imposing threshold criteria of $|r| > 0.5$ and $p < 0.05$. In both cases, any non-microbial contaminants or ambiguous taxa were manually removed.

RQ2–RQ3

Both research questions were addressed using a common three-stage pipeline, which was adjusted to meet task-specific needs. In the first stage, models were evaluated on the full dataset using default implementations to define baseline performance. In the second stage, the dataset was reduced to include microbial taxa identified during feature selection, and models were evaluated using default implementations to assess overfitting. Given the limited sample size, these pipeline stages used cross-validation (CV) instead of traditional train-test splitting to reduce variance in performance estimates caused by different split proportions. In the third stage, models were evaluated on the reduced dataset following nested cross-validation (NCV), which allowed for unbiased performance evaluation while, at the same time, performing hyperparameter tuning. Importantly, due to the current unavailability of a suitable external dataset, we used the same dataset to internally validate the ML models. While both research objectives share a common framework, task-specific implementation requirements were also taken into consideration. In the case of *RQ2*, SHAP analysis was performed to visualize the individual contribution of the selected features in the highest-performing model identified with NCV. In the case of *RQ3*, regression analysis required extending the dataset to incorporate the respective quantities of the dietary formulation ingredients reported in Terova et al. (2021).

Since most algorithms have been designed to handle 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 *RQ2*, 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), logistic regression (LOGREG), and MLP classifier (MLPC). In the case of *RQ3*, instead, as a regression task, we used the following algorithms: DT regressor (DTR), RF regressor (RFR), ET regressor (ETR), GB regressor (GBR), XGB regressor (XGBR), SVM regressor (SVR), *k*-NN regressor (KNR), and MLP regressor (MLPR). In both cases, when using SVM-based algorithms, we assessed model performance across different 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 higher model generalizability, effectively mitigating the overfitting risks associated with PK and RK as well as the unstable behavior of SK. Ultimately, to guarantee proper learning and unbiased performance, datasets were standardized with StandardScaler from scikit-learn when using scale-sensitive algorithms; in particular, StandardScaler applies *z*-score normalization independently to each feature, ensuring balanced contributions across features as well as improving convergence speed and model performance.

Results

To systematically characterize microbiota composition, we performed task-specific feature selection and applied a structured three-stage pipeline across taxonomic levels (family, genus, species) using microbial count data from skin and gut microbiota. For conciseness, we present results from the final pipeline stage, in which selected models underwent full optimization. The high accuracy values may reflect a close fit to the current dataset and do not necessarily indicate a confirmed biological distinction. Detailed assessments of earlier stages, such as baseline performance and overfitting probing, are provided as supplementary materials (Online Resource 1 and Online Resource 5 for classification and regression on gut microbiota; Online Resource 3 and Online Resource 6 for classification and regression on skin microbiota). In order to better contextualize ML-derived results, we summarize the main findings from traditional microbiota analyses, as detailed in Terova et al. (2021). This preliminary overview provides a necessary reference framework for interpreting the outcomes obtained from ML-based approaches.

Results from previous microbiota analysis

As reported in our parallel study article, when limited to the most representative taxa, gut microbiota consisted of 6 families and 2 genera. However, when assessing diet-related effects on taxa abundance variation, Neisseriaceae showed statistical significance ($p = 0.033$), as observed in higher taxonomic levels (Neisseriales, $p = 0.033$; Betaproteobacteria, $p = 0.012$; Proteobacteria, $p = 0.047$), with no statistical significance detectable

at the genus level. In relation to microbial diversity, alpha diversity indices revealed no statistically significant variation in response to diet, while beta diversity analyses indicated significant differences between microbial communities in gut mucosa in function of diet for weighted UniFrac analysis (Adonis, $p = 0.025$; ANOSIM, $p = 0.038$), but no effect on the microbial communities associated with skin mucus was found. For further information on the statistical analyses performed, we advise to refer to the reference study (Terova et al. 2021).

In contrast, skin microbiota, which is usually characterized by greater bacterial richness and diversity, consisted of 25 families and 20 genera. In this context, diet effect on microbial abundance variation was statistically significant for *Deefgea* ($p = 0.017$), with analogous significance detected at higher taxonomic levels (Neisseriaceae, $p = 0.013$; Neisseriales, $p = 0.013$); additionally, Clostridiaceae abundance exhibited significant association with diet ($p = 0.013$). On the other hand, with regard to microbial diversity, alpha diversity indices showed no statistically significant difference in response to diet, with beta diversity analyses equally showing no diet effect on skin microbiota composition.

Inferring dietary regimes from microbial abundance

Classification modeling of gut microbiota

Integrating RFE-based feature selection with nested cross-validation resulted in optimal classification performance across taxonomic levels (Table 1). In terms of average performance, MNB consistently outperformed other models, with KNC and MLPC emerging as strong competitors at the genus level. Among the highest-performing models, ensemble-based (RFC, ETC, GBC, CBC) and distance-based (KNC) approaches achieved perfect classification; in addition, margin-based (SVC-LK) and neural-network-based (MLPC) models demonstrated comparable accuracy at the genus levels. However, these values should be viewed as exploratory given the limited dataset size.

The elevated model accuracy was primarily attributable to the selective power of the RFE procedure, which effectively distilled influential microbial features into minimal subsets across taxonomic ranks (Table 2), enabling more nuanced interpretation through SHAP analysis. At the family level, separate groupings were identified: Chitinibacteraceae and Clostridiaceae, representing high-importance taxa; Lactobacillaceae and Neisseriaceae, representing low-importance taxa (Fig. 1a). At the genus level, although the high number of microbial features introduced interpretational complexities, sufficiently stable patterns emerged, with *Clostridium* and *Deefgea* exerting notable influence, and *Cutibacterium* consistently ranking among the least impactful taxa (Fig. 1b). At the species level, importance-based feature evaluation did not yield additional interpretative value, besides highlighting the relatively low contribution of *Lactobacillus aviarius* (Fig. 1c). For comprehensive information on the magnitude of feature impact at single-sample resolution across taxonomic levels, refer to Online Resource 2. However, these values should be interpreted as exploratory due to the limited dataset size and the potential for model overfitting.

Classification modeling of skin microbiota

Applying the same methodology used for gut microbiota, consistently high performance was achieved on skin microbiota (Table 3). Regarding overall reliability, KNC proved most effective at genus and species levels, while ETC, CBC, MLPC, and LOGREG

Table 1 Performance of optimized classification models for rainbow trout gut microbiota ($n = 12$) under FM-based (TM0) and TM-based (TM100) diets (metrics: AVG NCV, average classification accuracy during nested cross-validation and standard error; BM ACC, classification accuracy of the best-performing model returned by nested cross-validation on the whole dataset; BM MCC, Matthews correlation coefficient of the best-performing model returned by nested cross-validation on the whole dataset; classification algorithms: 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-LK, support vector machine classifier using the linear kernel; KNC, *k*-nearest neighbor classifier; LOGREG, logistic regression; MLPC, multilayer perceptron classifier)

		DTC	RFC	ETC	GBC	XGBC	CBC	MNB	SVC-LK	KNC	LOGREG	MLPC
Family	**AVG NCV**	0.33±0.14	0.67± 0.14	0.67±0.14	0.58±0.15	0.67±0.14	0.58±0.15	0.83±0.11	0.50±0.15	0.58±0.15	0.42± 0.15	0.58±0.15
	BM ACC	0.92	1.00	1.00	1.00	0.83	1.00	0.83	0.83	1.00	0.83	0.83
	BM MCC	0.85	1.00	1.00	1.00	0.67	1.00	0.71	0.67	1.00	0.67	0.67
Genus	**AVG NCV**	0.33±0.14	0.58±0.15	0.67±0.14	0.58±0.15	0.58±0.15	0.42±0.15	0.75±0.13	0.42±0.15	0.75±0.13	0.58±0.15	0.75±0.13
	BM ACC	0.92	1.00	1.00	1.00	0.92	1.00	0.75	1.00	0.83	0.92	1.00
	BM MCC	0.85	1.00	1.00	1.00	0.85	1.00	0.58	1.00	0.67	0.85	1.00
Species	**AVG NCV**	0.58±0.15	0.67±0.14	0.58±0.15	0.50±0.15	0.58±0.15	0.58±0.15	0.75±0.13	0.50±0.15	0.58±0.15	0.67±0.14	0.67±0.14
	BM ACC	0.92	1.00	1.00	1.00	0.92	1.00	0.75	0.83	1.00	0.83	0.83
	BM MCC	0.85	1.00	1.00	1.00	0.85	1.00	0.58	0.71	1.00	0.71	0.67

Table 2 RFE-selected microbial taxa across taxonomic levels in rainbow trout gut microbiota ($n = 12$) under FM-based (TM0) and TM-based (TM100) diets. The table columns contain the following information: (i) the taxonomic level; (ii) the number of OTU-related features before feature selection; (iii) the number of OTU-related features after feature selection; (iv) the name of the microbial taxa retained after feature selection

Taxonomic level	Total features	Selected features	Feature name
Family	46	4	Chitinibacteraceae, Clostridiaceae Lactobacillaceae, Neisseriaceae
Genus	63	8	*Aeromonas*, *Clostridium* sensu stricto 1, *Cutibacterium*, *Deefgea*, *Enhydrobacter*, *Escherichia-Shigella*, *Lactobacillus*, unknown Ruminococcaceae
Species	80	4	*Lactobacillus aviarius*, unknown *Clostridium*, unknown *Deefgea*, unknown *Mycoplasma*

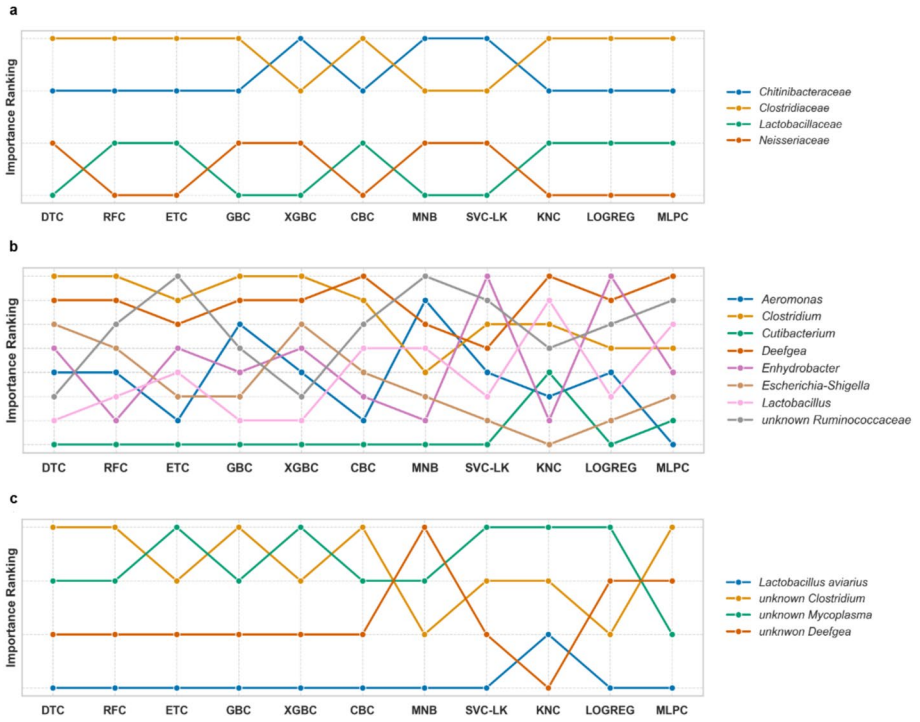

Fig. 1 Categorical parallel coordinate plot illustrating variation in feature importance across multiple classification models for rainbow trout gut microbiota ($n = 12$) under FM-based (TM0) and TM-based (TM100) diets. Each panel presents a summarized visualization derived from SHAP beeswarm plots, showing overall variation in feature importance among microbial taxa selected through feature selection across taxonomic levels (**a** family; **b** genus; **c** species). The *x*-axis indicates classification models, whereas the *y*-axis ranks microbial features in descending order of importance, with higher-importance ones at the top and lower-importance ones at the bottom (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-LK, support vector machine classifier using the linear kernel; KNC, *k*-nearest neighbor classifier; LOGREG, logistic regression; MLPC, multilayer perceptron classifier)

demonstrated superior accuracy at the family level. Among top-performing models, ensemble-based approaches (RFC, ETC, GBC, XGBC, CBC) achieved perfect accuracy at genus and species levels; however, at the family level, ensemble-based (RFC, GBC) and distance-based (KNC) methods performed equally well. However, accuracy values should be viewed as exploratory and may overrepresent model fit given the small sample size.

Although the skin microbiota remains largely unmapped in fish species, the RFE procedure enabled the identification of interesting feature subsets across taxonomic levels (Table 4). However, these taxa should be interpreted cautiously due to the limited dataset and potential variance in small-sample feature selection.

At the family level, Clostridiaceae and Streptococcaceae showed balanced contribution to sample classification (Fig. 2a). At the genus level, when compared to the gut microbiota, the lower number of microbial features facilitated interpretability, revealing two major groups: *Sphingopyxis* emerged as the most influential, whereas *Clostridium* and *Deefgea* were characterized by lower importance scores (Fig. 2b). For detailed insights into the

Table 3 Performance of optimized classification models for rainbow trout skin microbiota ($n = 12$) under FM-based (TM0) and TM-based (TM100) diets (metrics: AVG NCV, average classification accuracy during nested cross-validation and standard error; BM ACC, classification accuracy of the best-performing model returned by nested cross-validation on the whole dataset; BM MCC, Matthews correlation coefficient of the best-performing model returned by nested cross-validation on the whole dataset; classification algorithms: 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-LK, support vector machine classifier using the linear kernel; KNC, *k*-nearest neighbor classifier; LOGREG, logistic regression; MLPC, multilayer perceptron classifier)

		DTC	RFC	ETC	GBC	XGBC	CBC	MNB	SVC-LK	KNC	LOGREG	MLPC
Family	**AVG NCV**	0.83±0.11	0.83±0.11	0.92±0.08	0.67±0.14	0.83±0.11	0.92±0.08	0.83±0.11	0.50±0.15	0.58±0.15	0.92±0.08	0.92±0.08
	BM ACC	0.92	1.00	0.92	1.00	0.83	0.92	0.92	0.92	1.00	0.92	0.92
	BM MCC	0.85	1.00	0.85	1.00	0.67	0.85	0.85	0.85	1.00	0.85	0.85
Genus	**AVG NCV**	0.75±0.13	0.83±0.11	0.75±0.13	0.67±0.14	0.75±0.13	0.75±0.13	0.92±0.08	0.83±0.11	0.83±0.11	0.83±0.11	0.83±0.11
	BM ACC	0.92	1.00	1.00	1.00	1.00	1.00	0.92	0.92	0.83	0.83	0.83
	BM MCC	0.85	1.00	1.00	1.00	1.00	1.00	0.85	0.85	0.71	0.71	0.67
Species	**AVG NCV**	0.75±0.13	0.75±0.13	0.83±0.11	0.75±0.13	0.75±0.13	0.75±0.13	0.92±0.08	0.67±0.14	0.67±0.14	0.67±0.14	0.67±0.14
	BM ACC	0.92	1.00	1.00	1.00	1.00	1.00	0.92	0.83	0.92	0.92	0.92
	BM MCC	0.85	1.00	1.00	1.00	1.00	1.00	0.85	0.71	0.85	0.85	0.85

Table 4 RFE-selected microbial taxa across taxonomic levels in rainbow trout skin microbiota ($n = 12$) under FM-based (TM0) and TM-based (TM100) diets. The table columns contain the following information: (i) the taxonomic level; (ii) the number of OTU-related features before feature selection; (iii) the number of OTU-related features after feature selection; (iv) the name of the microbial taxa retained after feature selection

Taxonomic level	Total features	Selected features	Feature name
Family	112	2	Clostridiaceae, Streptococcaceae
Genus	206	3	*Clostridium* sensu stricto 1, *Deefgea*, *Sphingopyxis*
Species	279	2	unknown *Clostridium*, unknown *Sphingopyxis*

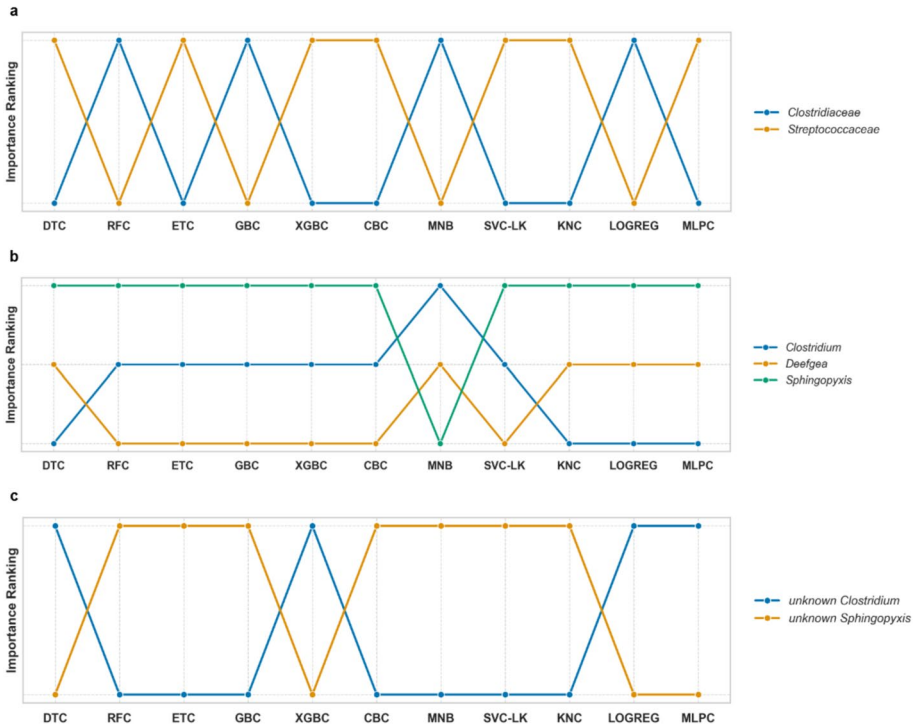

Fig. 2 Categorical parallel coordinate plot illustrating variation in feature importance across multiple classification models for rainbow trout skin microbiota ($n = 12$) under FM-based (TM0) and TM-based (TM100) diets. Each panel presents a summarized visualization derived from SHAP beeswarm plots, showing overall variation in feature importance among microbial taxa selected through feature selection across taxonomic levels (**a** family; **b** genus; **c** species). The *x*-axis indicates classification models, whereas the *y*-axis ranks microbial features in descending order of importance, with higher-importance ones at the top and lower-importance ones at the bottom (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-LK, support vector machine classifier using the linear kernel; KNC, *k*-nearest neighbor classifier; LOGREG, logistic regression; MLPC, multilayer perceptron classifier)

magnitude of feature influence at single-sample resolution across taxonomic levels, refer to Online Resource 4.

Predicting microbial variations based on feed composition changes

Regression modeling of gut microbiota

The integration of point-biserial correlation with nested cross-validation produced comparable results across taxonomic levels (Table 5). On average, most regression models demonstrated a moderate fit, whereas SVR-LK and MLPR consistently demonstrated the lowest performance, which is marked by the negative values of the R^2 metric, indicating that both models provide a poorer fit to the data than the baseline least-squares predictor, which relies solely on the mean estimate. In this regard, it should be noted that regression algorithms utilized the diet type as an intermediate variable for taxa selection via point-biserial correlation before replacing it with the feed ingredient quantities defined in Terova et al. (2021) for regression. In the case of gut microbiota, correlation analysis consistently identified hierarchically related taxa across taxonomic levels (Table 6). When evaluating correlation estimations and model performance metrics, despite suboptimal goodness of fit, error metrics remained elevated, particularly for *Clostridium*-related taxa, while error values were comparatively more acceptable for *Mycoplasma*-related ones. However, all estimates should be regarded as exploratory and not indicative of confirmed predictive relationships.

Regression modeling of skin microbiota

When using the same regression strategy applied to gut microbiota, similar outcomes were obtained for skin microbiota, with SVR-LK and MLPR continuously yielding the poorest performance (Table 7). In this case, point-biserial correlation analysis identified a broader array of microbial taxa, driven by the pronounced increase in variables across taxonomic levels, when compared to gut microbiota; nonetheless, taxa affiliated with Chitinibacteraceae and Enterobacteriaceae were frequently detected (Table 8). When combining correlation estimates and model performance metrics, R^2 values showed suboptimal predictive power, comparable to that observed for gut microbiota, although error metrics were noticeably reduced, with a marked decline observed when moving from family to species level. Despite this encouraging trend, error values remained relatively high for most taxa, except for those belonging to Chitinibacteraceae, which exhibited the highest recorded abundance. However, these associations should be considered exploratory and require confirmation in larger datasets.

Overall performance of predictive models in classification and regression

Given that multiple algorithmic families were leveraged, it is important to provide a general overview of their performance on microbiota data derived from different anatomical regions. On the one hand, for classification analyses (Tables 1 and 3), the selected classifiers exhibited broadly comparable performance when applied to data from both regions, though some differences merit closer attention. Considering average performance, measured in terms of accuracy, classification tended to perform better for skin microbiota than for gut microbiota. This trend was particularly evident for tree-based (DTC) and ensemble-based (RFC, ETC, XGBC, CBC) models, and, to a lesser extent, for generalized linear

Table 5 Performance of optimized regression models for rainbow trout gut microbiota ($n = 12$) under FM-based (TM0) and TM-based (TM100) diets (metrics: AVG NCV, average root mean squared error during nested cross-validation and standard error; BM R^2, coefficient of determination of the best-performing model returned by nested cross-validation on the whole dataset; BM MAE, mean absolute error of the best-performing model returned by nested cross-validation on the whole dataset; BM RMSE, root mean squared error of the best-performing model returned by nested cross-validation on the whole dataset; regression algorithms: DTR, decision tree regressor; RFR, random forest regressor; ETR, extra tree regressor; GBR, gradient boosting regressor; XGBR, extreme gradient boosting regressor; SVR-LK, support vector machine regressor using the linear kernel; KNR, *k*-nearest neighbor regressor; MLPR, multilayer perceptron regressor)

		DTR	RFR	ETR	GBR	XGBR	SVR-LK	KNR	MLPR
Family	**AVG NCV**	2348.78±640.00	2349.19±642.42	2348.78±640.00	2348.79±639.99	2348.78±640.00	2666.85±902.66	2348.78±640.00	7777.34±1061.76
	BM R^2	0.37	0.36	0.37	0.37	0.37	−0.09	0.37	−2.51
	BM MAE	1957.07	1959.76	1957.07	1957.08	1957.07	2545.90	1967.93	7777.58
	BM RMSE	2744.50	2753.25	2744.50	2744.49	2744.50	3999.65	2744.89	8543.24
Genus	**AVG NCV**	2348.78±640.00	2349.19±642.42	2348.78±640.00	2348.79±639.99	2348.78±640.00	2666.85±902.66	2348.78±640.00	7777.34±1061.76
	BM R^2	0.37	0.36	0.37	0.37	0.37	−0.09	0.37	−2.51
	BM MAE	1957.07	1959.76	1957.07	1957.08	1957.07	2545.90	1967.93	7777.58
	BM RMSE	2744.50	2753.25	2744.50	2744.49	2744.50	3999.65	2744.89	8543.24
Species	**AVG NCV**	2348.78±640.00	2349.19±642.42	2348.78±640.00	2348.79±639.99	2348.78±640.00	2666.85±902.66	2348.78±640.00	7777.34±1061.76
	BM R^2	0.37	0.36	0.37	0.37	0.37	−0.09	0.37	−2.51
	BM MAE	1957.07	1959.76	1957.07	1957.08	1957.07	2545.90	1967.93	7777.58
	BM RMSE	2744.50	2753.25	2744.50	2744.49	2744.50	3999.65	2744.89	8543.24

Table 6 Microbial taxa selected through point-biserial correlation across taxonomic levels in rainbow trout gut microbiota ($n = 12$) under FM-based (TM0) and TM-based (TM100) diets. The table columns contain the following information: (i) the taxonomic level; (ii) the number of OTU-related features before feature selection; (iii) the number of OTU-related features after feature selection; (iv) the name of the microbial taxa retained after feature selection; (v) the Pearson correlation coefficient (or *r* value) returned by point-biserial correlation for the selected feature; (vi) the *p*-value returned by point-biserial correlation for the selected feature; (vii) the highest microbial count within the dataset for the selected feature

Taxonomic level	Total features	Selected features	Feature name	*r* value	*p* value	*Max* value
Family	46	2	Clostridiaceae	0.581	0.047	83
			Mycoplasmataceae	0.662	0.019	21,108
Genus	63	2	*Clostridium* sensu stricto 1	0.581	0.047	83
			Mycoplasma	0.662	0.019	21,108
Species	80	2	unknown *Clostridium*	0.581	0.047	83
			unknown *Mycoplasma*	0.662	0.019	21,108

(LOGREG) and neural-network-based (MLPC) models. A notable exception included MNB, which achieved extremely similar accuracy across both anatomical sites. When focusing on the best-performing classifier, each model reached very high accuracy on gut and skin microbiota; however, ensemble-based models (RFC, ETC, GBC, CBC) attained perfect accuracy in the majority of cases. On the other hand, for regression analyses (Tables 5 and 7), when considering the R^2 of the best-performing model, regressors performed similarly in both gut and skin microbiota, with SVR-LK and MLPR consistently ranking as the lowest-performing ones.

These findings align with the advantages and disadvantages of the respective algorithmic families. Ensemble-based models consistently delivered better performance, particularly in classification tasks, due to their ability to reduce overfitting and maintain robustness to noise, albeit at the cost of greater computational requirements and reduced interpretability compared to single-tree models. Probability-based (restricted to classification), distance-based, and generalized linear models also performed well, offering fast training and high interpretability, though they remain sensitive to correlated or irrelevant features. While yielding acceptable results in classification, margin-based and neural-network-based models consistently underperformed in regression. This outcome could be related to their theoretical properties, as margin-based methods perform better in high-dimensional spaces and neural networks generally require large datasets to achieve optimal training, leading to reduced performance in smaller datasets such as the one analyzed in the present study.

Discussion

With aquafeed production progressively phasing out marine resources toward more environmentally sustainable solutions, insects have emerged as a suitable alternative, supporting sustainable farming practices as well as maintaining fish health. In particular, yellow mealworm has garnered particular attention due to its favorable nutritional profile and minimal environmental footprint when compared to conventional animal- and plant-based protein sources. In the context of aquaculture, the increasing integration of yellow mealworm in feeds has required a thorough assessment of FM-TM replacement diets. While the

Table 7 Performance of optimized regression models for rainbow trout skin microbiota ($n = 12$) under FM-based (TM0) and TM-based (TM100) diets (metrics: AVG NCV, average root mean squared error during nested cross-validation and standard error; BM R^2, coefficient of determination of the best-performing model returned by nested cross-validation on the whole dataset; BM MAE, mean absolute error of the best-performing model returned by nested cross-validation on the whole dataset; BM RMSE, root mean squared error of the best-performing model returned by nested cross-validation on the whole dataset; regression algorithms: DTR, decision tree regressor; RFR, random forest regressor; ETR, extra tree regressor; GBR, gradient boosting regressor; XGBR, extreme gradient boosting regressor; SVR-LK, support vector machine regressor using the linear kernel; KNR, *k*-nearest neighbor regressor; MLPR, multilayer perceptron regressor)

		DTR	RFR	ETR	GBR	XGBR	SVR-LK	KNR	MLPR
Family	**AVG NCV**	380.17±95.09	395.51±93.03	380.17±95.09	380.17±95.09	380.17±95.09	675.25±68.12	380.17±95.09	610.56±163.42
	BM R^2	0.38	0.36	0.38	0.38	0.38	−0.26	0.37	−1.26
	BM MAE	316.39	314.25	316.81	316.81	316.81	465.85	332.78	611.01
	BM RMSE	428.39	432.73	428.39	428.39	428.39	608.81	429.73	816.57
Genus	**AVG NCV**	144.58±35.32	149.95±34.58	144.58±35.32	144.58±35.32	144.58±35.32	242.55±24.72	144.58±35.32	230.40±61.49
	BM R^2	0.37	0.37	0.37	0.37	0.37	0.14	0.36	−0.71
	BM MAE	121.67	120.66	121.67	121.67	121.67	169.33	127.17	230.72
	BM RMSE	163.48	164.94	163.48	163.48	163.48	223.70	164.00	309.48
Species	**AVG NCV**	92.05±21.48	95.16±21.09	92.05±21.48	92.05±21.48	92.05±21.48	155.34±15.24	92.05±21.48	130.16±37.53
	BM R^2	0.34	0.34	0.34	0.34	0.34	0.12	0.34	−0.36
	BM MAE	77.39	76.11	77.39	77.39	77.39	105.25	80.03	130.21
	BM RMSE	103.15	104.07	103.15	103.15	103.15	141.92	103.36	182.27

Table 8 Microbial taxa selected through point-biserial correlation across taxonomic levels in rainbow trout skin microbiota ($n = 12$) under FM-based (TM0) and TM-based (TM100) diets. The table columns contain the following information: (i) the taxonomic level; (ii) the number of OTU-related features before feature selection; (iii) the number of OTU-related features after feature selection; (iv) the name of the microbial taxa retained after feature selection; (v) the Pearson correlation coefficient (or *r* value) returned by point-biserial correlation for the selected feature; (vi) the *p*-value returned by point-biserial correlation for the selected feature; (vii) the highest microbial count within the dataset for the selected feature

Taxonomic level	Total features	Selected features	Feature name	*r* value	*p* value	*Max* value
Family	112	1	Chitinibacteraceae	−0.652	0.022	1426
Genus	206	3	*Deefgea*	−0.652	0.022	1426
			Flectobacillus	0.603	0.038	17
			unknown Enterobacteriaceae	−0.648	0.023	207
Species	279	5	unknown *Chryseobacterium*	−0.594	0.042	39
			unknown *Deefgea*	−0.640	0.025	1370
			unknown Enterobacteriaceae	−0.648	0.023	207
			unknown *Flectobacillus*	0.603	0.038	17
			unknown *Lactococcus*	−0.590	0.043	10

impact of such diets on growth performance has been extensively documented, their influence on host-associated microbiomes remains comparatively underexplored.

In our parallel study (Terova et al. 2021), we examined the dietary effects of FM-TM replacement on rainbow trout microbiota composition, using 16S rRNA gene sequencing and bioinformatic tools to characterize gut and skin microbial communities. While this research demonstrated the suitability of yellow mealworm as protein source in aquaculture feeds, the data showed only marginal effects of full substitution on bacterial richness and diversity across both anatomical sites. In light of the limited diet-induced modulation observed through conventional microbiome profiling, we conducted a more in-depth analysis of microbial abundance patterns through artificial intelligence, specifically machine learning algorithms, to investigate the interrelationship between diet composition and host-associated microbiota, as outlined in our research objectives.

While well established in human and animal microbiome research, where it has provided promising insights, artificial intelligence has not yet been fully leveraged to unravel the intricacies of microbial communities characterizing the fish microbiota (Rizzi et al. 2025). Using a broad array of algorithmic families to perform classification (*RQ2*) and regression (*RQ3*) tasks, our study produced noteworthy results. On the one hand, classification analysis produced particularly encouraging results attributable to the binary nature of the classification objective (TM0, TM100), which has significantly simplified predictive modeling; indeed, in skin and gut microbiota, extremely favorable outcomes were achieved across multiple algorithms, with several models attaining perfect classification accuracy. On the other hand, regression analysis performance was generally suboptimal, with relatively acceptable deviation values only for a limited number of taxa identified through feature selection. Despite these limitations, regression results demonstrated potential, underscoring the need for further investigation to enhance predictive capacity. In this regard, understanding how variations in feed ingredient composition are able to influence microbial abundance could enable the development of targeted dietary interventions. Such ingredient-specific modulation strategies could in turn facilitate the deliberate manipulation

of microbial communities toward a desired composition, advancing precision nutrition in aquaculture.

Considering the promising outcomes observed for gut microbiota across both predictive tasks, feature selection had a pivotal role in shaping these results while also mitigating overfitting and enhancing generalizability. Indeed, when prioritizing family and genus as the most informative taxonomic ranks, this procedure enabled the identification of microbial taxa with high frequency (HF) as well as low frequency (LF), relative to those previously reported in Terova et al. (2021). At the family level, HF taxa (Aeromonadaceae, Clostridiaceae, Enterobacteriaceae, Mycoplasmataceae, Neisseriaceae, and Ruminococcaceae) outnumbered LF ones (Chitinibacteraceae, Lactobacillaceae, Moraxellaceae, and Propionibacteriaceae), albeit marginally. At the genus level, the distribution was more balanced: HF taxa with exact (*Deefgea*) and approximate (*Aeromonas*, *Clostridium*, and *Mycoplasma*) frequency were equally represented alongside LF ones (*Cutibacterium*, *Enhydrobacter*, *Escherichia-Shigella*, and *Lactobacillus*). Although ML analyses were extended to the species level, taxonomic resolution was predictably low, with most species still remaining unidentified, except for *Lactobacillus aviarius*. Despite the methodological differences, ML-based results were nevertheless in agreement with those derived from conventional microbiota analyses, which detected diet-associated statistical significance for Neisseriaceae, one of the HF taxa identified through our feature selection procedure.

Although the skin is a distinct anatomical region exposed to ecological pressures different from those of the intestinal environment, its associated microbial communities exhibited a comparably structured compositional profile across taxonomic levels when microbial taxa selected through feature selection methods were interpreted in light of the findings reported in Terova et al. (2021). At the family level, taxa distribution was relatively balanced, with LF families (Chitinibacteraceae, Sphingomonadaceae, Spirosomaceae, and Streptococcaceae) nearly equaling HF ones (Clostridiaceae, Enterobacteriaceae, and Weeksellaceae). At the genus level, HF genera (*Chryseobacterium* and *Deefgea*) were slightly outnumbered by LF ones (*Clostridium*, *Flectobacillus*, *Lactococcus*, and *Sphingopyxis*). Similar to the findings from gut microbiota, ML-derived results exhibited partial concordance with previous microbiota analyses, which identified *Deefgea* and Clostridiaceae as statistically significant features associated with dietary treatment.

Compared to the results presented in our parallel study (Terova et al. 2021), ML-based outcomes show overall consistency across both microbial ecosystems. The robustness of these results is further reaffirmed when compared with existing characterizations of the core microbiota in rainbow trout, as reported for both anatomical levels (Drosdowech et al. 2025; Hines et al. 2023; Lowrey et al. 2015). Notably, more detailed insights emerge when analyses are conducted at lower taxonomic ranks, with the genus level offering significant informational value, especially on LF members.

The gut microbiota revealed microbial genera of ecological and functional relevance. *Cutibacterium* (Propionibacteriaceae), which has been detected in aquatic environments such as zebrafish housing systems, may enter aquaculture systems through human skin contact given its role in the human skin microbiota and biofilm-forming capability (Ericsson et al. 2021). *Enhydrobacter* (Moraxellaceae) is dominant in teleost skin microbiota (Sylvain et al. 2020) but has also been reported in rainbow trout gut microbiota (Betiku et al. 2018), indicating broader distribution. *Lactobacillus* (Lactobacillaceae) is recognized for its role in gastrointestinal health and immune modulation, including the inhibition of pathogens (Govindaraj et al. 2021). In this study, taxonomic resolution also identified *L. aviarius*, a species originally described in avian hosts but also reported in finfish gastrointestinal tracts (Ringø et al. 2018).

The skin microbiota displayed notable ecological features. *Flectobacillus* (Spirosomaceae), detected on rainbow trout skin and gills (de Bruijn et al. 2018; Lowrey et al. 2015), frequently co-occurs with *Flavobacterium*, a genus connected to freshwater fish pathologies (Lee et al. 2023; Wahli and Madsen 2018). *Sphingopyxis* (Sphingomonadaceae), known for pollutant bioremediation in different aquatic habitats (Sharma et al. 2021), has also been reported in the skin microbiota of rainbow trout, zebrafish, and brook char (Coetzer et al. 2021). *Chryseobacterium* (Weeksellaceae) is broadly distributed across aquatic and terrestrial environments (Mwanza et al. 2022), frequently regarded as spoilage organisms (El-Saadony et al. 2021), with *C. shigense* and *C. piscicola* regarded as potential pathogens in rainbow trout (Zamora et al. 2012).

Among the observed similarities between gut and skin microbiota in rainbow trout, the occurrence of *Deefgea* (Chitinibacteraceae) in both anatomical sites supports the hypothesis of microbial exchange between host and aquatic environments, though further validation is needed (Bruno et al. 2023). Most species have been catalogued across fish hosts, including *D. salmonis* and *D. piscis* in rainbow trout (Chen MQ et al. 2022; Takeuchi and Sugahara 2025), *D. rivuli* in brown trout (*Salmo trutta*) (Carbajal-González et al. 2011), *D. tanakiae* in South Korean dark sleeper (*Odontobutis platycephala*) (Gim et al. 2022), and *D. chitinilytica* in gold tench (*Tinca tinca*) and goldfish (*Carassius auratus*) (Jung and Jung-Schroers 2011).

Methodological boundaries and future directions

While this study has deepened our understanding of gut and skin microbiota in rainbow trout, certain limitations remain from computational and biological perspectives.

Computational considerations

This study was conducted on a small sample size (12 fish), which may have resulted in findings that are restricted to this dataset. Although larger sample sizes are generally recommended to strengthen statistical robustness and improve diet-induced microbiome shift detection (Johnson et al. 2020; Jarett et al. 2021), we implemented appropriate computational strategies to minimize dataset-specific bias, including feature selection (to reduce dimensionality), NCV (to separate hyperparameter tuning from model evaluation), and LOOCV (to maximize data utilization) (Kirk et al. 2022).Furthermore, given the highly structured nature of biological data, particularly microbial abundance, we decided against data augmentation, as artificial manipulation may introduce biologically unrealistic distributions that fail to reflect ecological relationships and disregard biological constrains (e.g., co-occurrence patterns and phylogenetic relationships). Nonetheless, ongoing research is exploring specialized augmentation strategies to address these limitations, especially in microbial datasets (Gordon-Rodriguez et al. 2022; Wen et al. 2023, 2024). However, it is important to notice that existing microbiome studies in aquaculture have usually relied on smaller sample sizes than those used in growth-related analyses (Ruiz et al. 2024; Chen X et al. 2022; Turner et al. 2022), despite the definition of a suitable sample size for representative inference in metabarcoding analyses remaining debated. Recent combinatorial studies on European sea bass (*Dicentrarchus labrax*) and gilthead sea bream (*Sparus aurata*) indicate that nine specimens per dietary group may be sufficient to encompass approximately 90% of observed bacterial species richness, thereby providing representative insights into diet-induced microbial shifts (Panteli et al. 2020). In the current study,

while the sample size per diet (6 fish) was not sufficient for confirmatory results, consistent with conclusions by Panteli et al. (2020) for microbiome studies with similar numbers of biological samples (Cerezuela et al. 2013; Carda-Diéguez et al. 2014; Larsen et al. 2015), it nevertheless appeared sufficient to yield preliminary insights into the feasible application of our predictive pipeline. Small sample sizes have also been used in similar pilot studies on livestock (Liu et al. 2022; Wang et al. 2018, 2019) and human (David et al. 2014; Zhou et al. 2018) microbiome. Moreover, we acknowledge that conducting feature selection on the entire dataset before modeling carries a risk of data leakage, whereby selected features may be inadvertently biased toward the dataset from which they were derived. Although such risk can be prevented by performing feature selection within each fold of nested cross-validation, to obtain a common feature set across folds, we performed feature selection before model modeling, thus prioritizing interpretability and consistency. Given the promising results obtained, future work could involve implementing fold-specific feature selection to provide unbiased performance estimates, followed by consensus procedures to produce a common feature set for interpretability. When dealing with larger datasets, feature selection may be restricted to a dedicated subset, given that such data volumes generally guarantee representativeness (Karwowska et al. 2025; Marcos-Zambrano et al. 2021). Lastly, we acknowledge that the definition of a common feature set depends on feature selection procedure implementation; indeed, alternative algorithms, selection functions, and inclusion criteria may produce taxonomic profiles that differ from those reported in this study.

In the context of small datasets, microbiome analyses often rely on statistical and bioinformatic tools to identify and prioritize informative microbial features that differentiate between sample groups. The Analysis of Compositions of Microbiomes with Bias Correction (ANCOM-BC) is a statistical method for differential abundance testing, allowing for the identification of significantly varying taxa through the exclusion of non-significant features and adjustment for sampling bias (Lin and Peddada 2020). Linear Discriminant Analysis Effect Size (LEfSe) instead combines non-parametric statistical testing with supervised machine learning to detect features that exhibit statistical significance and biological consistency across groups (Segata et al. 2011). This approach is especially effective in discriminating among sample classes and facilitating biomarker discovery. For exploratory analyses, the Statistical Analysis of Metagenomic Profiles (STAMP) platform provides a comprehensive suite of tools for the visualization and statistical testing of taxonomic and functional profiles (Parks and Beiko 2010; Parks et al. 2014).

Although these tools are extensively used and provide interpretability through conventional statistical metrics (e.g., *p*-values, confidence intervals, and effect sizes), they generally remain domain-specific and typically assess features independently. As a result, these methodologies may fail to capture both interaction effects and co-occurrence patterns among features, in addition to not being designed for predictive modeling tasks, thereby limiting their applicability within supervised learning contexts. By contrast, ML algorithms offer several advantages. For instance, they can consider all features jointly, capturing complex interactions and nonlinear relationships that may be missed by univariate methods. ML models also provide direct feature importance measures, which facilitate taxa prioritization, and they are well suited to high-dimensional data handling, especially when regularization and ensemble techniques are employed. Given our research objectives, we selected ML models capable of jointly evaluating multiple features while integrating with explanatory frameworks. This strategy was able to balance predictive capability with biological interpretability, offering a complementary perspective to conventional microbiome analysis approaches.

While the present study relied on general-purpose machine-learning models owing to their suitability for predictive modeling, as supported by the noticeable shift in scientific literature toward ML-based approaches in microbiome research, future investigations may benefit from extending this framework into a hybrid analytical strategy that integrates domain-specific methods with ML algorithms within a unified pipeline. This integration has already been exemplified in recent work on colorectal cancer, where LEfSe and ANCOM-BC were applied to identify enriched taxa, followed by the application of a Bayesian ML model to validate and refine findings (Han et al. 2025). In this context, ML offers the scalability and flexibility required to manage and integrate high-dimensional multi-omics datasets (Li et al. 2022), while domain-specific frameworks contribute biological insights grounded in transparent assumptions. For transdisciplinary knowledge transfer, this hybrid implementation could be applied in aquaculture research to investigate microbiota alterations in response to various dietary and rearing conditions. The results obtained in this study offer preliminary support for the feasibility of ML-based approach, and it is increasingly evident that the convergence of general-purpose ML with specialized bioinformatic tools represents a promising paradigm in microbiome analytics. This hybrid framework could extend the current work, especially if larger datasets derived from similar experimental in vivo conditions become available.

Biological considerations

While gut microbiota interactions are comparatively well characterized, skin microbiota composition is shaped by interactions with microbial communities colonizing the rearing environment (Minich et al. 2020). These environmental communities commonly exhibit greater microbial diversity than those associated with fish hosts (Duarte et al. 2019), yet fish skin maintains distinct microbial profiles when compared to the surrounding water (Berggren et al. 2022). Skin microbiota structure is also influenced by captivity status (Uren Webster et al. 2020) as well as species-specific factors (Larsen et al. 2013). Species specificity plays a critical role in structuring gut microbiota, as evidenced in our parallel study (Terova et al. 2021), which reported minor dietary effects of FM-TM substitution on species richness and diversity in rainbow trout, with similar trends in sea trout (*Salmo trutta*) (Mikołajczak et al. 2020), but contrasting ones in Siberian sturgeon (*Acipenser baerii*) (Józefiak et al. 2019). At the same time, however, in rainbow trout gut, technical, environmental, and host-associated factor have been shown to influence alpha and beta diversity, with technical factors (target hypervariable region and DNA extraction kit) affecting beta diversity, and environmental (diet) and host-associated (initial fish weight) factors affecting alpha diversity (Cao et al. 2024). Moreover, methodological variability in 16S rRNA gene sequencing may introduce bias, potentially altering microbial abundance estimates and affecting accurate community diversity assessment. Because these biases depend on context and environment, establishing standardized protocols in microbiome analysis remains challenging, thus limiting knowledge transferability across studies. Rigorous experimental design and methodological consistency within individual studies are hence essential to ensure the reliability and reproducibility of findings (Pollock et al. 2018).

When considering the validation of our model and, by extension, the feature subset identified on an external dataset, it is important to note that, although rainbow trout is a widely used model organism in aquaculture, research on the influence of yellow mealworm on its microbiota is relatively recent, with few studies examining the effects of different diet formulations on the rainbow trout microbiome. While similar studies have explored

the impact of insect meals on microbial community composition, the present study specifically focused on the development of insect-specific predictive models. This design choice enabled models to learn patterns unique to this specific diet; consequently, meaningful external validation would require datasets derived from rainbow trout reared under similar conditions and fed the same insect diet. Such an approach would thus ensure that the models learn and validate features attributable to TM-induced microbiota shifts, rather than being confounded by differences arising from alternative insect diets, which could lead to feature misattribution, reduced performance, or misleading conclusions. This consideration is particularly relevant given that microbiota profiles can vary substantially between insect species due to differences in their composition (e.g., chitin, fatty acids, and antimicrobial peptides). In future work, training on datasets encompassing multiple insect types could account for inter-insect variability, thereby enhancing model generalizability. At present, however, to the best of our knowledge, no datasets exist within our insect-specific framework that originate from fish reared under the same experimental conditions as those described in the original study.

Last but not least, it is important to acknowledge that the microbiology community has not yet defined consensus naming practices, resulting in significant inconsistencies in species identification. Indeed, the ongoing discovery of novel microbial diversity has outpaced the capacity of existing taxonomic frameworks established by international organizations, thereby contributing to persistent uncertainty that best practices adherence could mitigate (Hugenholtz et al. 2021; Oren 2024; Sanford et al. 2021). Such an imbalance between the rate of discovery and the pace of formal classification is particularly evident in taxonomic assignment pipelines, which are dependent on sequence databases that may not be regularly updated (Edgar 2017; Jeske and Gallert 2022). Our reference article (Terova et al. 2021) performed OTU assignment through Greengenes 13.8, which, despite representing a commonly used reference database for 16S rRNA gene-based classification, is no longer actively maintained. Many researchers have thus transitioned to alternative solutions such as the Ribosomal Database Project (RDP), which provides curated rRNA sequence data, user-friendly taxonomic classifiers, and robust tools for high-throughput analysis data, and SILVA, which offers an extensive database covering all life domains, with high-quality alignments and frequent updates enhancing the accuracy of taxonomic assignments. In recent years, a completely redesigned and actively maintained version, Greengenes2, has been released, integrating multiple sources in an effort to harmonize 16S rRNA gene and shotgun metagenomic datasets and thus enhancing its versatility when compared to its predecessor (McDonald et al. 2024).

The use of the original Greengenes platform carries important limitations (Ceccarani and Severgnini 2023; Myer et al. 2020), including the absence of newly described taxa, misannotations, and missing entries resulting from taxonomic reclassifications and renamings. Compared to more recent databases such as SILVA or RDP, moreover, Greengenes offers limited representation of environmental and host-associated microbes, often leading to low-resolution assignments or OTUs classified only at the domain or phylum level. However, even with modern databases, the pace of microbial discovery can easily surpass the capacity of curators to maintain fully up-to-date taxonomies. To address this issue, an alternative dual nomenclature system for prokaryotes is currently in place: the International Code of Nomenclature of Prokaryotes (ICNP), which continues to serve as primary reference, and the Code of Nomenclature of Prokaryotes Described from Sequence Data (SeqCode), which has been designed to accelerate the classification of uncultured taxa by relaxing the stringent requirements imposed by the ICNP for official recognition. Although this dual

approach represents a considerable step forward, its long-term effectiveness remains to be demonstrated, given the potential risk of generating multiple valid names for the same taxa (Hitch et al. 2024).

Conclusion

As aquafeed production progresses toward more environmentally sustainable solutions, insect meals have emerged as viable alternative protein sources necessitating thorough assessment of their effects on the fish microbiota. In this work, we applied machine learning to elucidate gut and skin microbiota response in rainbow trout following full substitution of fish meal with yellow mealworm. Despite the limited dietary modulation detected by conventional bioinformatic analyses, ML models showed that microbial composition could be consistently traced back to dietary regimens, while variations in feed ingredient proportions could moderately predict shifts in microbial abundance. The observed patterns were influenced by microbial feature subsets functioning as diet- and tissue-specific indicators across taxonomic levels, although additional validation is necessary to confirm their discriminative accuracy in real biological differentiation scenarios owing to the inherent independence of ML algorithms from biological assumptions and the potential overfitting of ML models to limited data (Wang et al. 2015). While computational and biological limitations remain, as the first study, to our knowledge, to explore microbial abundance and dietary regimens in rainbow trout microbiota at gut and skin level using ML approaches, our findings, albeit still exploratory, offer a proof-of-concept methodological framework for future research across fish species, diet formulations, and aquaculture systems, thus underscoring the complementary value of integrating AI-driven methodologies with conventional microbiological and bioinformatic analyses.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10499-025-02401-1.

Acknowledgements Giulio Saroglia and Violeta Kalemi are doctoral candidates in the Ph.D. program in Life Sciences and Biotechnology at the University of Insubria, Varese, Italy.

Author contributions Genciana Terova and Silvio Rizzi contributed to the study conceptualization. Giulio Saroglia, Simona Rimoldi, and Silvio Rizzi contributed to the methodological development. Giulio Saroglia, Simona Rimoldi, Silvio Rizzi, and Violeta Kalemi contributed to data curation. Genciana Terova and Silvio Rizzi wrote the original manuscript. Genciana Terova, Giulio Saroglia, Simona Rimoldi, Silvio Rizzi, and Violeta Kalemi reviewed and edited the final manuscript. Genciana Terova was responsible for funding acquisition. All authors read and approved the final manuscript.

Funding Open access funding provided by Università degli Studi dell'Insubria within the CRUI-CARE Agreement. This work has been funded by I-FISH. Protocol Number 414352 (07/12/2023). Area di Orientamento Occupazionale (AOO)—Fondo per la Crescita Sostenibile (FCS)—Accordi per l'Innovazione (D.M. 31/12/2021 and D.D. 14/11/2022).

Data availability All sequencing data used in this study were previously deposited as FASTQ files into the European Nucleotide Archive (ENA) database under accession number PRJEB38845. No new sequencing data were generated in this study.

Declarations

Competing interests The authors declare no competing interests.

References

Abellan-Schneyder I, Matchado MS, Reitmeier S, Sommer A, Sewald Z, Baumbach J, List M, Neuhaus K (2021) Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. mSphere 6:e01202-20. https://doi.org/10.1128/msphere.01202-20

Barbedo JGA (2022) A review on the use of computer vision and artificial intelligence for fish recognition, monitoring, and management. Fishes 7:335. https://doi.org/10.3390/fishes7060335

Berggren H, Tibblin P, Yıldırım Y, Broman E, Larsson P, Lundin D, Forsman A (2022) Fish skin microbiomes are highly variable among individuals and populations but not within individuals. Front Microbiol 12:767770. https://doi.org/10.3389/fmicb.2021.767770

Betiku OC, Yeoman CJ, Gaylord TG, Americus B, Olivo S, Duff GC, Sealey WM (2018) Water system is a controlling variable modulating bacterial diversity of gastrointestinal tract and performance in rainbow trout. PLoS ONE 13:e0195967. https://doi.org/10.1371/journal.pone.0195967

Biada I, Santacreu MA, González-Recio O, Ibáñez-Escriche N (2025) Comparative analysis of Illumina, PacBio, and nanopore for 16S rRNA gene sequencing of rabbit's gut microbiota. Front Microbiomes 4:1587712. https://doi.org/10.3389/frmbi.2025.1587712

Biazi V, Marques C (2023) Industry 4.0-based smart systems in aquaculture: a comprehensive review. Aquac Eng 103:102360. https://doi.org/10.1016/j.aquaeng.2023.102360

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu YX, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857. https://doi.org/10.1038/s41587-019-0209-9

Bruno A, Sandionigi A, Panio A, Rimoldi S, Orizio F, Agostinetto G, Hasan I, Gasco L, Terova G, Labra M (2023) Aquaculture ecosystem microbiome at the water-fish interface: the case-study of rainbow trout fed with *Tenebrio molitor* novel diets. BMC Microbiol 23:248. https://doi.org/10.1186/s12866-023-02990-y

Budaev S, Cusimano GM, Rønnestad I (2025) FishMet: a digital twin framework for appetite, feeding decisions and growth in salmonid fish. Aquac Fish Fish 5:e70064. https://doi.org/10.1002/aff2.70064

Buetas E, Jordán-López M, López-Roldán A, D'Auria G, Martínez-Priego L, De Marco G, Carda-Diéguez M, Mira A (2024) Full-length 16S rRNA gene sequencing by PacBio improves taxonomic resolution in human microbiome samples. BMC Genomics 25:310. https://doi.org/10.1186/s12864-024-10213-5

Cao S, Dicksved J, Lundh T, Vidakovic A, Norouzitallab P, Huyben D (2024) A meta-analysis revealing the technical, environmental, and host-associated factors that shape the gut microbiota of Atlantic salmon and rainbow trout. Rev Aquac 16:1603–1620. https://doi.org/10.1111/raq.12913

Carbajal-González MT, Fregeneda-Grandes JM, Suárez-Ramos S, Rodríguez Cadenas F, Aller-Gancedo JM (2011) Bacterial skin flora variation and in vitro inhibitory activity against *Saprolegnia parasitica* in brown and rainbow trout. Dis Aquat Organ 96:125–135. https://doi.org/10.3354/dao02391

Carda-Diéguez M, Mira A, Fouz B (2014) Pyrosequencing survey of intestinal microbiota diversity in cultured sea bass (*Dicentrarchus labrax*) fed functional diets. FEMS Microbiol Ecol 87:451–459. https://doi.org/10.1111/1574-6941.12236

Ceccarani C, Severgnini M (2023) A comparison between Greengenes, SILVA, RDP, and NCBI reference databases in four published microbiota datasets. Biorxiv. https://doi.org/10.1101/2023.04.12.535864

Cerezuela R, Fumanal M, Tapia-Paniagua ST, Meseguer J, Moriñigo MÁ, Esteban MÁ (2013) 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 34:1063–1070. https://doi.org/10.1016/j.fsi.2013.01.015

Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. Proc ACM SIGKDD Int Conf Knowl Discov Data Min 785–794. https://doi.org/10.1145/2939672.2939785

Chen MQ, Tian Y, Zhang C, Zhou JS, Tashi L, Wang WL, Pan H (2022) *Deefgea salmonis* sp. nov., isolated from gills of rainbow trout (*Oncorhynchus mykiss*). Arch Microbiol 204:366. https://doi.org/10.1007/s00203-022-02980-0

Chen X, Sun C, Dong J, Li W, Tian Y, Hu J, Ye X (2022) Comparative analysis of the gut microbiota of mandarin fish (*Siniperca chuatsi*) feeding on compound diets and live baits. Front Genet 13:797420. https://doi.org/10.3389/fgene.2022.797420

Coetzer WG, Coetzee LM, Cason ED, Grobler JP, Schneider SR, Boucher CE (2021) A preliminary assessment of skin microbiome diversity of zebrafish (*Danio rerio*): South African pet shop fish. Indian J Microbiol 61:81–84. https://doi.org/10.1007/s12088-020-00900-8

Colombo SM, Turchini GM (2021) 'Aquafeed 3.0': creating a more resilient aquaculture industry with a circular bioeconomy framework. Rev Aquacult. https://doi.org/10.1111/raq.12567

Cooney R, Wan AHL, O'Donncha F, Clifford E (2021) Designing environmentally efficient aquafeeds through the use of multicriteria decision support tools. Curr Opin Environ Sci Health 23:100276. https://doi.org/10.1016/j.coesh.2021.100276

D'Abramo LR (2021) Sustainable aquafeed and aquaculture production systems as impacted by challenges of global food security and climate change. J World Aquac Soc 52:1162–1167. https://doi.org/10.1111/jwas.12867

David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ (2014) Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559–563. https://doi.org/10.1038/nature12820

de Bruijn I, Liu Y, Wiegertjes GF, Raaijmakers JM (2018) Exploring fish microbial communities to mitigate emerging diseases in aquaculture. FEMS Microbiol Ecol 94:fix161. https://doi.org/10.1093/femsec/fix161

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072. https://doi.org/10.1128/AEM.03006-05

Drosdowech SM, Bezner S, Daisley B, Chiasson M, Rooney N, Huyben D (2025) Insect species and inclusion level influence gut microbiota composition and predicted metabolic function in rainbow trout (*Oncorhynchus mykiss*). Available at SSRN. https://doi.org/10.2139/ssrn.5388361

Duarte LN, Coelho FJRC, Cleary DFR, Bonifacio D, Martins P, Gomes NCM (2019) Bacterial and microeukaryotic plankton communities in a semi-intensive aquaculture system of sea bass (*Dicentrarchus labrax*): a seasonal survey. Aquaculture 503:59–69. https://doi.org/10.1016/j.aquaculture.2018.12.066

Edgar RC (2017) Accuracy of microbial community diversity estimated by closed- and open-reference OTUs. PeerJ 5:e3889. https://doi.org/10.7717/peerj.3889

El-Saadony MT, Alagawany M, Patra AK, Kar I, Tiwari R, Dawood MAO, Dhama K, Abdel-Latif HMR (2021) The functionality of probiotics in aquaculture: an overview. Fish Shellfish Immunol 117:36–52. https://doi.org/10.1016/j.fsi.2021.07.007

Ericsson AC, Busi SB, Davis DJ, Nabli H, Eckhoff DC, Dorfmeyer RA, Turner G, Oswalt PS, Crim MJ, Bryda EC (2021) Molecular and culture-based assessment of the microbiome in a zebrafish (*Danio rerio*) housing system during set-up and equilibration. Anim Microbiome 3:55. https://doi.org/10.1186/s42523-021-00116-1

Eroldoğan OT, Glencross B, Novoveska L, Gaudêncio SP, Rinkevich B, Varese GC, de Fátima CM, Tasdemir D, Safarik I, Nielsen SL, Rebours C, Lada LB, Robbens J, Strode E, Haznedaroğlu BZ, Kotta J, Evliyaoğlu E, Oliveira J, Girão M, Vasquez MI, Čabarkapa I, Rakita S, Klun K, Rotter A (2023) From the sea to aquafeed: a perspective overview. Rev Aquac 15:1028–1057. https://doi.org/10.1111/raq.12740

Garcia-Launay F, Dusart L, Espagnol S, Laisse-Redoux S, Gaudré D, Méda B, Wilfart A (2018) Multi-objective formulation is an effective method to reduce environmental impacts of livestock feeds. Br J Nutr 120:1298–1309. https://doi.org/10.1017/S0007114518002672

Gim DH, Lee SY, Han JE, Lee JY, Kang SM, Bae JW (2022) Description of *Deefgea piscis* sp. nov., and *Deefgea tanakiae* sp. nov., isolated from the gut of Korean indigenous fish. J Microbiol 60:1061–1069. https://doi.org/10.1007/s12275-022-2250-5

Gkinali AA, Matsakidou A, Vasileiou E, Paraskevopoulou A (2022) Potentiality of *Tenebrio molitor* larva-based ingredients for the food industry: a review. Trends Food Sci Technol 119:495–507. https://doi.org/10.1016/j.tifs.2021.11.024

Gladju J, Kamalam BS, Kanagaraj A (2022) Applications of data mining and machine learning framework in aquaculture and fisheries: a review. Smart Agric Technol 2:100061. https://doi.org/10.1016/j.atech.2022.100061

Glencross BD, Baily J, Berntssen MHG, Hardy R, MacKenzie S, Tocher DR (2020) Risk assessment of the use of alternative animal and plant raw material resources in aquaculture feeds. Rev Aquac 12:703–758. https://doi.org/10.1111/raq.12347

Gomez JA, Primm TP (2021) A slimy business: the future of fish skin microbiome studies. Microb Ecol 82:275–287. https://doi.org/10.1007/s00248-020-01648-w

Gordon-Rodriguez E, Quinn TP, Cunningham JP (2022) Data augmentation for compositional data: advancing predictive models of the microbiome. Adv Neural Inf Process Syst 35:20551–20565

Govindaraj K, Samayanpaulraj V, Narayanadoss V, Uthandakalaipandian R (2021) Isolation of lactic acid bacteria from intestine of freshwater fishes and elucidation of probiotic potential for aquaculture application. Probiotics Antimicrob Proteins 13:1598–1610. https://doi.org/10.1007/s12602-021-09811-6

Han H, Li Y, Qi Y, Mangiola S, Ling W (2025) Deciphering gut microbiome in colorectal cancer via robust learning methods. Genes 16:452. https://doi.org/10.3390/genes16040452

Hines IS, Marshall MA, Smith SA, Kuhn DD, Stevens AM (2023) Systematic literature review identifying bacterial constituents in the core intestinal microbiome of rainbow trout (*Oncorhynchus mykiss*). Aquacult Fish Fish 3:393–406. https://doi.org/10.1002/aff2.127

Hitch TCA, Wylensek D, Bisdorf K, Buhl EM, Treichel N, Abt B, Overmann J, Clavel T (2024) Harmonious naming across nomenclature codes exemplified by the description of bacterial isolates from the mammalian gut. Syst Appl Microbiol 47:126543. https://doi.org/10.1016/j.syapm.2024.126543

Hua K, Cobcroft JM, Cole A, Condon K, Jerry DR, Mangott A, Praeger C, Vucko MJ, Zeng C, Zenger K, Strugnell JM (2019) The future of aquatic protein: implications for protein sources in aquaculture diets. One Earth 1:316–329. https://doi.org/10.1016/j.oneear.2019.10.018

Huang M, Zhou YG, Yang XG, Gao QF, Chen YN, Ren YC, Dong SL (2025) Optimizing feeding frequencies in fish: a meta-analysis and machine learning approach. Aquaculture 595:741678. https://doi.org/10.1016/j.aquaculture.2024.741678

Hugenholtz P, Chuvochina M, Oren A, Parks DH, Soo RM (2021) Prokaryotic taxonomy and nomenclature in the age of big sequence data. ISME J 15:1879–1892. https://doi.org/10.1038/s41396-021-00941-x

Hunter JD (2007) Matplotlib: a 2D graphics environment. Comput Sci Eng 9:90–95. https://doi.org/10.1109/MCSE.2007.55

Iaconisi V, Bonelli A, Pupino R, Gai F, Parisi G (2018) Mealworm as dietary protein source for rainbow trout: body and fillet quality traits. Aquaculture 484:197–204

Ido A, Hashizume A, Ohta T, Takahashi T, Miura C, Miura T (2019) Replacement of fish meal by defatted yellow mealworm (*Tenebrio molitor*) larvae in diet improves growth performance and disease resistance in red seabream (*Pargus major*). Animals 9:100. https://doi.org/10.3390/ani9030100

Jankauskienė A, Aleknavičius D, Kiseliovienė S, Antanaitis Š, Falkauskas R, Šumskienė M, Juknienė I, Kabašinskienė A (2024) The influence of different sustainable substrates on the nutritional value of *Tenebrio molitor* larvae. Foods 13:365. https://doi.org/10.3390/foods13030365

Jarett JK, Kingsbury DD, Dahlhausen KE, Ganz HH (2021) Best practices for microbiome study design in companion animal research. Front Vet Sci 8:644836. https://doi.org/10.3389/fvets.2021.644836

Jeske JT, Gallert C (2022) Microbiome analysis via OTU and ASV-based pipelines - a comparative interpretation of ecological data in WWTP systems. Bioengineering 9:146. https://doi.org/10.3390/bioengineering9040146

Johnson AJ, Zheng JJ, Kang JW, Saboe A, Knights D, Zivkovic AM (2020) A guide to diet-microbiome study design. Front Nutr 7:79. https://doi.org/10.3389/fnut.2020.00079

Józefiak A, Nogales-Mérida S, Rawski M, Kierończyk B, Mazurkiewicz J (2019) Effects of insect diets on the gastrointestinal tract health and growth performance of Siberian sturgeon (*Acipenser baerii* Brandt, 1869). BMC Vet Res 15:348. https://doi.org/10.1186/s12917-019-2070-y

Jung A, Jung-Schroers V (2011) Detection of *Deefgea chitinilytica* in freshwater ornamental fish. Lett Appl Microbiol 52:497–500. https://doi.org/10.1111/j.1472-765x.2011.03030.x

Karwowska Z, Aasmets O, Estonian Biobank research team, Kosciolek T, Org E (2025) Effects of data transformation and model selection on feature importance in microbiome classification data. Microbiome 13:2. https://doi.org/10.1186/s40168-024-01996-6

Kirk D, Kok E, Tufano M, Tekinerdogan B, Feskens EJM, Camps G (2022) Machine learning in nutrition research. Adv Nutr 13:2573–2589. https://doi.org/10.1093/advances/nmac103

Larsen A, Tao Z, Bullard SA, Arias CR (2013) Diversity of the skin microbiota of fishes: evidence for host species specificity. FEMS Microbiol Ecol 85:483–494. https://doi.org/10.1111/1574-6941.12136

Larsen AM, Mohammed HH, Arias CR (2015) Comparison of DNA extraction protocols for the analysis of gut microbiota in fishes. FEMS Microbiol Lett 362:fnu031. https://doi.org/10.1093/femsle/fnu031

Lee BH, Nicolas P, Saticioglu IB, Fradet B, Bernardet JF, Rigaudeau D, Rochat T, Duchaud E (2023) Investigation of the genus *Flavobacterium* as a reservoir for fish-pathogenic bacterial species: the case of *Flavobacterium collinsii*. Appl Environ Microbiol 89:e0216222. https://doi.org/10.1128/aem.02162-22

Li P, Luo H, Ji B, Nielsen J (2022) Machine learning for data integration in human gut microbiome. Microb Cell Fact 21:241. https://doi.org/10.1186/s12934-022-01973-4

Lin H, Peddada SD (2020) Analysis of compositions of microbiomes with bias correction. Nat Commun 11:3514. https://doi.org/10.1038/s41467-020-17041-7

Liu Y, Wu H, Chen W, Liu C, Meng Q, Zhou Z (2022) Rumen microbiome and metabolome of high and low residual feed intake angus heifers. Front Vet Sci 9:812861. https://doi.org/10.3389/fvets.2022.812861

Lowrey L, Woodhams DC, Tacchi L, Salinas I (2015) Topographical mapping of the rainbow trout (*Oncorhynchus mykiss*) microbiome reveals a diverse bacterial community with antifungal properties in the skin. Appl Environ Microbiol 81:6915–6925. https://doi.org/10.1128/aem.01826-15

Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 30:4765–4774. https://doi.org/10.48550/arXiv.1705.07874

Marcos-Zambrano LJ, Karaduzovic-Hadziabdic K, Loncar Turukalo T, Przymus P, Trajkovik V, Aasmets O, Berland M, Gruca A, Hasic J, Hron K, Klammsteiner T, Kolev M, Lahti L, Lopes MB, Moreno V, Naskinova I, Org E, Paciência I, Papoutsoglou G, Shigdel R, Stres B, Vilne B, Yousef M, Zdravevski E, Tsamardinos I, de Carrillo Santa Pau E, Claesson MJ, Moreno-Indias I, Truu J (2021) Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment. Front Microbiol 12:634511. https://doi.org/10.3389/fmicb.2021.634511

McDonald D, Jiang Y, Balaban M, Cantrell K, Zhu Q, Gonzalez A, Morton JT, Nicolaou G, Parks DH, Karst SM, Albertsen M, Hugenholtz P, DeSantis T, Song SJ, Bartko A, Havulinna AS, Jousilahti P, Cheng S, Inouye M, Niiranen T, Jain M, Salomaa V, Lahti L, Mirarab S, Knight R (2024) Greengenes2 unifies microbial data in a single reference tree. Nat Biotechnol 42:715–718. https://doi.org/10.1038/s41587-023-01845-1

McKinney W (2010) Data structures for statistical computing in Python. Proc Python Sci Conf 9:51–56. https://doi.org/10.25080/Majora-92bf1922-00a

Mikołajczak Z, Rawski M, Mazurkiewicz J, Kierończyk B, Józefiak D (2020) The effect of hydrolyzed insect meals in sea trout fingerling (*Salmo trutta* m. *trutta*) diets on growth performance, microbiota and biochemical blood parameters. Animals 10:1031. https://doi.org/10.3390/ani10061031

Minich JJ, Poore GD, Jantawongsri K, Johnston C, Bowie K, Bowman J, Knight R, Nowak B, Allen EE (2020) Microbial ecology of Atlantic salmon (*Salmo salar*) hatcheries: impacts of the built environment on fish mucosal microbiota. Appl Environ Microbiol 86:e00411-20. https://doi.org/10.1128/aem.00411-20

Mwanza EP, Hugo A, Charimba G, Hugo CJ (2022) Pathogenic potential and control of *Chryseobacterium* species from clinical, fish, food and environmental sources. Microorganisms 10:895. https://doi.org/10.3390/microorganisms10050895

Myer PR, McDaneld TG, Kuehn LA, Dedonder KD, Apley MD, Capik SF, Lubbers BV, Harhay GP, Harhay DM, Keele JW, Henniger MT, Clemmons BA, Smith TPL (2020) Classification of 16S rRNA reads is improved using a niche-specific database constructed by near-full length sequencing. PLoS ONE 15:e0235498. https://doi.org/10.1371/journal.pone.0235498

Ng WK, Liew FL, Ang LP, Wong KW (2001) Potential of mealworm (*Tenebrio molitor*) as an alternative protein source in practical diets for African catfish, *Clarias gariepinus*. Aquac Res 32:273–280. https://doi.org/10.1046/j.1355-557x.2001.00024.x

Noyens I, Van Peer M, Goossens S, Ter Heide C, Van Miert S (2024) The nutritional quality of commercially bred yellow mealworm (*Tenebrio molitor*) compared to European Union nutrition claims. Insects 15:769. https://doi.org/10.3390/insects15100769

Olsen RL, Hasan MR (2012) A limited supply of fishmeal: impact on future increases in global aquaculture production. Trends Food Sci Technol 27:120–128. https://doi.org/10.1016/j.tifs.2012.06.003

Oren A (2024) On validly published names, correct names, and changes in the nomenclature of phyla and genera of prokaryotes: a guide for the perplexed. NPJ Biofilms Microbiomes 10:20. https://doi.org/10.1038/s41522-024-00494-9

Panteli N, Mastoraki M, Nikouli E, Lazarina M, Antonopoulou E, Kormas KA (2020) Imprinting statistically sound conclusions for gut microbiota in comparative animal studies: a case study with diet and teleost fishes. Comp Biochem Physiol Part D Genomics Proteomics 36:100738. https://doi.org/10.1016/j.cbd.2020.100738

Parks DH, Beiko RG (2010) Identifying biologically relevant differences between metagenomic communities. Bioinformatics 26:715–721. https://doi.org/10.1093/bioinformatics/btq041

Parks DH, Tyson GW, Hugenholtz P, Beiko RG (2014) STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics 30:3123–3124. https://doi.org/10.1093/bioinformatics/btu494

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Pollock J, Glendinning L, Wisedchanwet T, Watson M (2018) The madness of microbiome: attempting to find consensus "best practice" for 16S microbiome studies. Appl Environ Microbiol 84:e02627-17. https://doi.org/10.1128/aem.02627-17

Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2018) CatBoost: unbiased boosting with categorical features. Adv Neural Inf Process Syst 31:6638–6648. https://doi.org/10.48550/arXiv.1706.09516

Ravzanaadii N, Kim SH, Choi WH, Hong SJ, Kim NJ (2012) Nutritional value of mealworm, *Tenebrio molitor* as food source. Int J Indust Entomol Biomater 25:93–98. https://doi.org/10.7852/ijie.2012.25.1.093

Rema P, Saravanan S, Armenjon B, Motte C, Dias J (2019) Graded incorporation of defatted yellow mealworm (*Tenebrio molitor*) in rainbow trout (*Oncorhynchus mykiss*) diet improves growth performance and nutrient retention. Animals 9:187. https://doi.org/10.3390/ani9040187

Ringø E, Hoseinifar SH, Ghosh K, Doan HV, Beck BR, Song SK (2018) Lactic acid bacteria in finfish - an update. Front Microbiol 9:1818. https://doi.org/10.3389/fmicb.2018.01818

Rizzi S, Saroglia G, Kalemi V, Rimoldi S, Terova G (2025) Artificial intelligence in microbiome research and beyond: connecting human health, animal husbandry, and aquaculture. Appl Sci 15:9781. https://doi.org/10.3390/app15179781

Ruiz A, Torrecillas S, Kashinskaya E, Andree KB, Solovyev M, Gisbert E (2024) 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 11:1374803. https://doi.org/10.3389/fvets.2024.1374803

Saad A, Baikas A, Remen M, Bjørnson FO (2024) Optimizing feeding strategies in aquaculture using machine learning: ensuring sustainable and economically viable fish farming practices. Procedia Comput Sci 246:4712–4721. https://doi.org/10.1016/j.procs.2024.09.336

Sanford RA, Lloyd KG, Konstantinidis KT, Löffler FE (2021) Microbial taxonomy run amok. Trends Microbiol 29:394–404. https://doi.org/10.1016/j.tim.2020.12.010

Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C (2011) Metagenomic biomarker discovery and explanation. Genome Biol 12:R60. https://doi.org/10.1186/gb-2011-12-6-r60

Sharma M, Khurana H, Singh DN, Negi RK (2021) The genus *Sphingopyxis*: systematics, ecology, and bioremediation potential - a review. J Environ Manage 280:111744. https://doi.org/10.1016/j.jenvman.2020.111744

Sharma R, Barange M, Agostini V, Barros P, Gutierrez NL, Vasconcellos M, Fernandez Reguera D, Tiffay C, Levontin P (2025) Review of the state of world marine fishery resources–2025. FAO Fish Aquac Tech Pap 721. https://doi.org/10.4060/cd5538en

Soriano B, Hafez AI, Naya-Català F, Moroni F, Moldovan RA, Toxqui-Rodríguez S, Piazzon MC, Arnau V, Llorens C, Pérez-Sánchez J (2023) SAMBA: structure-learning of aquaculture microbiomes using a Bayesian approach. Genes 14:1650. https://doi.org/10.3390/genes14081650

Sylvain FÉ, Holland A, Bouslama S, Audet-Gilbert É, Lavoie C, Val AL, Derome N (2020) Fish skin and gut microbiomes show contrasting signatures of host species and habitat. Appl Environ Microbiol 86:e00789-20. https://doi.org/10.1128/aem.00789-20

Takeuchi M, Sugahara K (2025) Systematic literature review identifying core genera in the gut microbiome of rainbow trout (*Oncorhynchus mykiss*) and species-level microbial community analysis using long-read amplicon sequencing. Aquac Fish Fish 5:e70054. https://doi.org/10.1002/aff2.70054

Terova G, Gini E, Gasco L, Moroni F, Antonini M, Rimoldi S (2021) Effects of full replacement of dietary fishmeal with insect meal from *Tenebrio molitor* on rainbow trout gut and skin microbiota. J Anim Sci Biotechnol 12:1–14. https://doi.org/10.1186/s40104-021-00551-9

Turchini GM, Trushenski JT, Glencross BD (2019) Thoughts for the future of aquaculture nutrition: realigning perspectives to reflect contemporary issues related to judicious use of marine resources in aquafeeds. N Am J Aquac 81:13–39. https://doi.org/10.1002/naaq.10067

Turner JW Jr, Cheng X, Saferin N, Yeo JY, Yang T, Joe B (2022) Gut microbiota of wild fish as reporters of compromised aquatic environments sleuthed through machine learning. Physiol Genomics. https://doi.org/10.1152/physiolgenomics.00002.2022

Uren Webster TM, Rodriguez-Barreto D, Castaldo G, Gough P, Consuegra S, Garcia de Leaniz C (2020) Environmental plasticity and colonisation history in the Atlantic salmon microbiome: a translocation experiment. Mol Ecol 29:886–898. https://doi.org/10.1111/mec.15369

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17:261–272. https://doi.org/10.1038/s41592-019-0686-2

Wahli T, Madsen L (2018) Flavobacteria, a never ending threat for fish: a review. Curr Clin Microbiol Rep 5:26–37. https://doi.org/10.1007/s40588-018-0086-x

Wakita Y, Shimomura Y, Kitada Y, Yamamoto H, Ohashi Y, Matsumoto M (2018) Taxonomic classification for microbiome analysis, which correlates well with the metabolite milieu of the gut. BMC Microbiol 18:188. https://doi.org/10.1186/s12866-018-1311-8

Wang X, Su X, Cui X, Ning K (2015) MetaBoot: a machine learning framework of taxonomical biomarker discovery for different microbial communities based on metagenomic data. PeerJ 3:e993. https://doi.org/10.7717/peerj.993

Wang Y, Zhang H, Zhu L, Xu Y, Liu N, Sun X, Hu L, Huang H, Wei K, Zhu R (2018) Dynamic distribution of gut microbiota in goats at different ages and health states. Front Microbiol 9:2509. https://doi.org/10.3389/fmicb.2018.02509

Wang L, Jin L, Xue B, Wang Z, Peng Q (2019) Characterizing the bacterial community across the gastrointestinal tract of goats: composition and potential function. Microbiologyopen 8:e00820. https://doi.org/10.1002/mbo3.820

Waskom ML (2021) Seaborn: statistical data visualization. J Open Source Softw 6:3021. https://doi.org/10.21105/joss.03021

Wen LY, Zhang XM, Li QF, Min F (2023) KGA: integrating KPCA and GAN for microbial data augmentation. Int J Mach Learn & Cyber 14:1427–1444. https://doi.org/10.1007/s13042-022-01707-3

Wen LY, Chen Z, Xie XN, Min F (2024) Microbial data augmentation combining feature extraction and transformer network. Int J Mach Learn & Cyber 15:2539–2550. https://doi.org/10.1007/s13042-023-02047-6

Yang X, Zhang S, Liu J, Gao Q, Dong S, Zhou C (2021) Deep learning for smart fish farming: applications, opportunities and challenges. Rev Aquac 13:66–90. https://doi.org/10.1111/raq.12464

Zamora L, Vela AI, Palacios MA, Domínguez L, Fernández-Garayzábal JF (2012) First isolation and characterization of *Chryseobacterium shigense* from rainbow trout. BMC Vet Res 8:77. https://doi.org/10.1186/1746-6148-8-77

Zhang B, Xiao J, Liu H, Zhai D, Wang Y, Liu S, Xiong F, Xia M (2024) Vertical habitat preferences shape the fish gut microbiota in a shallow lake. Front Microbiol 15:1341303. https://doi.org/10.3389/fmicb.2024.1341303

Zhou Y, Xu ZZ, He Y, Yang Y, Liu L, Lin Q, Nie Y, Li M, Zhi F, Liu S, Amir A, González A, Tripathi A, Chen M, Wu GD, Knight R, Zhou H, Chen Y (2018) Gut microbiota offers universal biomarkers across ethnicity in inflammatory bowel disease diagnosis and infliximab response prediction. mSystems 3:10–1128. https://doi.org/10.1128/msystems.00188-17

Authors and Affiliations

Silvio Rizzi[1] · Giulio Saroglia[2] · Violeta Kalemi[1] · Simona Rimoldi[1] · Genciana Terova[1]

✉ Genciana Terova
genciana.terova@uninsubria.it

[1] Department of Biotechnology and Life Sciences, University of Insubria, Via Jean Henry Dunant 3, 21100 Varese, Italy

[2] Medical Devices 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