

Article

Prediction of Feed Efficiency and Performance-Based Traits in Fish via Integration of Multiple Omics and Clinical Covariates

Tim Young ^{1,2,*}, Olivier Laroche ³ , Seumas P. Walker ³, Matthew R. Miller ^{3,4} , Paula Casanovas ³, Konstanze Steiner ³, Noah Esmaili ⁴ , Ruixiang Zhao ⁴, John P. Bowman ⁵, Richard Wilson ⁶ , Andrew Bridle ⁴, Chris G. Carter ^{4,7} , Barbara F. Nowak ⁴ , Andrea C. Alfaro ¹ and Jane E. Symonds ^{3,4,*} 

- ¹ Aquaculture Biotechnology Research Group, Department of Environmental Science, School of Science, Private Bag 92006, Auckland 1142, New Zealand
 - ² The Centre for Biomedical and Chemical Sciences, School of Science, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand
 - ³ Cawthron Institute, Nelson 7010, New Zealand
 - ⁴ Institute for Marine and Antarctic Studies, University of Tasmania, Hobart Private Bag 49, Hobart 7005, Australia
 - ⁵ Tasmanian Institute of Agricultural Research, University of Tasmania, Hobart 7005, Australia
 - ⁶ Central Science Laboratory, Research Division, University of Tasmania, Hobart 7001, Australia
 - ⁷ Blue Economy Cooperative Research Centre, Launceston 7250, Australia
- * Correspondence: tim.young@outlook.co.nz (T.Y.); jane.symonds@cawthron.org.nz (J.E.S.)

Simple Summary: Aquaculture plays a key role in many emerging economies. Sources of fish for human consumption now exceed those from capture fisheries. However, the high cost of food to feed fish limits investment returns. Physiological inefficiencies in the way fish digest and assimilate nutrients and energy from food can be improved to lower the amount of feed required to farm fish whilst reducing waste and enhancing environmental sustainability. Being able to measure this ‘feed efficiency’ is crucial to develop strategies to improve it, yet we do not currently have a good means of doing this. With a focus on salmon, this research uses detailed physiological information involved in fish metabolic processes to better understand the mechanisms that can contribute towards better feed efficiency and growth, and at the same time enable development of models to accurately predict such measures. Our findings can be used to improve the efficiency of the salmon aquaculture industry and reduce the impact of farming practices on the environment.

Abstract: Fish aquaculture is a rapidly expanding global industry, set to support growing demands for sources of marine protein. Enhancing feed efficiency (FE) in farmed fish is required to reduce production costs and improve sector sustainability. Recognising that organisms are complex systems whose emerging phenotypes are the product of multiple interacting molecular processes, systems-based approaches are expected to deliver new biological insights into FE and growth performance. Here, we establish 14 diverse layers of multi-omics and clinical covariates to assess their capacities to predict FE and associated performance traits in a fish model (*Oncorhynchus tshawytscha*) and uncover the influential variables. Inter-omic relatedness between the different layers revealed several significant concordances, particularly between datasets originating from similar material/tissue and between blood indicators and some of the proteomic (liver), metabolomic (liver), and microbiomic layers. Single- and multi-layer random forest (RF) regression models showed that integration of all data layers provide greater FE prediction power than any single-layer model alone. Although FE was among the most challenging of the traits we attempted to predict, the mean accuracy of 40 different FE models in terms of root-mean square errors normalized to percentage was 30.4%, supporting RF as a feature selection tool and approach for complex trait prediction. Major contributions to the integrated FE models were derived from layers of proteomic and metabolomic data, with substantial influence also provided by the lipid composition layer. A correlation matrix of the top 27 variables in the models highlighted FE trait-associations with faecal bacteria (*Serratia* spp.), palmitic and nervonic acid moieties in whole body lipids, levels of free glycerol in muscle, and N-acetylglutamic acid



Citation: Young, T.; Laroche, O.; Walker, S.P.; Miller, M.R.; Casanovas, P.; Steiner, K.; Esmaili, N.; Zhao, R.; Bowman, J.P.; Wilson, R.; et al. Prediction of Feed Efficiency and Performance-Based Traits in Fish via Integration of Multiple Omics and Clinical Covariates. *Biology* **2023**, *12*, 1135. <https://doi.org/10.3390/biology12081135>

Academic Editor: Patricia Pereira

Received: 18 May 2023

Revised: 7 August 2023

Accepted: 8 August 2023

Published: 15 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

content in liver. In summary, we identified subsets of molecular characteristics for the assessment of commercially relevant performance-based metrics in farmed Chinook salmon.

Keywords: integrated multi-omics; metabolomics; proteomics; microbiomics; machine learning; random forest; feed efficiency; aquaculture

1. Introduction

Meeting global demands for sources of sustainable food protein is one of our greatest challenges. Aquaculture is uniquely poised to help resolve this for future populations. Creation of a thriving and viable ‘blue economy’ will greatly expand resource availability, capacity, and opportunity. However, increased seafood production will need to be complemented by reductions in environmental impact and improvements in resource efficiency [1,2]. Fish nutrition and feed development are key areas of focus to support sector advancement and drive aquaculture growth.

Feed efficiency (FE) is the efficiency with which food is utilised by animals for productive purposes such as growth. Commonly expressed as the Feed Conversion Ratio (FCR), this measure is the ratio between feed intake and body weight gain over a particular period. FE is the inverse of FCR, with lower FCR values indicating higher (enhanced) FE; from herein the term FE is primarily used in discussion for simplicity. In salmonid aquaculture, feed typically comprises 50–60% of total production costs, and excess provisions through uneaten and/or poorly digested food can cause adverse environmental impact [3,4]. FE is thus a highly important consideration for farming fish, and trait enhancement can improve sector profitability and sustainability. Yet, FE trait improvement is often overlooked by selective breeding programmes because it is difficult to measure in individual fish and the genetic underpinnings are not well understood [5,6].

FE is a complex trait which is influenced by numerous inextricably linked physiological processes, including social behaviour [7], feed intake [8], digestion capacity [9,10], nutrient absorption and assimilation [11,12], metabolism [13], and genetics [5,14,15]; processes which are further governed by a myriad of environmental factors [16,17]. Feed efficient fish tend to eat smaller meals whilst maintaining higher growth rates than inefficient conspecifics. We recently demonstrated in farmed Chinook salmon (*Oncorhynchus tshawytscha*) that fish displaying higher FE also have higher capacities for protein, lipid, and energy retention, and lower minimal routine metabolic rates [18]. These findings reveal that salmon with higher FE are also more efficient in their energy expenditure, requiring less energy for maintenance processes and presumably allowing more energy to be allocated to growth.

Providing further mechanistic insight into FE in farmed salmon, we have applied exploratory metabolomic- and proteomic-based approaches to screen for metabolite and protein biomarkers associated with divergent FE phenotypes. Using secondary bioinformatics techniques to reveal functional biochemical pathway information, findings highlighted that fish with high FE display greater capacity for protein synthesis in white muscle tissue, intestine, and liver, and increased energy production from elevated rates of lipid metabolism in the liver and brain [19–21]. Conversely, salmon with low FE displayed signatures indicative of protein degradation process (proteolysis and amino acid catabolism) being prominent in muscle, liver, and the intestine [21,22].

While these outcomes inform some physiological processes coupled to FE, additional information is believed obtainable through incorporation of broader biological features with opportunity to develop FE trait prediction capability. To this end, we have subsequently assessed a range of samples in Chinook salmon displaying variable FE phenotypes, including clinical covariates (i.e., multiple health indices, haematology parameters, and

known diagnostic blood biomarkers), crude chemical composition data, lipid composition, multi-sample proteomics, dual-technique metabolomics, and faecal microbiomics. Associations between FE and clinical biomarkers, the digestive microbiome, and body composition in salmon are currently unknown. Integration of these data form the basis of the current study.

Integration of multi-omics data is challenging due to their heterogeneity, yet is considered essential to decipher complex mechanisms underpinning phenotypic variations and gain holistic biological understanding [23]. Strategies to investigate cross-omic associations are developing rapidly and a wide array of methods are available [23–27]. Various machine learning principals and/or analyses of network architecture are typically applied under one of two primary frameworks—data-driven or knowledge-based integration—with method selection being guided by characteristics of the data at hand.

Knowledge-based approaches combine omics data from a functional perspective via pathway-based modelling [25–27]. Leveraging the collective wealth of genomic and biochemical knowledge in the public domain, genome-scale metabolic models and networks can be interrogated to link experimental data and extract unique information. Such approaches are suited to studies involving homogeneous sample types and sets of omics data with known functional relationships (e.g., transcriptomic—proteomic—metabolomic). Data-driven approaches on the other hand typically rely on machine learning algorithms to select relevant features from multi-omics datasets for interpretative and/or predictive purposes without a priori assumptions [23,28,29].

Our current data are particularly well-suited to a data-driven integration strategy due to collection of different sample types, inclusion of non-omics-based data, use of a high number of diverse datasets, and a lack of known biological associations among them. Multi-omics and clinical covariate data have different modalities and are by their nature meta-dimensional. These different data subsets, referred from now as ‘layers’, may each contribute information valuable for providing insight into, or predicting, phenotypic traits. However, challenges in the development of data-driven integration strategies are associated with overlapping predictive information within individual layers, differences in levels of predictive information between layers (which may also be dependent on the phenotypic traits under study), and requisite to consider existing interactions between variables across data layers [29]. To overcome these issues, Hornung and Wright [29] recently developed a prediction modelling approach for combining multi-omics and clinical covariates called ‘Random Block’; a variant of the random forest (RF) decision tree algorithm.

RF is an ensemble-based machine learning method able to capture complex dependency patterns between covariates and phenotypic traits [29]. RF is widely applicable to classification and regression problems, is robust to outliers, easily handles heterogenous data types (normalisation not required), is generally resilient to overfitting, does not rely on data distribution assumptions, can discern non-linear relationships, and computes feature importances in model predictions [30,31]. However, early use of RF in multi-omics integration relied on concatenated single matrix datasets and suffered from problems associated with unequal layer representation. The new Random Block variant addresses this integration problem by keeping the data layers separate and randomising the layer choice for each node split. When succession of the layers is randomised, predictions of the individual trees differ more widely. Furthermore, since predictive information contained within each layer overlap, the omission of particular layers at each split has a limited negative impact on the forecasting ability of each tree. Block-based RF has shown to out-perform numerous alternative data-driven multi-omics integration strategies for prediction modelling of phenotypic traits [29,32–34]. The algorithm’s versatility allows both the importance of individual features and the contributions of entire layers towards prediction accuracy to be assessed.

Herein, we utilise the powerful Random Block regression-based method to integrate 14 layers of covariate information obtained from the analysis of matched samples from individual fish displaying a continuum of FE. Our primary aim was to explore utility of leveraging multiple omics and clinical covariates for the prediction of FE phenotypes in farmed salmon. We further sought to extend application of this rich meta-dimensional data for the prediction of other performance-based traits related to feed intake and growth. Our specific objectives were to: (1) establish associations among the different data modalities, (2) construct a suite of trait prediction models and benchmark their accuracies, (3) evaluate relative contributions each information layer provides for predicting the different traits, and (4) gain new insights into mechanisms underpinning feed efficiency.

2. Materials and Methods

2.1. Experimental Design

Fish and their corresponding layers of multiple omics and clinical information were derived from a larger trial designed to establish the effects of water temperature and dietary ration on various performance metrics, such as feed efficiency (FE) and growth; see ‘Supplementary Methods S1’ for detailed descriptions of the fish, experimental design, and rearing conditions. Briefly, an all-female cohort of juvenile Chinook salmon were obtained from a commercial hatchery (Mt Cook Alpine Salmon; Canterbury, New Zealand), individually tagged, and reared for eight months at the Cawthron Aquaculture Park Finfish Research Centre (Nelson, New Zealand) in nine 8000 L tanks (176–187 fish per tank) on a freshwater recirculating aquaculture system at 17 °C. To promote FE variation, tanks were assigned one of three feed rations (60%, 80%, and 100% (satiation)), with three replicate tanks per ration; the 60 and 80% rations were calculated daily based on the satiation ration adjusted for biomass differences.

During the experimental period, fish were assessed four times and individual FE was estimated in three sequential phases. On completion of the trial, salmon from each tank were selected based on their overall health status and consistencies of their FE trait ranks across the FE assessments. Various samples from these fish were assigned to a broad range of evaluations involving omics (i.e., microbiome, proteome, metabolome, proximate and lipid composition) and/or assessments with clinical relevance (i.e., macroscopic observations, blood chemistry parameters, haematology). We acquired matched information for 28 fish across all samples and layers of multiple omics and clinical covariates (Figure 1).

2.2. Fish Handling and Sampling

Fish were handled in accordance with protocols approved by the Animal Ethics Committee of the Nelson Marlborough Institute of Technology (Ref: AEC2018 CAW01). During phenotypic trait measurements, fish were either sedated with tricaine methanesulfonate (mid-trial assessments) or euthanised first by anaesthetic overdose with AQUI-S® (end-of trial assessments). Immediately following euthanasia, blood samples were withdrawn for haematological and plasma-based biochemical analyses, anatomical health assessments were made, and dissections were performed to obtain liver, muscle, and digesta/faecal samples.

2.3. Phenotypic Traits

Performance-based traits were assessed following [19], yielding four measures of growth (i.e., fork length, body weight, condition factor, specific growth rate), three measures of feed intake (i.e., mean daily feed intake (DFI), mean daily feed intake per bodyweight (DFI BW⁻¹), the mean proportion of the feed intake (PFI) of an individual fish based on its share of the meal within a tank), and three measures of feed efficiency (i.e., feed conversion ratio (FCR)).

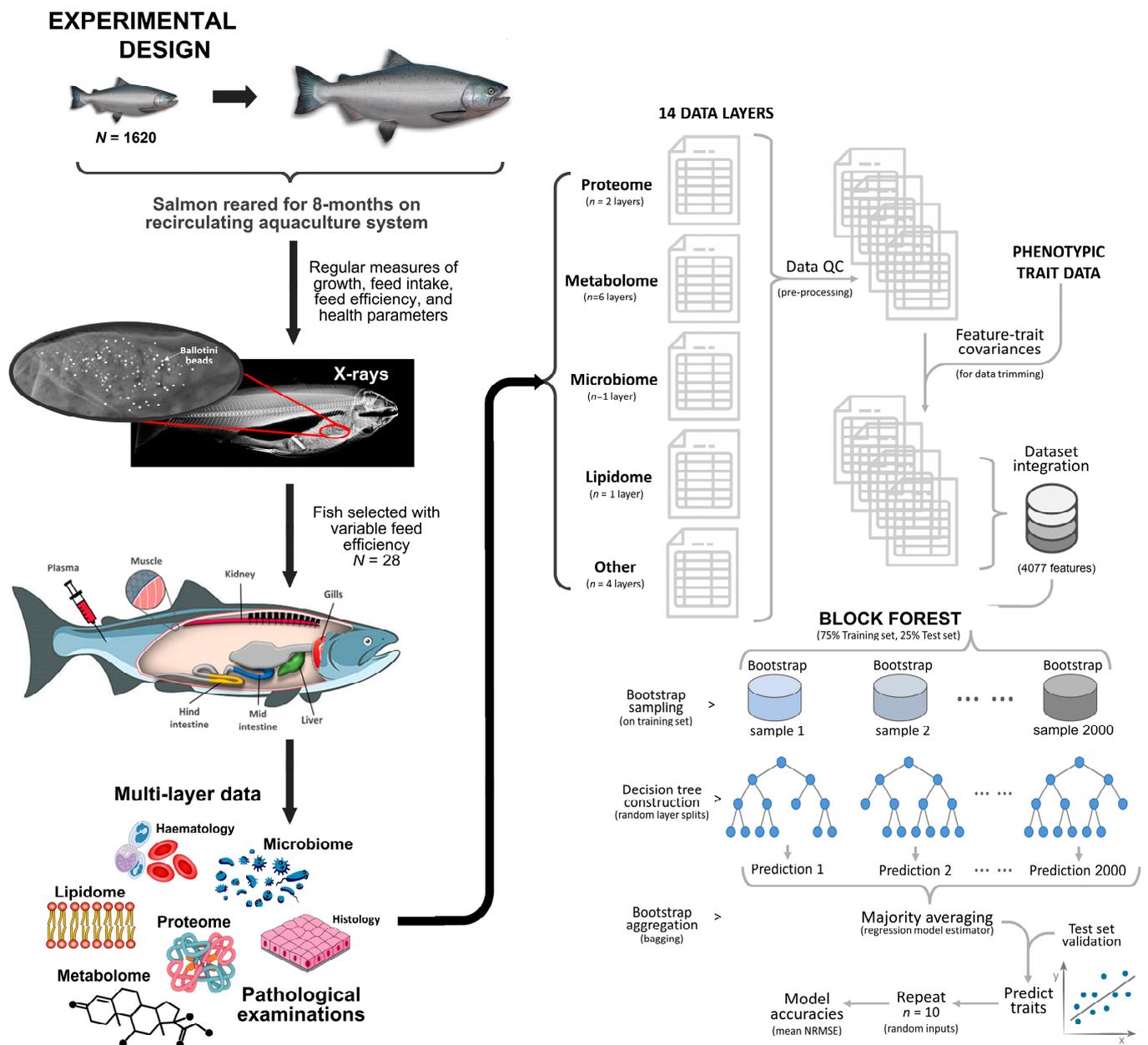


Figure 1. Schematic overview of the experimental design and data analysis workflow.

2.3.1. Growth

Fork length (FL (mm)) and body weight (BW (g)) were measured on four occasions during the trial, with Fulton’s K condition factors (CF) being calculated from these values after each assessment:

$$CF = 100 \times \frac{BW}{FL^3}$$

End-of-trial FL and BW values were used as dependent variable traits in prediction models, whereas for CF the mean of all four measures were used. Specific growth rates (SGR (% mass per day)) between assessments were calculated as:

$$SGR = \frac{\ln BW \text{ gain}}{N \text{ Days}} \times 100$$

The mean of the three SGR measures were used as model responses.

2.3.2. Feed Intake

Tank-level feed intake (tFI (g)) was measured daily through quantification of uneaten feed and subtraction from the amount originally provided. Individual feed intake (iFI (g)) assessments were performed using X-radiography, according to [35]. Briefly, on four occasions during the trial, fish were fed customised extruded pellets containing a known concentration of X-ray opaque beads. Fish were X-rayed following these feeding events and numbers of beads within the gastrointestinal tract were counted to provide accurate iFI appraisals for each fish. Mean DFI and mean DFI BW⁻¹ were subsequently calculated as:

$$\text{mean DFI} = \frac{\sum \text{iFI measures}}{4}$$

$$\text{mean DFI BW}^{-1} = \frac{\sum [\text{iFI}_{(i..j)} / \text{BW}_{(i..j)}]}{4}$$

Measured percentage 'share of the meal' (SOM (%)) at each assessment was calculated for each fish from the iFI and tFI values by dividing the individual feed intake by the total feed intake for the tank based on X-radiography; the mean proportion of the feed intake was then calculated as the mean % SOM:

$$\% \text{ SOM} = \frac{\text{iFI}}{\text{tFI}} \times 100$$

$$\text{mean PFI} = \frac{\sum \% \text{ SOMs}}{4}$$

2.3.3. Feed Efficiency

Feed conversion ratio (FCR) was used as the measure of individual FE, where FCR is the ratio between the mass of food consumed and the BW gain over a particular period. FCR was calculated using two different methods. The first method (FCR1) used the total food eaten by each tank (accounting for the feed recovered), the mean percent share of the meal of each fish calculated using the daily feed intake, and the total weight gain of the fish between assessment periods.

First, we calculated the percent share of the meal for each fish (% SOM) as previously described. The mean percentage of the meal eaten for each fish was then determined based on the assessments at the start and end of each period, for example for Assessment 1 (A1) to Assessment 2 (A2):

$$\text{mean \% SOM}_{(A1 \text{ to } A2)} = \frac{\% \text{ SOM}_{(A1)} + \% \text{ SOM}_{(A2)}}{2}$$

To estimate the total amount of feed eaten by an individual fish from A1 to A2, the total amount of feed consumed by the tank over this period (calculated as the total feed given to the tank minus the total feed recovered (i.e., \sum tFI values A1 to A2)) was multiplied by the mean % SOM for that period, for each individual fish:

$$\text{Total feed intake} = \sum \text{tFI}_{(A1 \text{ to } A2)} \times \text{mean \% SOM}_{(A1 \text{ to } A2)}$$

Individual feed conversion ratio (FCR1) was then calculated as the total feed eaten by an individual (A1 to A2) based on the mean % SOM divided by the total weight gain (g) of the individual (A1 to A2):

$$\text{FCR1} = \frac{\text{Total feed intake}}{\text{Total weight gain}}$$

The second method (FCR2) used the mean daily feed intake estimated at the time of the two assessments and the daily weight gain during that period. Individual FCR was calculated as the mean daily feed intake by an individual (for A1 and A2) divided by the

daily weight gain of the individual (weight difference between A1 and A2, divided by the number of days between assessments).

$$\text{FCR2} = \frac{\text{mean DFI}_{(A1 \text{ and } A2)}}{\text{Daily weight gain}}$$

‘Mean FCR1’ and ‘Mean FCR2’ are the means of the three FCR values obtained for each assessment period using the FCR1 and FCR2 methods, respectively. ‘Last FCR1’ and ‘Last FCR2’ are the FCR values for the last (third) measurement period using the FCR1 and FCR2 methods, respectively. These response variables were selected to represent long-term FE phenotypes (i.e., the Mean FCRs) and a snapshot of those reflective of the FE phenotype at the time of sampling for tissues and biofluids (i.e., the Last FCRs).

2.4. Model Covariates

Tissues and biofluids were subjected to a range of analyses to generate the model covariates for this study. Across all samples and measurements, a total of 4077 features were retained for inclusion as explanatory variables in the Random Block models (Table 1). Complete lists of all model covariates are provided in Supplementary Data.

Table 1. Summary of explanatory variables used in this study for the prediction of feed efficiency and performance-based traits in fish (ASVs = amplicon sequence variants; F = faeces; L = liver; M = white muscle; P = plasma; WB = whole body homogenate; Fi = fillet; B = blood).

	Covariate Class	Feature Type	Sample Type	N Layers	N Features
Non-clinical	Microbiome	Microbial ASVs	F	1	135
	Proteome	Proteins	L, M	2	3223
	Metabolome	Metabolites	L, M P	6	556
	General composition	Macromolecules	WB, Fi	1	28
	Lipid composition	Fatty acids	WB	1	32
Clinical	Blood biomarkers	Proteins and Metabolites	B	1	24
	Haematology indices	Blood cell profiles	B	1	12
	Health indices	Organs	O	1	5
	Total			14	4077

2.4.1. Microbiome Layer

Microbial profiling of faecal samples was performed as described by [36]; refer to ‘Supplementary Methods S2’ for a detailed description. Briefly, bacterial DNA was extracted using a ‘NucleoSpin Soil’ kit according to the manufacturers’ instructions. Polymerase chain reaction was performed using the bacterial universal primers 27F and 519R. To evaluate the bacterial DNA, 2 × 300 bp pair-ended amplicon sequencing of the V1–V3 region of the 16S rRNA gene was performed using a MiSeq Illumina platform.

Data were trimmed, joined, and demultiplexed, then filtered for low quality reads. Taxonomic assignments were made using the ‘Seed 2’ pipeline, according to [37]. Following sequence alignment, denoising, chimera checking, and clustering, operational taxonomic units were optimised and defined then classified against the Silva non-redundant 16S rRNA database (www.arb-silva.de/, accessed on 16 December 2019). Microbial data were pre-processed by filtering out rare amplicon sequence variants (ASVs; those with less than two reads in at least two samples, leaving a total of 135 ASVs) and by transforming their read counts to centred-log ratios following recommendations from [38] when dealing with compositional data.

2.4.2. Proteome Layers

Samples of liver and muscle were processed for proteomic profiling as described by Esmaeili et al. [19]; refer to ‘Supplementary Methods S3’ for a detailed description. Briefly, tissues were homogenised in denaturation buffer and protease inhibitor cocktail. Crude protein was reduced and alkylated via standard techniques, then digested using the ‘SP3’

method according to [39]. Following acidification and de-salting, digests were analysed on a liquid chromatography mass spectrometry (LC-MS) system (LC: Thermo Scientific™ Ultimate 3000 RSLCnano) equipped with a reverse phase (RP) column, nanospray ion source, and a hybrid quadrupole-orbitrap mass spectrometer (MS: Thermo Scientific™ Q-Exactive HF). Proteomic analysis deployed a combination of data-dependent acquisition (DDA) and data-independent acquisition (DIA) mass spectrometry approaches.

Spectral libraries for the tissue peptides were generated by off-line micro fractionation of pooled liver and pooled muscle digests into 16 concatenated fractions using microflow high-pH RP high performance liquid chromatography (HPLC), followed by DDA-MS. Individual samples were then analysed by DIA-MS using the same HPLC gradients and extracted median MS2 ion intensities were used for label-free quantitation of protein abundance between samples. DDA- and DIA-MS raw files were processed using Spectronaut software. Each tissue-specific library was generated using the Pulsar search engine to search DDA MS2 spectra against *Oncorhynchus tshawytscha* protein sequences in the National Center for Biotechnology Information (NCBI) database. Spectronaut protein quantitation pivot reports, including protein description, gene names, and NCBI accession numbers, from the 28 fish were uploaded into Perseus software for further data processing. Quantitative values were log₂ transformed, and proteins filtered according to the number of valid values (N = 3223 variables). Batch effects in the data were removed via an empirical Bayes framework, the 'ComBat' method [40], using MetaboAnalyst v5.0 software [41].

2.4.3. Metabolome Layers

Plasma, liver, and muscle samples were extracted and derivatised according to established protocols [42,43], with minor modification; refer to 'Supplementary Methods S4' for a detailed description. Briefly, metabolites were co-extracted with internal standards (d4-alanine and ribitol) using methanol as a solvent. Sub-aliquots of each extract were simultaneously derivatised using methyl chloroformate (MCF) and trimethyl silylation (TMS) to target organic acid and carbohydrate functional group chemistries, respectively—thus providing two metabolomics datasets for each sample type (i.e., six metabolomics datasets in total). Derivatised samples were analysed according to [44] on a gas chromatography mass spectrometry (GC-MS) system (GC: Agilent 7890B) equipped with a mid-polarity column, an electron impact ion source, and a quadrupole analyser (MS: Agilent 5977B). Spectral data were deconvoluted using Automated Mass Spectral Deconvolution and Identification System (AMDIS) software [45].

In-house libraries of pure standards and the mass spectral library of the National Institute of Standards and Technology (NIST) were used to identify metabolites where possible; unidentified features were retained as variables within the metabolomics data layers. Pooled quality control normalisation was performed using an RF-based algorithm (Systematic Error Removal using Random Forest [46]) to remove systematic errors associated with post-extraction processes (e.g., batch effects); internal standard-based normalisation was subsequently applied to account for error sources associated with the extractions (e.g., variable metabolite recoveries). Aberrant records were filtered via blank subtraction and manual screening. Lastly, data were adjusted to sample-specific biomass where appropriate, missing values were estimated via the 'half minimum' approach, and resultant metabolomics entries (N = 556 variables) were autoscaled.

2.4.4. General Composition Layer

Selected elements, crude chemical composition, and some fatty acid class information in whole-body (WB) homogenates and/or the fillet portion of the fish were measured using established methods [47]; refer to 'Supplementary Methods S5' for detailed descriptions. Briefly, levels of calcium and phosphorus were measured in whole-body (WB) homogenates via inductively coupled plasma mass spectrometry (ICP-MS). WB crude chemical composition was assessed using validated procedures of the Association of Official Analytical Chemists.

Briefly, values were obtained for crude protein via total Kjeldahl nitrogen (AOAC 981.10), total fat via acid hydrolysis (AOAC 948.15), moisture via gravimetry at 105 °C (AOAC 950.46), and ash via incineration (AOAC 920.153); total carbohydrate content was estimated by difference (i.e., 100 – % protein – % fat – % moisture – % ash). These wet chemistry data were supplemented with WB proximate composition data estimated using near infrared (NIR) spectroscopy and species-specific NIR models we previously validated [48]. NIR models were also employed to obtain proximate composition and information on selected lipid-based characteristics (e.g., fatty acid classes) in fish fillets, extending the ‘general composition’ data layer (N = 28 variables).

2.4.5. Lipid Composition Layer

Lipids in WB homogenates were extracted via a cold-solvent-modified Folch extraction [49], and their fatty acid profiles (fatty acid methyl esters [FAMES]; N = 32 variables) were analysed by gas chromatography mass spectrometry (GC-MS); refer to ‘Supplementary Methods S6’ for a detailed description. Briefly, samples were homogenised in a chloroform/methanol/water solvent mix to partition the lipid component into the organic phase. Following removal and concentration of the organic phase, the crude lipid extracts were methylated using a Gerstel MPS (Multipurpose sampler) via saponification with methanolic sodium hydroxide. FAME samples were analysed according to AOAC official methods 963.22.

2.4.6. Blood Biomarker Layer

A targeted suite of clinically relevant blood biomarkers was profiled according to [50]; refer to ‘Supplementary Methods S7’ for a detailed description. Briefly, snap-frozen plasma samples were sent frozen on dry ice to an accredited commercial laboratory (Gribbles Veterinary (GV); Christchurch, New Zealand) for analysis of blood chemistry indices (N = 24 variables), comprising various proteins, metabolites, ions, and their ratios. Analytes were quantified using automated chemistry analysers with assay kits developed by the manufacturer, or via enzyme-linked immunosorbent assays.

2.4.7. Haematology Layer

Haematology indices (N = 12 variables) were evaluated according to [51]; refer to ‘Supplementary Methods S8’ for a detailed description. Briefly, haematocrit values were recorded immediately after blood withdrawal. Differential cell counts were obtained using two whole blood smears per fish. Air-dried slides were stained (Leishman) and cells were enumerated under the microscope at the GV laboratory. Haemoglobin content was determined on an automated analyser.

2.4.8. Health Layer

Selected health-relevant indices (N = 5 variables) were integrated as potential trait predictors, comprising: visceral fat score (VFS) according to [51], faecal appearance score (FAS) according to [52], and somatic indices for the heart, liver, and gonad (i.e., organ weight/BW × 100).

2.5. Data Analyses

2.5.1. Concordance Analysis

Pair-wise associations among the different multivariate data modalities were examined via Procrustes analyses and followed by procrustean randomisation tests (PROTEST [53]) to test significance of the goodness-of-fit of correspondence statistics. Procrustes-based concordance analysis essentially computes reduced dimensions (e.g., multidimensional scaling (MDS) configurations) for two multivariate datasets and transforms one of the matrices until it has maximum similarity with the other when superimposed. This is achieved by minimising the sum of the squared residual deviations between points for

each observation and the identical observation in the target matrix. Procrustes is widely used for measuring global associations between multi-omics datasets.

Each dataset was standardised and the non-metric dimensional scaling matrices calculated on the Euclidean distance: `monoMDS(vegdist(decostand(x = dataset, method = "standardize"), "euclidean"))`, followed by PROTEST (scores = "sites", permutations = `how(nperm = 999)m` scale = T, symmetric = T) using the 'vegan' R package. Briefly, PROTEST assesses the correspondence between matching points of two datasets using least-squares orthogonal mapping and tests for concordance significance using a permutational approach. Results were visualised with a heatmap using the 'ggplot2' R package (v3.3.5) [54].

2.5.2. Random Block-Based Prediction of Phenotypes

The random forest variant Random Block differs from its predecessor simply with respect to the selection of the split points in the decision trees which make up the forests. Rather than concatenate the data into a single matrix for integration, Random Block keeps the layers separate as inputs and integrates them during tree construction. Layers were separated by covariate class and by sample type to ensure that overlapping features (e.g., the same protein being detected in muscle and liver) were given simultaneous opportunity for inclusion in the models. At the split points, a block (layer) of data is randomly selected and a subset of features contained within that block is then considered for splitting [29]. Split point selection is subsequently performed as in standard random forest using the sampled subset of features. Randomising succession of the blocks used for splitting promotes the rendering of feature interactions across blocks.

Prior to Random Block prediction, each dataset was trimmed to keep only the most relevant variables associated with the phenotype of interest using the minimal unbiased variable selection (MUVR) algorithm [55]. MUVR uses machine learning to identify minimal-optimal and all-relevant variable sets for regression and classification problems. The algorithm minimises overfitting and false positives in multivariate analysis and improves predictive performance. The MUVR package (v0.0.973) was implemented in R with the following settings: `nRep = 35`, `nOuter = 8`, `varRatio = 0.9`, `method = "RF"`, and `model = "mid"`.

Trimmed datasets were then used as input (with each dataset representing a block) to the Random Block algorithm (v0.2.4) [29] to create predictive models (`block.method = "RandomBlock"`, `num.trees = 2000`, `nsets = 300`, `num.trees.pre = 1500`, `splitrule = "extratrees"`, `importance = "impurity"`). Prioritisation of particular layers were not applied during tree construction since relationships between different covariates and the responses were unknown a priori.

For each phenotype, ten iterations of the process were executed using 75% of the data to create the models (`seeds = 30–39`), and the remainder to test their performance. This process allowed us to assess the stability of the models and identify potential issues that could affect the results. Prediction accuracies were evaluated and benchmarked by the mean normalised root-mean square errors (NRMSE), also called a scatter index. The root-mean square error (RMSE) is basically a measure of the differences between predicted and observed values and normalising the RMSE enables comparison between datasets of different scales. The lower the NRMSE value, the lesser the deviation of the predicted values from the true response. Here, the RMSE was normalised by the mean and transformed to a percentage following the formula below, where y represents observed values:

$$\text{NRMSE} = \frac{\text{RMSE}}{\bar{y}} \times 100$$

The mean contribution of each layer (block) for phenotype prediction (based on block/layer selection probabilities) was visualised with stacked bar plots using the 'ggplot2' R package. To evaluate the general performance of models based on the integration of all data versus individual datasets, Random Block models were also created per dataset and

the performance of each ranked per phenotype and summed across all traits. Results were then displayed as bar plots using the ‘ggplot2’ R package.

2.5.3. Network Analysis and Inter-Layer Associations

Focusing specifically now on FE as the trait of interest, variables selected by the best models for the FE predictions were examined and compared between FCR methodologies (i.e., mean and last FCR1 and FCR2) using an UpSet figure and the ‘UpSetR’ (v1.4.0) R package [56]. Variables consistently found to be important in predicting the FE sub-phenotypes were further investigated by creating a correlation matrix and a co-occurrence network. First, all variables of interest, including the FCR phenotypes, were merged into a single dataframe and normalised. Secondly, Pearson correlations were performed between each variable using the rcorr function of the ‘Hmisc’ (v4.5.0) R package [57]; a heatmap matrix of these values were plotted. Finally, highly significant relationships (p -value ≤ 0.001) were displayed in a co-occurrence network using the ‘ggraph’ (v2.0.5) R package [58].

3. Results and Discussion

3.1. Phenotypic Traits

The fish phenotypic traits used in this study as Random Block model response variables are displayed as density plots in Figure 2, highlighting their distributions. Traits were generally unimodal, although a relatively small proportion of fish had poor FE values, which were also associated with lower SGR and CF values.

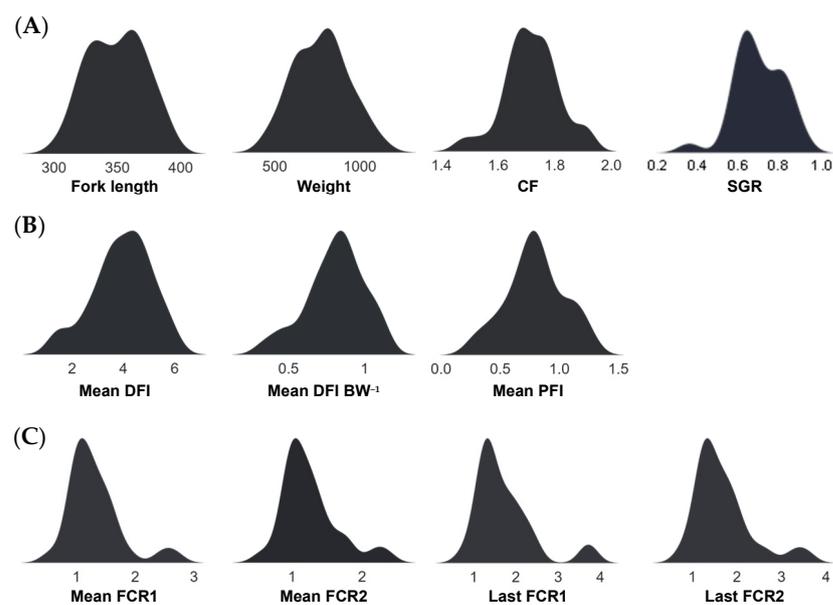


Figure 2. Performance-based trait distributions of salmon (*O. tshawytscha*) used as response variables in the Random Block models, comprising measures of growth (A), feed intake (B), and feed efficiency (C).

3.2. Covariate Layer Associations

To investigate inherent cross-layer association, an initial survey of the multi-omic and clinical covariate data was performed without considering the phenotypic traits of fish. Global covariate associations between the different layers were assessed using Procrustes analysis, which aims to find the best possible correspondence or correlation between two sets of multivariate matrices. Procrustes analysis uses ordination methods to compare high-dimensional datasets by optimally aligning their lower-dimensional projections. Dissimilarity between configurations is iteratively minimised based on a distance metric and the sum of square differences between corresponding points.

Significant concordance between many of the layers were detected (Figure 3; bold values only). Highest concordances were observed between layers of similar omics approach (i.e., metabolomics plasma TMS vs. metabolomics muscle MCF data ($r = 0.49$)) and between layers originating from the same sample (i.e., metabolomic plasma TMS vs. metabolomic plasma MCF [$r = 0.43$] and metabolomics muscle MCF vs. muscle proteomics), suggesting associations between these different levels of biological organisation within sample types. In addition, strong and significant correlation between blood biomarkers and health ($r = 0.44$), and between the former and a few of the omics datasets (i.e., proteomics liver ($r = 0.44$), metabolomics liver MCF ($r = 0.39$), and microbiomics ($r = 0.41$)) were observed.

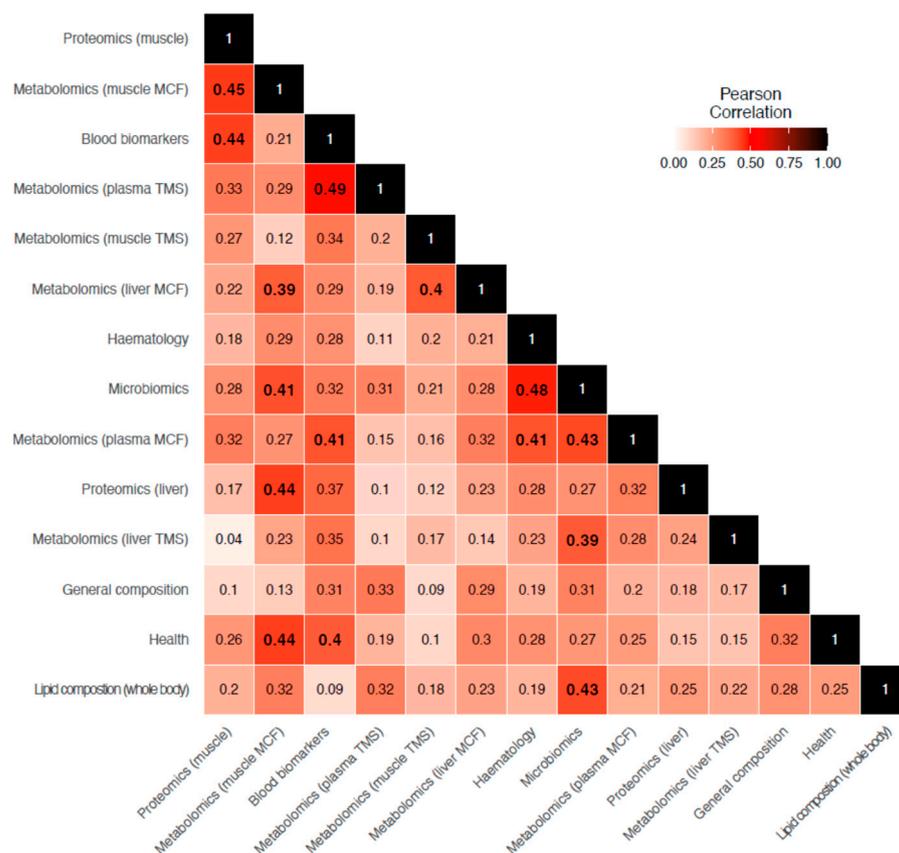


Figure 3. Heatmap of Pearson correlation between datasets derived from PROTEST results. Significant ($p < 0.05$) associations are highlighted in bold.

Our study highlights the utility of concordance-based analysis to examine relationships between multi-omics datasets, extending its use beyond some existing microbiome-metabolome exemplars in human models [59–61]. Omics-based investigations of fish health is a rapidly evolving area of research important for the development of strategies to confer disease resistance and produce therapeutic interventions [62–65]. Unravelling cross-omic associations which underly various health conditions in fish is a challenging but crucial task for mapping molecular networks.

3.3. Integrated Random Block Prediction Accuracies

Using the integrated multiple omics and clinical covariates, a suite of Random Block trait prediction models was constructed to benchmark their performance accuracies and reveal scope for potential biomarker identification and interpretation (Figure 4). The most predictable phenotypes included fork length, condition factor and SGR with average mean normalised root mean squared errors (NRMSE) of 5.3%, 5.5%, and 11.5%, respectively. Although practical utilities of these growth models are considered limited since the traits themselves are easy to measure directly, the identities of important covariates within the

models may expose underpinning mechanistic foundations. FE measures (our primary trait of interest) were among the most difficult to predict with models showing moderate–high variance in performance (NRMSE's 11.4–54.9%). However, 23 out of 40 of these model iterations had NRMSE's < 30%, which provides an encouraging outlook for discovering putative covariate–trait associations.

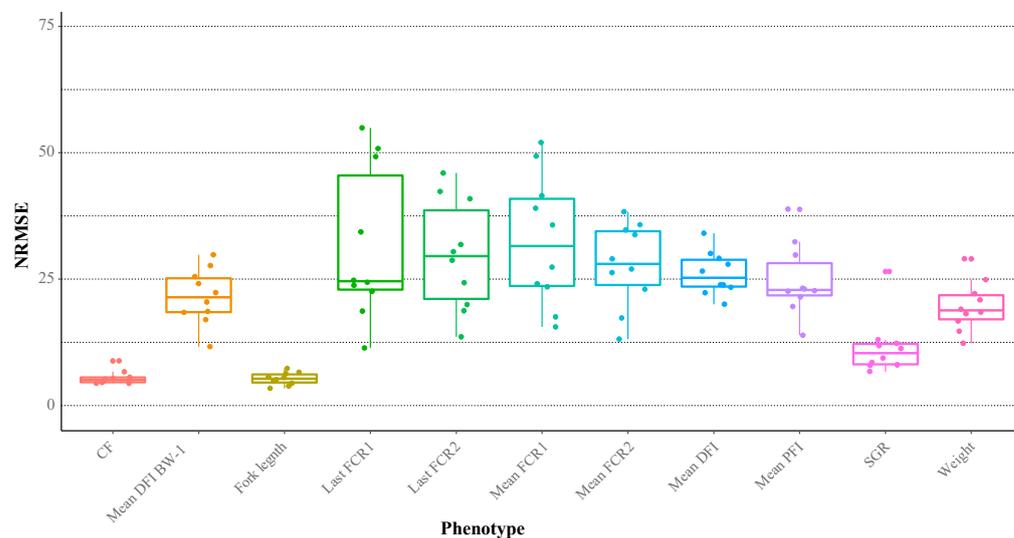


Figure 4. Boxplots of the normalised root-mean square error (NRMSE) in the Random Block phenotype prediction models based on all layers. Data points correspond to the result of each model iteration.

3.4. Covariate Layer Contributions

During construction of the integrated Random Block models, specific selection probabilities are calculated for each of the different layers. Using these values, the average relative contribution each information layer provides for predicting the phenotypic traits in the Random Block models were assessed (Figure 5). Overall, the datasets which contributed most to trait prediction were the liver and muscle proteomics data, as well as the muscle and plasma MCF and TMS metabolomics data. This aligns with the importance of metabolic processes underlying growth performance variation. Whole body lipid composition (fatty acid (FAME) data) played a greater role in predicting FE phenotypes compared to other performance metrics in general, and faecal microbial composition played a larger role in model performances in two of the four FE measures. Associations between fatty acid profiles and FE have previously been found in cattle [66], yet such relationships in fish are currently unknown [19].

Global performance of the integrated datasets with combined modalities versus individual dataset in predicting phenotypes was assessed by ranking the NRMSE results per dataset per phenotype and by summing these ranks (Figure 6). The integrated data and the general composition layer had by far the lowest ($N = 40$) and highest sum of ranks ($N = 123$), respectively, while differences between the other datasets were more subtle. Interestingly, the models from the integrated data rank first only once but were consistently among the top five performers. In a relatively similar study on the prediction of phenotypic traits in potato using multi-omics data [67], integration of data layers did not enhance Random Forest model accuracies much more than when using individual layers. Layer contributions to trait prediction accuracies will undoubtedly depend on the biological system under investigation. In the current study, although many inter-layer covariate correlations exist, the substantial gain made by integrating the datasets indicate that some of these layers harbour unique information and are not redundant towards prediction of growth performance traits in fish. To summarise, considerable improvements in trait prediction accuracies were made by utilising all data rather than any single data modality alone.

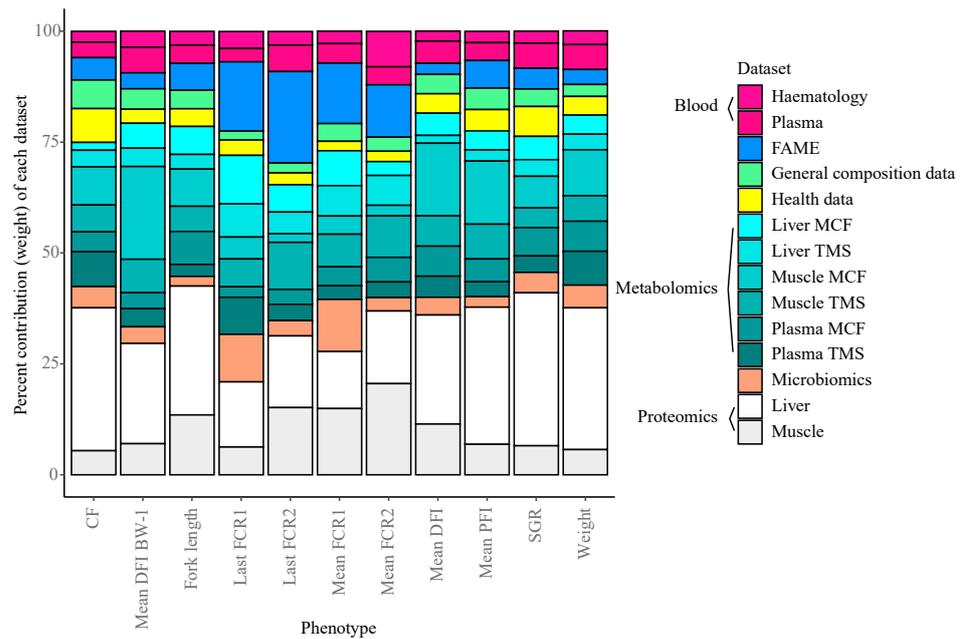


Figure 5. Mean dataset contribution (%) of the Random Block models per phenotype.

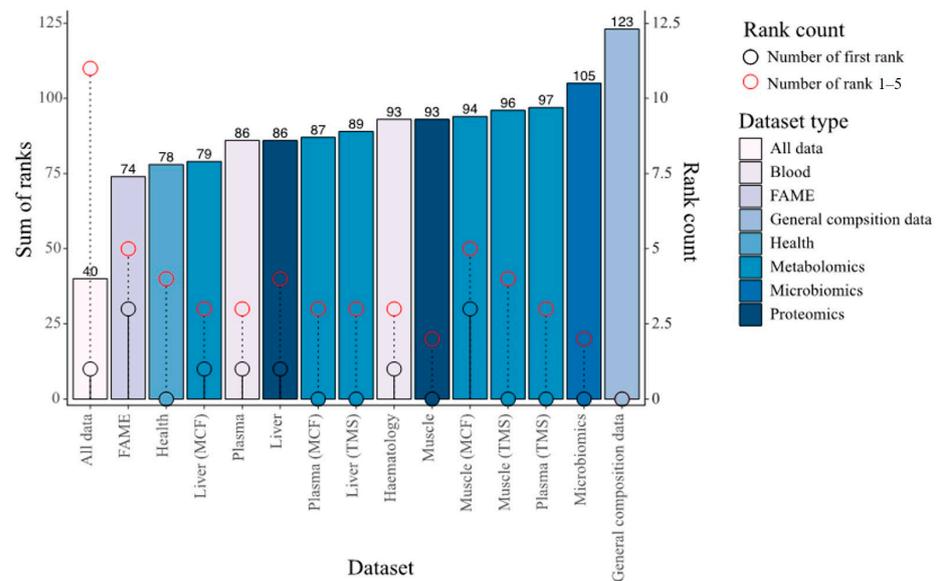


Figure 6. Overall performance (by ranking of NMRSE values per phenotype) of Random Block models trained per dataset. The dataset with the lowest NMRSE per phenotype was given a rank of one while the dataset with highest NMRSE was given a rank of 15 (Random Block models were trained on the 15 different datasets in the figure). The barplot represent the sum of the ranks allocated per phenotype per dataset. The black lollipops indicate the number of times models from the dataset ranked first while the red lollipops show the number of times they ranked within the five best datasets.

3.5. Important Random Block Predictor Variables for Feed Efficiency

Here we focus attention on the FE-specific models that were constructed in Figure 4, and interrogate them to identify a subset of the most predictive features. In Random Forest (RF) regression, the measure of ‘variable importance’ is used for feature selection. Analogous to the distinction between predictive importance and explanatory/causal importance in conventional regression models, RF-based feature selection serves two objectives: (1) to identify a smaller subset of variables sufficient for response/trait prediction, and

(2) to identify important variables highly related to the response/trait for explanatory and interpretation purposes. For the first objective, RF is highly applicable for constructing parsimonious trait prediction models since the algorithm is resilient to overfitting, is not impacted by high predictor multicollinearity, has versatility towards different variable types, is relatively robust to outliers, and can efficiently handle missing data and non-linear parameters. To enable biologically relevant interpretations for the second objective, RF has some limitations since magnification of all explanatory variables, including those with high redundancy, would ideally be achieved. Collinear variables which may be associated with a trait of interest can become down weighted in their variable importance scores if they do not enhance predictive accuracy, which means the full extent of informative features are often not realised [68]. Although variables which were most important to the accuracies of the Random Block models in predicting FE may only be a subset of those with significant biological meaning, functional evaluation of their roles are however warranted.

Looking more closely at the highest ranked co-occurring variables important to the best-performing Random Block models for FE, scope exists to gain mechanistic insight into FE variation. Many of these important predictor variables ($N = 27$) which originated from the integrated data were shared among all four FE trait prediction models (Figure 7). Interactions between this subset of variables and FE values were further assessed via correlation analyses (Figure 8) and a co-occurrence network (Figure 9), which additionally highlight some feature associations across the different data layers. These analyses notably show that three bacterial ASVs (i.e., 'x8993', 'ed55c' and 'x2846') from the family *Yersiniaceae* (genus *Serratia*) were negatively correlated with overall mean FE ($r = -0.51$ to -0.64 ; p -values < 0.01). *Serratia* spp. also formed positive correlations with N-acetylglutamic acid, hepatosomatic/liver index, and quinazoline, and a negative correlation with palmitic acid; all are significantly correlated to FE values. Members of *Yersiniaceae* are known fish pathogens [69–73], and their higher abundance in faeces of salmon with poor FE may indicate the existence of an intestinal infection/disease-related association.

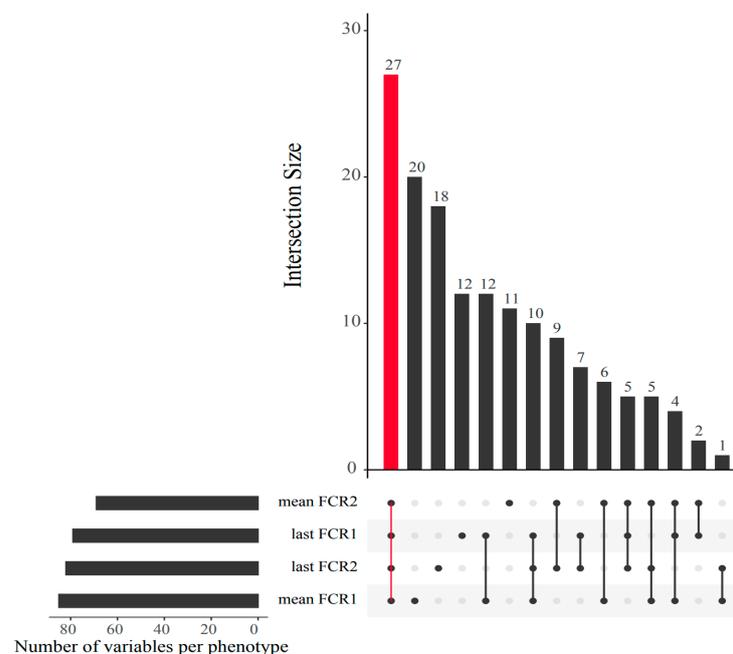


Figure 7. UpSet plot displaying shared predictor covariates between the best Random Block model for each of the four feed efficiency measures. The bar highlighted in red shows the number of covariates shared across all feed efficiency measures.

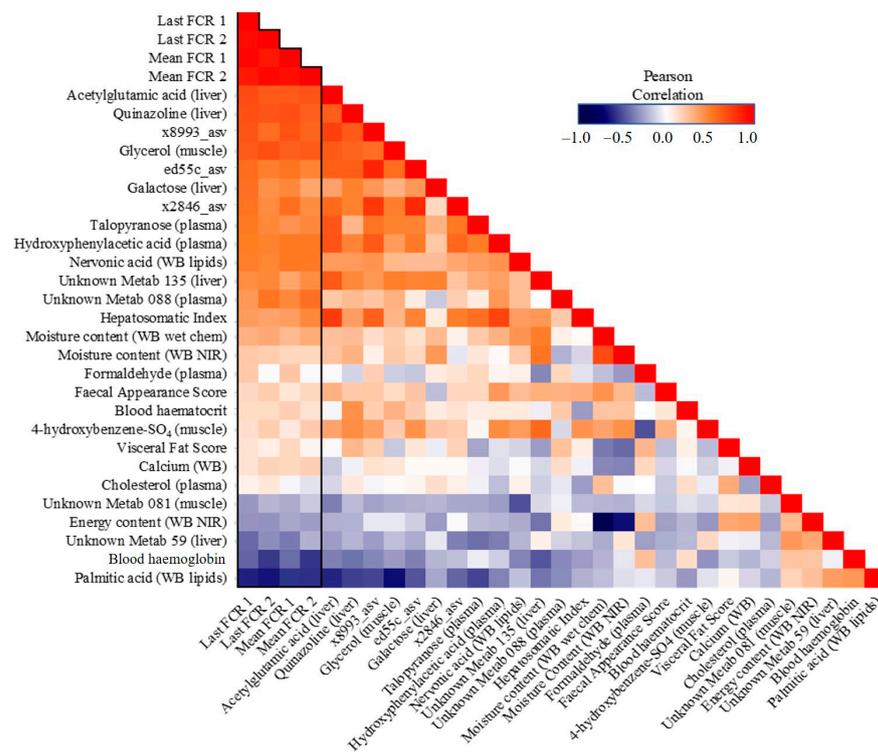


Figure 8. Correlation heatmap of important shared features (N = 27) in the Random Block prediction models for feed efficiency across all four measures.

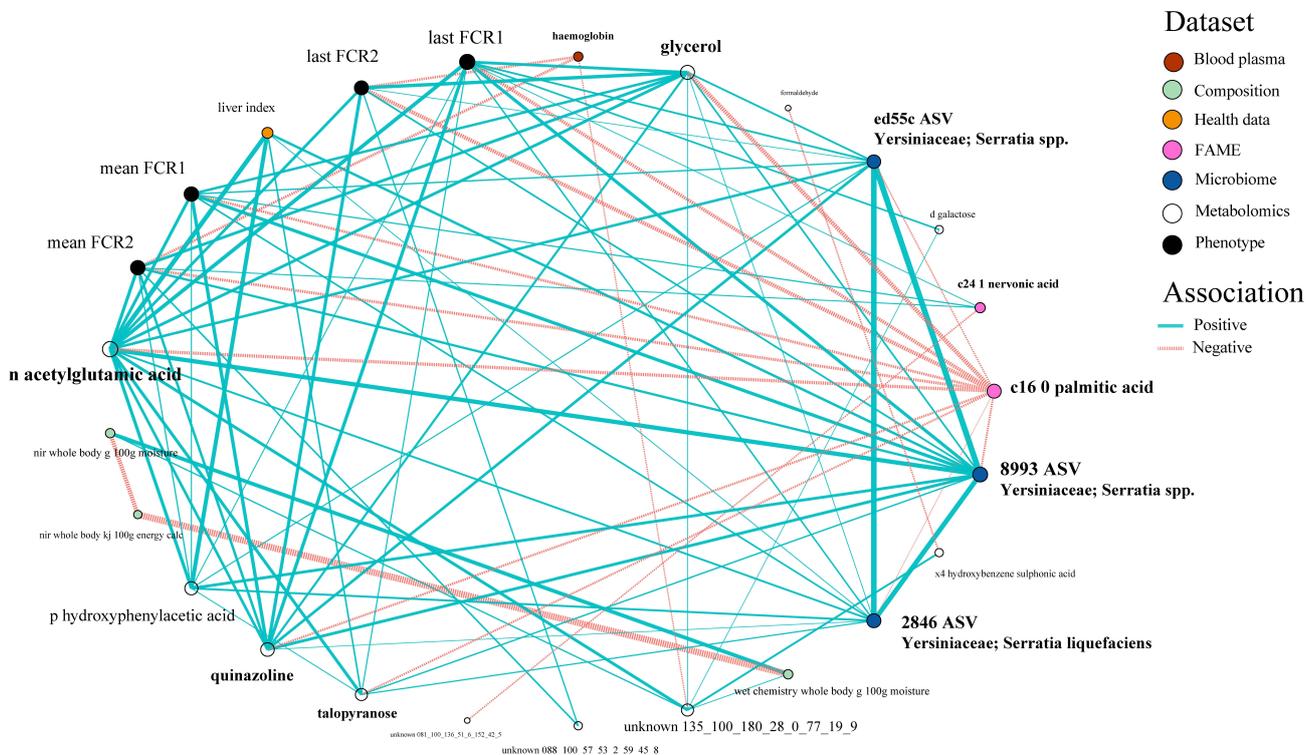


Figure 9. Associations between the 27 covariates which were consistently selected by the Random Block models in predicting FE, visualised via a trimmed co-occurrence network highlighting 21 major relationships.

Putative links between some clinical covariates and FE in fish were signalled, although weak–moderate linear correlations with the trait may be suggestive of more complex re-

relationships. FE predictors selected as important by the Random Block models included hepatosomatic index, visceral fat score, faecal appearance score, plasma cholesterol, blood haemoglobin, and haematocrit value. Interestingly, haemoglobin and haematocrit were recently identified as potential indicators of FE phenotypes in swine, with mechanistic underpinnings involving oxygen carrying capacity [74]. Further analysis of these parameters is warranted to substantiate links with FE in fish. All clinical covariates measured in this study were considered within the typical range.

Muscle glycerol showed a negative association ($r = -0.66$; $p < 0.001$) with overall mean FE. Higher glycerol concentrations in fish with poor FE may suggest increased energy utilisation of lipolysis substrates. Glycerol is currently receiving much attention as a feed supplement for the aquaculture of various fish species, with a proposed role in substrate provision for gluconeogenesis; thus, outcompeting dietary amino acids so they are more available for other physiological functions, such as growth [75,76]. Lower glycerol levels in fish with better FE may also reflect its increased substrate use for energy production, and differential turnover and partitioning of this metabolite should be considered an area for further study.

Levels of liver N-acetylglutamic acid (NAG) were also negatively associated ($r = -0.68$; $p < 0.001$) with FE. NAG is an essential allosteric cofactor for the first enzyme of the urea cycle in higher vertebrates. Most teleost fish directly clear the majority of nitrogenous waste as ammonia through the gills; however, *Oncorhynchus* spp. are known to produce ornithine-urea cycle enzymes as a supplementary route for nitrogen excretion and arginine synthesis [77–79]. FE in poultry can be modulated by NAG supplementation [80] and dietary N-carbamylglutamate, a metabolically stable analogue of NAG, improves protein efficiency ratios and growth in fish [81,82]. Variation in liver NAG may reflect a link between nitrogenous waste removal processes and FE in *O. tshawytscha*. This finding led us to look for a potential waste product relationship; levels of liver urea were also negatively correlated ($r = -0.68$; $p < 0.001$) with FE which supports a nitrogenous-based link. Urea excretion has a higher energetic cost than clearance of ammonia which may partly underpin this inverse association with FE.

Urea production in some fish taxa can occur through three routes: the ornithine-urea cycle (ammonia detoxification), the uricolytic pathway (purine degradation), and the arginolytic pathway (degradation of dietary arginine). Relative contributions of these processes towards ureagenesis depend on species [83]. We screened the liver proteomics data layer for proteins relevant to urea production, but no significant associations were identified between enzyme abundances and FE. Mechanistic foundations for the urea variations we observed in the current study should focus on enzyme activities.

Quinazoline is a predominant scaffold in many naturally occurring bioactive alkaloids, and levels in the liver of *O. tshawytscha* were negatively associated with overall mean FE ($r = -0.68$; $p < 0.001$). Marine sources of quinazoline derivatives are diverse, being synthesised by numerous bacterial, fungal, and marine invertebrate taxa, and a few fish species [84]. However, a functional role for this metabolite in salmon FE is currently uncertain.

Key fatty acid (FA) constituents of whole-body lipids which were associated with FE were nervonic acid (C24:1[n-9]) and palmitic acid (C16:0). Nervonic acid, a sphingomyelin-enriched long chain monounsaturated FA common in fish oils [85,86], displayed a negative relationship with FE ($r = -0.51$; $p < 0.01$). Nobrega et al. [87] recently identified a negative association between nervonic acid content and daily feed intake in Nile tilapia and suggested that selective incorporation of this FA into cell membranes during periods of food limitation may aid in winter survival, although a mechanism was not proposed. Whole-body palmitic acid content was positively associated ($r = 0.69$; $p < 0.001$) with FE. Palmitic acid is abundant in the diet of farmed salmon and is constitutively used as a saturated pairing to long-chain polyunsaturated FAs in polar lipids.

4. Conclusions

Improving feed efficiency (FE) is a strategy to reduce production costs and improve sustainability in fish aquaculture. Measurement of individual FE in fish is logistically challenging, time-consuming, and costly. The identification of reliable predictors of FE could reduce phenotyping efforts. FE is governed by a myriad of intrinsic and environmental factors, which multi-omic analyses are anticipated to contribute greater mechanistic understanding. This study demonstrated the novel use of the Random Block algorithm to integrate a diverse set of multimodal covariates in a salmonid (*Oncorhynchus tshawytscha*) model to predict a range of growth performance metrics, including FE. Our findings revealed the existence of strong concordance among covariate layers and demonstrated the unique contributions each layer provides towards trait prediction accuracies. FE-associations were made with some bacterial taxa, metabolites, and biochemical pathways, pointing towards key areas for subsequent hypothesis testing. Results further emphasise that FE is a highly complex trait; no single parameter we measured was able to reliably predict FE alone, but rather necessitated multiple variables for accurate prediction. Random Forest-based algorithms are useful for integrating multiple omics and clinical covariates and provide unique application for predicting growth and efficiency phenotypes in farmed Chinook salmon.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biology12081135/s1>, Methods S1: Experimental design, Methods S2: Microbiomics, Methods S3: Proteomics, Methods S4: Metabolomics, Methods S5: General composition, Methods S6: Fatty acids, Methods S7: Blood biomarkers, and Methods S8 Haematology.

Author Contributions: Conceptualization, J.E.S. and S.P.W.; Investigation, J.E.S., S.P.W., T.Y., M.R.M. and K.S.; Data curation, T.Y., J.E.S., S.P.W., O.L., P.C., M.R.M., K.S., N.E., R.Z., J.P.B., R.W. and A.B.; Data analysis, O.L., T.Y. and P.C.; Visualisation, O.L. and T.Y.; Writing—original draft preparation, T.Y. Writing—review and editing, J.E.S., S.P.W., O.L., M.R.M., P.C., K.S., N.E., R.Z., J.P.B., R.W., A.B., C.G.C., B.F.N. and A.C.A.; Project administration, J.E.S. and S.P.W.; Funding acquisition, J.E.S., S.P.W., C.G.C., B.F.N. and A.C.A. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the New Zealand Ministry of Business, Innovation and Employment (MBIE) Efficient Salmon research programme (CAWX1606). The research is a collaboration between the Cawthron Institute (New Zealand), Auckland University of Technology (New Zealand), University of Tasmania (Australia), and the Blue Economy Cooperative Research Centre (Australia).

Institutional Review Board Statement: The animal study protocol was conducted according to the guidelines of The Australian and New Zealand Council for the Care of Animals in Research and Teaching (ANZCCART), and was approved by the Nelson Marlborough Institute of Technology Animal Ethics Committee (AEC 2018 CAW01).

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available on reasonable request from the corresponding authors. The data are not publicly available due to privacy.

Acknowledgments: We would like to thank our team of colleagues at the Cawthron Aquaculture Park Finfish Research Centre (CAP FRC) for their dedicated assistance with systems maintenance, husbandry, and sampling: Nicholas Hearn, Gareth Nicholson, Chris Ensor, Michael Scott, Jordan Elvy, Chaya Bandaranayake, and Chris Chamberlain. We also thank Mount Cook Alpine Salmon for providing fish for this research.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Naylor, R.L.; Hardy, R.W.; Bureau, D.P.; Chiu, A.; Elliott, M.; Farrell, A.P.; Forster, I.; Gatlin, D.M.; Goldberg, R.J.; Hua, K.; et al. Feeding aquaculture in an era of finite resources. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15103–15110. [CrossRef] [PubMed]
2. Naylor, R.L.; Hardy, R.W.; Buschmann, A.H.; Bush, S.R.; Cao, L.; Klinger, D.H.; Little, D.C.; Lubchenco, J.; Shumway, S.E.; Troell, M. A 20-year retrospective review of global aquaculture. *Nature* **2021**, *591*, 551–563. [CrossRef] [PubMed]
3. New Zealand King Salmon. *New Zealand King Salmon Operations Report*; New Zealand King Salmon Ltd.: Nelson, New Zealand, 2019. Available online: <https://www.mpi.govt.nz/dmsdocument/16102-New-Zealand-King-Salmon-Operations-report> (accessed on 28 March 2023).
4. Iversen, A.; Asche, F.; Hermansen, Ø.; Nystøyl, R. Production cost and competitiveness in major salmon farming countries 2003–2018. *Aquaculture* **2020**, *522*, 735089. [CrossRef]
5. De Verdal, H.; Komen, H.; Quillet, E.; Chatain, B.; Allal, F.; Benzie, J.A.; Vandeputte, M. Improving feed efficiency in fish using selective breeding: A review. *Rev. Aquac.* **2018**, *10*, 833–851. [CrossRef]
6. Zhang, X.; Luan, P.; Cao, D.; Hu, G. A high-density genetic linkage map and fine mapping of QTL for feed conversion efficiency in Common carp (*Cyprinus carpio*). *Front. Genet.* **2021**, *12*, 778487. [CrossRef]
7. De Verdal, H.; O’Connell, C.M.; Mekki, W.; Vandeputte, M.; Chatain, B.; Bégout, M.L.; Benzie, J.A. Agonistic behaviour and feed efficiency in juvenile Nile tilapia *Oreochromis niloticus*. *Aquaculture* **2019**, *505*, 271–279. [CrossRef]
8. Martins, C.I.; Conceição, L.E.; Schrama, J.W. Feeding behavior and stress response explain individual differences in feed efficiency in juveniles of Nile tilapia *Oreochromis niloticus*. *Aquaculture* **2011**, *312*, 192–197. [CrossRef]
9. Carter, C.; Houlihan, D.; Keissling, A.; Medale, F.; Jobling, M. Physiological effects of feeding. In *Food Intake in Fish*; Houlihan, D., Boujard, T., Jobling, M., Eds.; Blackwell Science: Oxford, UK, 2001; pp. 297–331.
10. Lemieux, H.; Blier, P.; Dutil, J.D. Do digestive enzymes set a physiological limit on growth rate and food conversion efficiency in the Atlantic cod (*Gadus morhua*)? *Fish Physiol. Biochem.* **1999**, *20*, 293–303. [CrossRef]
11. Dawood, M.A.; Magouz, F.I.; Salem, M.F.; Elbially, Z.I.; Abdel-Daim, H.A. Synergetic effects of *Lactobacillus plantarum* and β -glucan on digestive enzyme activity, intestinal morphology, growth, fatty acid, and glucose-related gene expression of genetically improved farmed tilapia. *Probiotics Antimicrob. Proteins* **2020**, *12*, 389–399. [CrossRef]
12. Kaushik, S.J.; Schrama, J.W. Bioenergetics. In *Fish Nutrition*, 4th ed.; Hardy, R.W., Kaushik, S.J., Eds.; Academic Press: Cambridge, MA, USA, 2022; pp. 17–56.
13. Dvergedal, H.; Ødegård, J.; Øverland, M.; Mydland, L.T.; Klemetsdal, G. Selection for feed efficiency in Atlantic salmon using individual indicator traits based on stable isotope profiling. *Genet. Sel. Evol.* **2019**, *51*, 13. [CrossRef]
14. Dvergedal, H.; Harvey, T.N.; Jin, Y.; Ødegård, J.; Grønvold, L.; Sandve, S.R.; Våge, D.I.; Moen, T.; Klemetsdal, G. Genomic regions and signaling pathways associated with indicator traits for feed efficiency in juvenile Atlantic salmon (*Salmo salar*). *Genet. Sel. Evol.* **2020**, *52*, 66. [CrossRef] [PubMed]
15. Scholtens, M.; Dodds, K.; Walker, S.; Clarke, S.; Tate, M.; Slattery, T.; Preece, M.; Arratia, L.; Symonds, J. Opportunities for improving feed efficiency and spinal health in New Zealand farmed Chinook salmon (*Oncorhynchus tshawytscha*) using genomic information. *Aquaculture* **2023**, *563*, 738936. [CrossRef]
16. Carter, C.G.; Houlihan, D.F.; Buchanan, B.; Mitchell, A.I. Growth and feed utilization efficiencies of seawater Atlantic salmon, *Salmo salar* L., fed a diet containing supplementary enzymes. *Aquac. Res.* **1994**, *25*, 37–46. [CrossRef]
17. Hardy, R.W.; Kaushik, S.J. (Eds.) *Fish Nutrition*, 4th ed.; Academic Press: Cambridge, MA, USA, 2022; p. 905.
18. Elvy, J.E.; Symonds, J.E.; Hilton, Z.; Walker, S.P.; Tremblay, L.A.; Casanovas, P.; Herbert, N.A. The relationship of feed intake, growth, nutrient retention, and oxygen consumption to feed conversion ratio of farmed saltwater Chinook salmon (*Oncorhynchus tshawytscha*). *Aquaculture* **2022**, *554*, 738184. [CrossRef]
19. Esmaeili, M.; Carter, C.G.; Wilson, R.; Walker, S.P.; Miller, M.R.; Bridle, A.; Symonds, J.E. Proteomic investigation of liver and white muscle in efficient and inefficient Chinook salmon (*Oncorhynchus tshawytscha*): Fatty acid metabolism and protein turnover drive feed efficiency. *Aquaculture* **2021**, *542*, 736855. [CrossRef]
20. Esmaeili, N.; Carter, C.G.; Wilson, R.; Walker, S.P.; Miller, M.R.; Bridle, A.R.; Young, T.; Alfaro, A.C.; Laroche, O.; Symonds, J.E. An integrated proteomics and metabolomics investigation of feed efficiency in seawater reared Chinook salmon (*Oncorhynchus tshawytscha*). *Aquaculture* **2023**, *562*, 738845. [CrossRef]
21. Esmaeili, M.; Carter, C.G.; Wilson, R.; Walker, S.P.; Miller, M.R.; Bridle, A.R.; Symonds, J.E. Proteomic investigation of brain, liver and intestine in high feed intake and low feed intake Chinook salmon (*Oncorhynchus tshawytscha*). *Aquaculture* **2022**, *551*, 737915. [CrossRef]
22. Esmaeili, M.; Carter, C.G.; Wilson, R.; Walker, S.P.; Miller, M.R.; Bridle, A.R.; Symonds, J.E. Protein metabolism in the liver and white muscle is associated with feed efficiency in Chinook salmon (*Oncorhynchus tshawytscha*) reared in seawater: Evidence from proteomic analysis. *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2022**, *42*, 100994. [CrossRef] [PubMed]
23. Picard, M.; Scott-Boyer, M.P.; Bodein, A.; Périn, O.; Droit, A. Integration strategies of multi-omics data for machine learning analysis. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 3735–3746. [CrossRef]
24. Love, M. *Awesome Multi-Omics*; Github: San Francisco, CA, USA, 2023; Available online: <https://github.com/mikelove/awesome-multi-omics> (accessed on 15 July 2023).
25. Maghsoudi, Z.; Nguyen, H.; Tavakkoli, A.; Nguyen, T. A comprehensive survey of the approaches for pathway analysis using multi-omics data integration. *Brief. Bioinform.* **2022**, *23*, bbac435. [CrossRef]

26. Zhou, G.; Li, S.; Xia, J. Network-based approaches for multi-omics integration. *Comput. Methods Data Anal. Metab.* **2020**, *2104*, 469–487.
27. Agamah, F.E.; Bayjanov, J.R.; Niehues, A.; Njoku, K.F.; Skelton, M.; Mazandu, G.K.; Ederveen, T.H.; Mulder, N.; Chimusa, E.R.; t Hoen, P.A. Computational approaches for network-based integrative multi-omics analysis. *Front. Mol. Biosci.* **2022**, *9*, 1214. [[CrossRef](#)] [[PubMed](#)]
28. Bartmanski, B.J.; Rocha, M.; Zimmermann-Kogadeeva, M. Recent advances in data- and knowledge-driven approaches to explore primary microbial metabolism. *Curr. Opin. Chem. Biol.* **2023**, *75*, 102324. [[CrossRef](#)] [[PubMed](#)]
29. Hornung, R.; Wright, M.N. Block Forests: Random forests for blocks of clinical and omics covariate data. *BMC Bioinform.* **2019**, *20*, 358. [[CrossRef](#)]
30. Petralia, F.; Wang, P.; Yang, J.; Tu, Z. Integrative random forest for gene regulatory network inference. *Bioinformatics* **2015**, *31*, 197–205. [[CrossRef](#)]
31. Sarica, A.; Cerasa, A.; Quattrone, A. Random forest algorithm for the classification of neuroimaging data in Alzheimer’s disease: A systematic review. *Front. Aging Neurosci.* **2017**, *9*, 329. [[CrossRef](#)] [[PubMed](#)]
32. Wiedemann, C. On the Multiplicity of Analysis Strategies and the Resulting Biased Interpretations of a Large-Scale Benchmark Study. Ph.D. Dissertation, Ludwig Maximilian University of Munich, Munich, Germany, 2020.
33. Herrmann, M.; Probst, P.; Hornung, R.; Jurinovic, V.; Boulesteix, A.L. Large-scale benchmark study of survival prediction methods using multi-omics data. *Brief. Bioinform.* **2021**, *22*, bbaa167. [[CrossRef](#)] [[PubMed](#)]
34. Wissel, D.; Rowson, D.; Boeva, V. Clinically driven knowledge distillation for sparsifying high-dimensional multi-omics survival models. *bioRxiv* **2022**. [[CrossRef](#)]
35. Walker, S.P.; Ingram, M.; Bailey, J.; Dodds, K.G.; Fisher, P.J.; Amer, P.R.; Symonds, J.E. Chinook salmon (*Oncorhynchus tshawytscha*) feed conversion efficiency: Evaluation and potential for selection. *Proc. N. Z. Soc. Anim. Prod.* **2012**, *72*, 227–230.
36. Zhao, R.; Symonds, J.E.; Walker, S.P.; Steiner, K.; Carter, C.G.; Bowman, J.P.; Nowak, B.F. Effects of feed ration and temperature on Chinook salmon (*Oncorhynchus tshawytscha*) microbiota in freshwater recirculating aquaculture systems. *Aquaculture* **2021**, *543*, 736965. [[CrossRef](#)]
37. Větrovský, T.; Baldrian, P.; Morais, D. SEED 2: A user-friendly platform for amplicon high-throughput sequencing data analyses. *Bioinformatics* **2018**, *34*, 2292–2294. [[CrossRef](#)] [[PubMed](#)]
38. Gloor, G.B.; Macklaim, J.M.; Pawlowsky-Glahn, V.; Egozcue, J.J. Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* **2017**, *8*, 2224. [[CrossRef](#)] [[PubMed](#)]
39. Hughes, C.S.; Sorensen, P.H.; Morin, G.B. A standardized and reproducible proteomics protocol for bottom-up quantitative analysis of protein samples using SP3 and mass spectrometry. *Proteom. Biomark. Discov. Methods Protoc.* **2019**, *1959*, 65–87.
40. Johnson, W.E.; Li, C. Adjusting batch effects in microarray experiments with small sample size using empirical Bayes methods. In *Batch Effects and Noise in Microarray Experiments: Sources and Solutions*; John Wiley & Sons: Hoboken, NJ, USA, 2009; pp. 113–129.
41. Pang, Z.; Chong, J.; Zhou, G.; de Lima Morais, D.A.; Chang, L.; Barrette, M.; Gauthier, C.; Jacques, P.É.; Li, S.; Xia, J. MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **2021**, *49*, W388–W396. [[CrossRef](#)] [[PubMed](#)]
42. Villas-Bôas, S.G.; Smart, K.F.; Sivakumaran, S.; Lane, G.A. Alkylation or silylation for analysis of amino and non-amino organic acids by GC-MS? *Metabolites* **2011**, *1*, 3–20. [[CrossRef](#)]
43. Pinu, F.R.; Edwards, P.J.; Jouanneau, S.; Kilmartin, P.A.; Gardner, R.C.; Villas-Boas, S.G. Sauvignon blanc metabolomics: Grape juice metabolites affecting the development of varietal thiols and other aroma compounds in wines. *Metabolomics* **2014**, *10*, 556–573. [[CrossRef](#)]
44. Smart, K.F.; Aggio, R.B.; Van Houtte, J.R.; Villas-Bôas, S.G. Analytical platform for metabolome analysis of microbial cells using methyl chloroformate derivatization followed by gas chromatography–mass spectrometry. *Nat. Protoc.* **2010**, *5*, 1709–1729. [[CrossRef](#)]
45. Aggio, R.; Villas-Bôas, S.G.; Ruggiero, K. Metab: An R package for high-throughput analysis of metabolomics data generated by GC-MS. *Bioinformatics* **2011**, *27*, 2316–2318. [[CrossRef](#)]
46. Fan, S.; Kind, T.; Cajka, T.; Hazen, S.L.; Tang, W.W.; Kaddurah-Daouk, R.; Irvin, M.R.; Arnett, D.K.; Barupal, D.K.; Fiehn, O. Systematic error removal using random forest for normalizing large-scale untargeted lipidomics data. *Anal. Chem.* **2019**, *91*, 3590–3596. [[CrossRef](#)]
47. Association of Official Agricultural Chemists (AOAC). *International Official Methods of Analysis*, 21st ed.; AOAC International: Rockville, MD, USA, 2019.
48. Miller, M.R.; Puddick, J.; Symonds, J.E.; Walker, S.P.; Tian, H. Application of a Fourier transform—Near infrared reflectance spectroscopy method for the rapid proximate analysis of the greenshell mussel (*Perna canaliculus*) and king (Chinook) salmon (*Oncorhynchus tshawytscha*). *Aquac. Res.* **2019**, *50*, 1668–1677. [[CrossRef](#)]
49. Folch, J.; Lees, M.; Sloane Stanley, G.H. A simple method for the isolation and purification of total lipids from animal tissues. *J. Biol. Chem.* **1957**, *226*, 497–509. [[CrossRef](#)] [[PubMed](#)]
50. Casanovas, P.; Walker, S.P.; Johnston, H.; Johnston, C.; Symonds, J.E. Comparative assessment of blood biochemistry and haematology normal ranges between Chinook salmon (*Oncorhynchus tshawytscha*) from seawater and freshwater farms. *Aquaculture* **2021**, *537*, 736464. [[CrossRef](#)]

51. Araújo, B.C.; Symonds, J.E.; Walker, S.P.; Miller, M.R. Effects of fasting and temperature on the biological parameters, proximal composition, and fatty acid profile of Chinook salmon (*Oncorhynchus tshawytscha*) at different life stages. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* **2022**, *264*, 111113. [CrossRef]
52. Zarkasi, K.Z.; Taylor, R.S.; Abell, G.C.; Tamplin, M.L.; Glencross, B.D.; Bowman, J.P. Atlantic salmon (*Salmo salar* L.) gastrointestinal microbial community dynamics in relation to digesta properties and diet. *Microb. Ecol.* **2016**, *71*, 589–603. [CrossRef]
53. Jackson, D.A. PROTEST: A PROcrustean Randomization TEST of community environment concordance. *Écoscience* **1995**, *2*, 297–303. [CrossRef]
54. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016; Available online: <http://ggplot2.org> (accessed on 5 October 2022).
55. Shi, L.; Westerhuis, J.A.; Rosén, J.; Landberg, R.; Brunius, C. Variable selection and validation in multivariate modelling. *Bioinformatics* **2019**, *35*, 972–980. [CrossRef] [PubMed]
56. Conway, J.R.; Lex, A.; Gehlenborg, N. UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* **2017**, *33*, 2938–2940. [CrossRef]
57. Harrell, F.E.; Dupont, C. *Hmisc: Harrell Miscellaneous*; R Package Version 4.5-0; HBIOSSTAT: Nashville, TN, USA, 2021.
58. Pedersen, T.L. Ggraph: An Implementation of Grammar of Graphics for Graphs and Networks (2.0.5). 2021. Available online: <https://cran.r-project.org/package=ggraph%7D> (accessed on 5 October 2022).
59. Melnik, A.V.; da Silva, R.R.; Hyde, E.R.; Aksenov, A.A.; Vargas, F.; Bouslimani, A.; Protsyuk, I.; Jarmusch, A.K.; Tripathi, A.; Alexandrov, T.; et al. Coupling targeted and untargeted mass spectrometry for metabolome-microbiome-wide association studies of human fecal samples. *Anal. Chem.* **2017**, *89*, 7549–7559. [CrossRef]
60. Ayeni, F.A.; Biagi, E.; Rampelli, S.; Fiori, J.; Soverini, M.; Audu, H.J.; Cristino, S.; Caporali, L.; Schnorr, S.L.; Carelli, V.; et al. Infant and adult gut microbiome and metabolome in rural Bassa and urban settlers from Nigeria. *Cell Rep.* **2018**, *23*, 3056–3067. [CrossRef]
61. Yuan, C.; Graham, M.; Staley, C.; Subramanian, S. Mucosal microbiota and metabolome along the intestinal tract reveal a location-specific relationship. *Msystems* **2020**, *5*, e00055-20. [CrossRef]
62. Raposo de Magalhães, C.S.F.; Cerqueira, M.A.C.; Schrama, D.; Moreira, M.J.V.; Boonanuntanasarn, S.; Rodrigues, P.M.L. A Proteomics and other omics approach in the context of farmed fish welfare and biomarker discovery. *Rev. Aquac.* **2020**, *12*, 122–144. [CrossRef]
63. Natnan, M.E.; Low, C.F.; Chong, C.M.; Bunawan, H.; Baharum, S.N. Integration of omics tools for understanding the fish immune response due to microbial challenge. *Front. Mar. Sci.* **2021**, *8*, 668771. [CrossRef]
64. Natnan, M.E.; Mayalvanan, Y.; Jazamuddin, F.M.; Aizat, W.M.; Low, C.F.; Goh, H.H.; Azizan, K.A.; Bunawan, H.; Baharum, S.N. Omics strategies in current advancements of infectious fish disease management. *Biology* **2021**, *10*, 1086. [CrossRef] [PubMed]
65. Shivam, S.; El-Matbouli, M.; Kumar, G. Development of fish parasite vaccines in the OMICs era: Progress and opportunities. *Vaccines* **2021**, *9*, 179. [CrossRef] [PubMed]
66. Zhang, F.; Ekine-Dzivenu, C.; Vinsky, M.; Basarab, J.A.; Aalhus, J.L.; Dugan, M.E.R.; Li, C. Phenotypic and genetic relationships of residual feed intake measures and their component traits with fatty acid composition in subcutaneous adipose of beef cattle. *J. Anim. Sci.* **2017**, *95*, 2813–2824. [CrossRef]
67. Acharjee, A.; Kloosterman, B.; Visser, R.G.; Maliepaard, C. Integration of multi-omics data for prediction of phenotypic traits using random forest. *BMC Bioinform.* **2016**, *17*, 363–373. [CrossRef]
68. Degenhardt, F.; Seifert, S.; Szymczak, S. Evaluation of variable selection methods for random forests and omics data sets. *Brief. Bioinform.* **2019**, *20*, 492–503. [CrossRef]
69. McIntosh, D.; Austin, B. Recovery of an extremely proteolytic form of *Serratia liquefaciens* as a pathogen of Atlantic salmon, *Salmo solar*, in Scotland. *J. Fish Biol.* **1990**, *36*, 765–772. [CrossRef]
70. Aydin, S.; Erman, Z.; Bilgin, Ö.C. Investigations of *Serratia liquefaciens* infection in rainbow trout (*Oncorhynchus mykiss* Walbaum). *Turk. J. Vet. Anim. Sci.* **2001**, *25*, 643–650.
71. Crouse, C.; Davidson, J.; May, T.; Summerfelt, S.; Good, C. Production of market-size European strain Atlantic salmon (*Salmo salar*) in land-based freshwater closed containment aquaculture systems. *Aquac. Eng.* **2021**, *92*, 102138. [CrossRef]
72. Matter, A.F.; El Asely, A.M.; Shaheen, A.A.; El-Gawad, E.A.A.; El-Abd, H.; Abbass, A.A. Phenotypic and molecular characterization of bacterial pathogens isolated from diseased freshwater fishes. *Int. J. Fish. Aquat. Stud.* **2018**, *6*, 34–41.
73. Lane, H.S.; Brosnahan, C.L.; Poulin, R. Aquatic disease in New Zealand: Synthesis and future directions. *N. Z. J. Mar. Freshw. Res.* **2022**, *56*, 1–42. [CrossRef]
74. Lindholm-Perry, A.K.; Kuehn, L.A.; Wells, J.E.; Rempel, L.A.; Chitko-McKown, C.G.; Keel, B.N.; Oliver, W.T. Hematology parameters as potential indicators of feed efficiency in pigs. *Transl. Anim. Sci.* **2021**, *5*, txab219. [CrossRef] [PubMed]
75. Palma, M.; Tavares, L.C.; Rito, J.; Henriques, L.F.; Silva, J.G.; Ozório, R.; Pardal, M.A.; Magnoni, L.J.; Viegas, I. Metabolic effects of dietary glycerol supplementation in muscle and liver of European seabass and rainbow trout by ¹H NMR metabolomics. *Metabolites* **2019**, *9*, 202. [CrossRef]
76. Mauerwerk, M.T.; Zadelino, I.V.; Meurer, F. Use of glycerol in fish nutrition: A review. *Rev. Aquac.* **2021**, *13*, 853–861. [CrossRef]
77. Todgham, A.E.; Anderson, P.M.; Wright, P.A. Effects of exercise on nitrogen excretion, carbamoyl phosphate synthetase III activity and related urea cycle enzymes in muscle and liver tissues of juvenile rainbow trout (*Oncorhynchus mykiss*). *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* **2001**, *129*, 527–539. [CrossRef]

78. Clark, T.C.; Tinsley, J.; Macqueen, D.J.; Martin, S.A.M. Rainbow trout (*Oncorhynchus mykiss*) urea cycle and polyamine synthesis gene families show dynamic expression responses to inflammation. *Fish Shellfish. Immunol.* **2019**, *89*, 290–300. [[CrossRef](#)]
79. Jung, E.H.; Smich, J.; Rubino, J.G.; Wood, C.M. An in vitro study of urea and ammonia production and transport by the intestinal tract of fed and fasted rainbow trout: Responses to luminal glutamine and ammonia loading. *J. Comp. Physiol. B* **2021**, *191*, 273–287. [[CrossRef](#)]
80. Wang, J.; Lin, J.; Wang, J.; Wu, S.; Qi, G.; Zhang, H.; Song, Z. Effects of in ovo feeding of N-acetyl-L-glutamate on early intestinal development and growth performance in broiler chickens. *Poult. Sci.* **2020**, *99*, 3583–3593. [[CrossRef](#)]
81. Cheng, W.; Zhang, L.; Xu, G.; Wu, Q.; Xiong, D.; Guo, Y.; Tan, W.; Xu, D. Effects of arginine on the regulation of the growth, the blood amino acid composition and the fat deposition in Nile tilapia (*Oreochromis niloticus*). *Acta Hydrobiol. Sin.* **2015**, *39*, 490–497.
82. Li, W.; Liu, B.; Liu, A.; Han, M.; Yin, Y.; Xu, G.; Cheng, W.; Xie, L. Dietary supplementation of N-carbamylglutamate promotes growth performance by modulating the homeostasis of gut microbiota in tilapia (*Oreochromis niloticus*). *Aquac. Rep.* **2021**, *20*, 100750. [[CrossRef](#)]
83. Wright, P.A.; Land, M.D. Urea production and transport in teleost fishes. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* **1998**, *119*, 47–54. [[CrossRef](#)] [[PubMed](#)]
84. D'yakonov, A.L.; Telezhenetskaya, M.V. Quinazoline alkaloids in nature. *Chem. Nat. Compd.* **1997**, *33*, 221–267. [[CrossRef](#)]
85. Ratnayake, W.N.; Galli, C. Fat and fatty acid terminology, methods of analysis and fat digestion and metabolism. *Ann. Nutr. Metab.* **2009**, *55*, 8–43. [[CrossRef](#)] [[PubMed](#)]
86. Gylfason, G.A.; Knútsdóttir, E.; Ásgeirsson, B. Nervonic acid (24: 1n-9) is a dominant unsaturated fatty acid in the intestinal brush border of Atlantic cod. *Lipid Insights* **2012**, *5*, LPI-S10291. [[CrossRef](#)]
87. Nobrega, R.O.; Banze, J.F.; Batista, R.O.; Fracalossi, D.M. Improving winter production of Nile tilapia: What can be done? *Aquac. Rep.* **2020**, *18*, 100453. [[CrossRef](#)]

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