RESEARCH

SALARECON connects the Atlantic salmon genome to growth and feed efficiency

Maksim Zakhartsev[1†], Filip Rotnes[1†], Marie Gulla[1†], Ove Øyås[1,2†], Jesse C. J. van Dam[3], Maria Suarez-Diez[3], Fabian Grammes[2], Róbert Anton Hafþórsson[2], Wout van Helvoirt[2], Jasper J. Koehorst[3], Peter J. Schaap[3], Yang Jin[2], Liv Torunn Mydland[2], Arne B. Gjuvsland[2], Simen R. Sandve[2], Vitor A. P. Martins dos Santos[3] and Jon Olav Vik[1*]

[*]Correspondence: jon.vik@nmbu.no
[1]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences (NMBU), Ås, Norway
Full list of author information is available at the end of the article
[†]Equal contributor.

Abstract

Background: Atlantic salmon (*Salmo salar*) is the most valuable farmed fish globally and there is much interest in optimizing its genetics and rearing conditions for growth and feed efficiency. Marine feed ingredients must be replaced to meet global demand, with challenges for fish health and sustainability. Metabolic models can address this by connecting genomes to metabolism, which converts nutrients in the feed to energy and biomass, but such models are currently not available for major aquaculture species such as salmon.

Results: We present SALARECON, a metabolic model that links the Atlantic salmon genome to metabolic fluxes and growth. It performs well in standardized tests and captures expected metabolic (in)capabilities. We show that it can explain observed hypoxic growth in terms of metabolic fluxes and apply it to aquaculture by simulating growth with commercial feed ingredients. Predicted limiting amino acids and feed efficiencies agree with data, and the model suggests that marine feed efficiency can be achieved by supplementing a few amino acids to plant- and insect-based feeds.

Conclusion: SALARECON is a high-quality model that makes it possible to simulate Atlantic salmon metabolism and growth. It can be used to explain Atlantic salmon physiology and address key challenges in aquaculture such as development of sustainable feeds.

Keywords: Atlantic salmon; *Salmo salar*; systems biology; aquaculture; metabolism; metabolic model

Background

Salmonid aquaculture has grown in volume and economic importance over the past several decades, and Atlantic salmon (*Salmo salar*) has become the world's most valuable fish commodity [1]. This is largely thanks to selective breeding, which has improved both growth rate and feed efficiency [2]. The increase in fish farming has also increased demand for feed, and insufficient marine resources has led to a switch to plant-based ingredients [3]. This has reduced production costs and exploitation of fish stocks, but salmon are not adapted to eating plants and current plant-based feeds have a negative impact on fish health and the environment [4, 5]. Also, plant-based feeds are complex, the ingredient market is fluctuating, and feeding trials are demanding. Thus, developing feeds that minimize cost and environmental impact while providing necessary nutrients to the fish is an important challenge [6].

The metabolic network of a cell or organism converts nutrients that are present in the environment to the energy and building blocks that are required to live and grow. It consists of metabolites that are interconverted by metabolic reactions, most of which are catalyzed by enzymes that are encoded by the genome, and it can be translated to a metabolic model, which allows mathematical analysis of network functionality through methods such as flux balance analysis (FBA) [7]. Specifically, metabolic models allow prediction of growth and metabolic fluxes (steady-state reaction rates) that are linked to the genome through logical gene-protein-reaction (GPR) associations, making them promising tools for addressing challenges in aquaculture such as breeding for feed efficiency and sustainable feed development [8]. Large databases of metabolic reactions and models [9, 10, 11] and methods for metabolic network reconstruction from annotated genomes [12, 13] have made such models available for organisms ranging from microbes to animals [14]. However, there are still very few metabolic models of fish available [15, 16, 17, 18] and none of Atlantic salmon or other important farmed fish species.

Here, we present SALARECON: a metabolic model built from the Atlantic salmon genome [19] that predicts growth and metabolic fluxes. SALARECON is a high-quality model according to community-standardized tests, and it captures expected metabolic (in)capabilities such as amino acid essentiality. Using oxygen-limited growth under hypoxia as an example, we show that model predictions can explain salmon physiology in terms of metabolic fluxes that are, in turn, tied to genes and pathways. Furthermore, we demonstrate an important application for aquaculture by predicting growth-limiting amino acids and feed efficiencies for commercial feed ingredients in agreement with data.

Results

We built a metabolic model of Atlantic salmon (SALARECON) from its genome [19], metabolic reaction and model databases, and literature (**Fig. 1**). The model focuses on energy metabolism and covers 1,104 genes, which amounts to 2% of the 47,329 annotated genes in the genome and 48% of the 2,281 Atlantic salmon genes that are associated with metabolic reactions in KEGG [11]. The genes are mapped to a metabolic network of 718 reactions and 530 metabolites (**Fig. 2a**) with node degree distributions that are typical for metabolic and other biological networks [20] (**Fig. S1**). Reactions and metabolites are divided between six compartments: cytosol, mitochondrion, inner mitochondrial membrane, extracellular environment, peroxisome, and nucleus (**Fig. 2b**). The compartments are connected by 175 transport reactions that allow metabolite exchange through the cytosol, and 86 boundary reactions allow metabolites to move in and out of the system through the extracellular environment. There are 357 unique metabolites when those occurring in multiple compartments are counted once. A salmon-specific biomass reaction based on whole-body composition [21] allows growth rate prediction by accounting for production of the proteins, lipids, carbohydrates and nucleic acids that consitute biomass from metabolites supplied by the metabolic network (**Fig. 2c**).

To investigate whether SALARECON is likely to be an accurate representation of Atlantic salmon metabolism, we first compared it to the only existing high-quality metabolic model of a fish as well as all models of multicellular eukaryotes currently available in the BiGG database [10] (**Fig. 3a** and **Fig. S2**). Specifically, we

hierarchically clustered the reaction contents of SALARECON and models of zebrafish (*Danio rerio*) [17], mouse (*Mus musculus*) [22], chinese hamster ovary (CHO, *Cricetulus griseus*) [23], human (*Homo sapiens*) [24], and the diatom *Phaeodactylum tricornutum* [25] based on Jaccard distance. The models clustered by phylogeny with fish and mammals forming distinct groups and the diatom as an outlier, indicating that SALARECON captures fish- and likely salmon-specific metabolism. SALARECON also performed well in community-standardized MEMOTE tests [26], which evaluate model consistency and annotation (**Fig. 3b**). It achieved an overall MEMOTE score of 95% (best possible score is 100%) with subscores of 94% for model consistency, 100% for Systems Biology Ontology (SBO) annotation, 97% for metabolite annotation, 87% for reaction annotation, and 53% for gene annotation. Finally, we evaluated the ability of SALARECON to perform 210 metabolic tasks grouped into seven metabolic systems (**Fig. 3c**) and 73 metabolic subsystems (**Fig. S3**). These tasks were originally defined for mammalian cells [27] but we changed the expected outcomes of amino acid synthesis tests to match known essentiality in Atlantic salmon. SALARECON correctly captured all expected metabolic (in)capabilities for the three metabolic systems within the scope of the model (energy, amino acid, and nucleotide metabolism). It also succeeded in 44% of vitamin and cofactor tasks, 43% of carbohydrate tasks, and 15% of lipid tasks, reflecting the fact that these parts of metabolism are simplified in the model. The only system completely outside the scope of SALARECON was glycan metabolism, in which no tasks were successfully performed. In total, SALARECON succeeded in 66% of all metabolic tasks.

In our first application of SALARECON, we predicted oxygen-limited growth rates under hypoxia on a minimal feed containing essential amino acids and choline, using uniform random sampling to account for uncertainty in feed nutrient ratios and flux capacities (**Fig. 4a** and **Fig. S4**). Assuming that relative oxygen uptake rate is a linear function of water oxygen saturation (percent air saturation), we fitted our predictions to experimental data [28, 29, 30, 31] along with a logistic model and a Monod model (**Fig. 4b**). The choice of a linear model for the metabolic fit was motivated by the fact that diffusive oxygen uptake in fish gills is governed by Fick's law and therefore proportional to the oxygen gradient [32]. Also, replacing the linear model by a Michaelis-Menten model would make the metabolic and Monod fits virtually identical because the Monod and Michaelis-Menten equations have the same form. We found that the metabolic, logistic, and Monod models fitted the data about equally well ($R^2 \approx 0.6$) but they differed in their parameter estimates (**Fig. 4b**). All the models estimated the minimal oxygen saturation required for growth, but the logistic estimate was low with high standard error ($x_0 = 0.11 \pm 0.16$) and the Monod fit was high with low standard error ($x_0 = 0.45 \pm 0.04$). The metabolic model gave an intermediate estimate and standard error ($x_0 = 0.35 \pm 0.08$), and it also allowed estimation of the minimal oxygen saturation required for *maximal* growth ($x_1 = 1.43 \pm 0.17$). The metabolic fit was closer than the two other fits to the expected relationship between water oxygen saturation and growth rate presented by Thorarensen et al. [32], both in terms of the shape of the fitted curve and the estimated parameter values. The SALARECON estimates were within one and two standard errors, respectively, of the values $x_0 \approx 0.3$ and $x_1 \leq 1.2$ suggested

by Thorarensen et al. The logistic estimate was within two standard errors of the suggested x_0, but this confidence interval also included zero.

In contrast to the simple growth models, SALARECON is mechanistic and makes it possible to explain predictions in terms of metabolic fluxes (**Fig. 4d**). Assuming that organisms have evolved to grow as efficiently as possible, i.e. with minimal enzymatic cost, we used parsimonius flux balance analysis (pFBA) [33] to minimize overall flux through the metabolic network while requiring maximal growth rate for each randomly sampled condition and oxygen level. We identified six clusters of reactions whose pFBA fluxes made distinct contributions to oxygen-limited growth (**Fig. 4e**). Connecting clusters to the Atlantic salmon genome through GPR associations, we also identified enriched metabolic pathways among the genes associated with each cluster (**Fig. 4f**). In two clusters, fluxes were closely correlated with relative growth rate, indicating that they contained reactions that were virtually always necessary for growth. Indeed, these clusters were enriched in lipid and nucleotide metabolism, which produce biomass precursors, and nitrogen metabolism, which is involved in amino acid biosynthesis and disposal of nitrogen-containing waste products. The fluxes of another cluster tracked this first cluster closely at the lowest oxygen levels but then decreased relative to the first cluster, leveling off and remaining more or less constant at all higher oxygen levels. Notably, this cluster was enriched in glycolysis, pyruvate, and thiamine metabolism, indicating that flux through glycolysis was maximized at low oxygen levels and kept constant while other energy-generating pathways were activated at higher oxygen levels. For two of the three remaining clusters, fluxes increased slightly less than in the first three at low oxygen levels before increasing linearly across higher oxygen levels in one cluster and plateauing and decreasing slightly in the other. These clusters were enriched in pathways related to metabolism of fatty acids and amino acids, suggesting that these compounds become important energy sources after saturation of glycolysis at low oxygen levels. The final cluster consisted of reactions with no or very little flux, even at the highest oxygen levels, and was enriched in metabolism of pyrimidines, β-alanine, and the essential amino acids tryptophan and histidine.

Finally, to demonstrate the potential of SALARECON to address key challenges in aquaculture, we used it to predict growth-limiting amino acids and feed efficiencies for three commercial feed ingredients: fish, soybean, and insect meal (**Table 1**). For each feed ingredient, we iteratively identified and supplemented the most limiting amino acid until all amino acid limitations had been lifted, computing feed efficiency at each iteration. Comparing predicted limiting amino acids in fish meal to soybean and insect meal, we found that lysine and threonine were more limiting in both soybean and insect meal, methionine was more limiting in soybean meal, and arginine was more limiting in insect meal (**Fig. 5a**, **Fig. 5b**, and **Fig. S5**). The feed efficiency predictions suggest that the baseline feed efficiency of fish meal can be achieved by supplementing one and three amino acids for soybean and insect meal, respectively (**Fig. 5c**). For soybean meal, major increases in feed efficiency were predicted for lysine, threonine, and methionine supplementation, while lysine had the largest impact on insect meal (**Fig. S5**). The predictions from SALARECON agree well with expected baseline feed efficiencies [34, 35] as well as reports that lysine, methionine, threonine, and arginine are more limiting in plant-based than in marine feeds [36, 37].

Discussion

SALARECON is the first metabolic model of a production animal, bridging the gap between production and systems biology and initiating a framework for adapting Atlantic salmon breeding and nutrition strategies to modern feeds. By explicitly representing connections between metabolites, reactions, and genes, it connects the genome to metabolism and growth in a way that can be tuned to specific genetic and environmental contexts by integration of domain nowledge and experimental data [8]. Thus, SALARECON forms a transdiciplinary framework for diverse disciplines and data sets involved in Atlantic salmon research and aquaculture. Tools developed for constraint-based modeling of microbes and well-studied plants and animals can now be applied in production biology, providing a sharper lens through which to interpret omics data by requiring consistency with flux balances and other known constraints. This enables clearer analysis than classical multivariate statistics, which does not incorporate such mechanistic knowledge.

Although laborious and time-consuming, our bottom-up manual reconstruction of the Atlantic salmon metabolic network was necessary to make SALARECON a high-quality predictive model. Automatically built models work well for microbes but are still outperformed by models that are built by manual iteration, and reconstruction of eukaryotes is more challenging due to larger genomes, less knowledge, and compartmentalization [13, 12]. However, semi-automated annotation and curation combined with automated MEMOTE tests [26] and metabolic tasks [27] allowed faster iteration, and future reconstructions of related species [38] can benefit from our efforts by using SALARECON as a template. MEMOTE and metabolic tasks were instrumental in the development of SALARECON, and we highly recomend integrating testing in model development. Tests help catch mistakes that arise when modifying a model and do triple duty by specifying what it should be capable of, identifying broken functionality, and forming a basis for comparison with other models, e.g. new versions or models of different tissues or species. Clearly formulated tests also make the model more accessible to non-modelers, speaking the same language as nutritionists or physiologists. Such experts can point out missing or ill-formulated tests, which in turn contribute to improvement

We have strived to make SALARECON an accurate model of Atlantic salmon metabolism and growth, but it does not aim to capture salmon physiology exhaustively or perfectly. It covers 2% of the genes in the genome, which amounts to 48% of Atlantic salmon genes mapped to reactions in KEGG [11], and its focus is on core metabolism generating energy and biomass. This covers pathways that connect feed to fillet, which is a primary focus of research and aquaculture, but obviously excludes many other interesting processes such as synthesis of long-chain polyunsaturated fatty acids. Still, SALARECON performs very well according to all of our metrics: it is more similar to the latest zebrafish model [17] than to any other multicellular eukaryote for which a model is available in BiGG [22, 23, 24, 25], achieves a MEMOTE score of 95%, which is better than all models in BiGG [10], and performs all metabolic tasks within the scope of the model (amino acid, nucleotide, and energy metabolism). It also correctly classifies amino acids as essential [21]. SALARECON has a low MEMOTE subscore for gene annotation compared to subscores for reaction and metabolite annotation, but this is because Atlantic salmon genes can be mapped to fewer databases than generic biochemical components.

The extensive annotation of genes, metabolites, and reactions is a key strength of SALARECON that facilitates use with existing models, tools, and data. In particular, identifiers from BiGG [10] make it easy to compare and combine SALARECON with state-of-the-art models [39, 40], e.g. to predict interactions between Atlantic salmon and its gut microbiota. It also allows direct application of implemented methods such as evaluation of metabolic tasks [27]. Another crucial feature is the salmon-specific biomass reaction, which makes SALARECON a more realistic representation of salmon metabolism than a simple network reconstruction [12]. It enables prediction of growth and related fluxes and is based on organism-specific data [21] rather than copied from mammalian models, as has been done for zebrafish [17]. As demonstrated for Atlantic cod [18], even getting to this stage is challenging for non-model animals. Notably, SALARECON predicts growth in a minimal environment with only essential amino acids and choline as a precursor for lipids.

Our analysis of growth under oxygen limitation shows that phenotypes predicted by SALARECON can be fitted to experimental data and produce mechanistic explanations of Atlantic salmon physiology. Specifically, SALARECON explained hypoxic metabolism and growth in terms of metabolic fluxes with implications for fish welfare and productivity in aquaculture. The growth predictions depend on unknown environmental conditions and flux capacities, but SALARECON can be used to account for such uncertainty through random sampling. Average growth predictions from SALARECON fitted the available data [28, 29, 30, 31] as well as simple growth models and gave accurate estimates of critical water oxygen saturations in agreement with observations [32]. The predicted metabolic fluxes defined clusters of reactions with distinct pathway enrichments and contributions to hypoxic growth, notably suggesting that glycolysis becomes saturated at low oxygen levels and that amino and fatty acids become important energy sources with increasing oxygen. Predictions contrasting growth-limiting amino acids in three commercial feed ingredients also agreed well with data [36, 37] and showed that SALARECON can be used to evaluate the efficiency of sustainable feeds, a key challenge for modern aquaculture. Feed efficiencies predicted by SALARECON lie within reported ranges [34, 35] and suggest that the feed efficiency of fish meal can be achieved by supplementing one amino acid for insect meal and three for soybean meal. This shows that SALARECON can be used to evaluate both current and novel feeds, potentially reducing the need for expensive fish experiments *in vitro* or *in vivo*.

In future work, we will expand SALARECON to cover more processes such as lipid and carbohydrate metabolism in full detail, and we will tailor it to gut, liver, muscle, and other tissues using omics data and metabolic tasks [27]. We will also leverage automated metabolic reconstruction tools for microbes to build models the Atlantic salmon gut microbiota [39]. By coupling tissue-specific models to each other and to gut microbiota models, we can make detailed and partially dynamic whole-body models [42]. This would be a major leap from available dynamic models [43] and provide a mechanistic alternative to state-of-the-art bioenergetics models [44], opening up new possibilities for understanding fish physiology and rational engineering of feeds, conditions, and genetics.

Conclusions

SALARECON covers half of the annotated metabolic genes in the Atlantic salmon genome and can predict metabolic fluxes and growth with a salmon-specific biomass reaction. It has been extensively annotated, curated, and evaluated, and it can be used to tackle research questions from fish physiology to aquaculture. In particular, SALARECON is a promising new tool for predicting breeding strategies and novel feeds that optimize for production parameters such as feed efficiency and impact on fish health and environment. Future work will expand SALARECON and integrate it with omics data to make tissue-specific and partially dynamic whole-body models. SALARECON should facilitate systems biology studies of Atlantic salmon and other salmonids, and we hope that it will be widely used by modelers as well as biologists.

Methods

Building the metabolic model

First, we manually built a draft model of Atlantic salmon core metabolism using the genome [19] with KEGG [11] annotations and the software Insilico Discovery (Insilico Biotechnology, Stuttgart, Germany). We used WoLF PSORT [45] and SAPP [46] to assign metabolites and reactions to six different compartments (cytosol, mitochondrion, inner mitochondrial membrane, extracellular environment, peroxisome, and nucleus). Exchange reactions were added to allow metabolite import (negative flux) and export (positive flux). Second, we used COBRApy [47] to annotate and curate the draft model. We iteratively converted the model to the BiGG [10] namespace and added annotations from MetaNetX [9], BiGG [10], KEGG [11], and UniProt [48]. We also added and removed metabolites, reactions, and genes, mapped genes to reactions using AutoKEGGRec [49], and added a salmon-specific biomass reaction. Due to lack of knowledge, we formulated all gene-protein-reaction (GPR) associations as OR relations, effectively ignoring enzyme complexes and assuming that all genes mapped to a reaction encode isozymes. To build the biomass reaction, we estimated the fractional composition of macromolecules in 1 g dry weight biomass (gDW) from Atlantic salmon whole-body composition [21]. We mapped macromolecules to metabolites and estimated the fractional composition of amino acids in proteins and nucleoside triphosphates in nucleic acids from proteome and genome sequences, respectively. Finally, we evaluated the quality of the model as described below and then alternated semi-automated annotation and curation with quality evaluation until we saw no further opportunities to improve it without expanding its scope. The final model was exported to Systems Biology Markup Language (SBML) format [50].

Evaluating the quality of the metabolic model

First, we compared the reaction contents in SALARECON to other models of multicellular eukaryotes available in the BiGG [10] namespace (*Danio rerio* [17], *Mus musculus* [22], *Cricetulus griseus* [23], *Homo sapiens* [24], and *Phaeodactylum tricornutum* [25]). We excluded boundary reactions and computed the Jaccard distance $J(A, B)$ between each pair of models A and B by dividing the size of the intersection of reactions in the two models by the size of the union:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

The Jaccard distances were used for complete-linkage agglomerative hierarchical clustering of the models. Second, we tested SALARECON's consistency and annotation using the community standard MEMOTE [26] and its metabolic (in)capabilities using tasks defined for mammalian cells [27]. We adapted tasks to Atlantic salmon by moving metabolites from compartments not included in SALARECON to the cytoplasm and modifying the expected outcomes of amino acid synthesis tests to match known essentiality [21]. Finally, we used the model to predict growth in the absence of individual amino acids. We allowed both uptake and secretion of all extracellular metabolites, disabled uptake of each amino acid separately, and maximized growth rate using FBA. Amino acids were classified as essential if they were required for growth and non-essential otherwise, and the predicted essentiality was compared to experimental data [21].

Analyzing oxygen-limited growth

We used parsimonious FBA (pFBA) [33] to find maximal growth rates and minimal flux distributions for 100 randomized conditions and 100 linearly spaced oxygen uptake rates in the range $r \in (0, r_{\max})$ where r is uptake rate and $r_{\max}$ is the minimal oxygen uptake rate at maximal growth. For each condition, we uniformly sampled random ratios (1–100) of nutrients in a minimal feed (essential amino acids and choline) that were normalized to 1 g gDW^{-1} h^{-1} and used as coefficients in a boundary reaction representing feed uptake. This normalization ensured that conditions were comparable, but the absolute value was arbitrary as only relative predictions were needed. We allowed unlimited uptake of phosphate and disabled all other uptakes as well as secretion of feed nutrients. To account for uncertainty in relative flux capacities, we uniformly sampled random bounds for all reactions (1–100 mmol gDW^{-1} h^{-1}) for each condition but kept the original reaction reversibilities. Reversible reactions were assigned the same random bound in both directions.

For each oxygen uptake rate, we computed mean growth rate with 95% confidence band from bootstrapping with 1,000 samples. We fitted the means to experimental data [28, 29, 30, 31] by assuming a simple piecewise linear relationship between water oxygen saturation (x) and relative oxygen uptake rate:

$$\frac{r}{r_{\max}} = \begin{cases} 0 & x \leq x_0 \\ \frac{x - x_0}{x_1 - x_0} & x \in (x_0, x_1) \\ 1 & x \geq x_1 \end{cases}, \tag{2}$$

where x_0 and x_1 are the oxygen saturations at which the relative growth rate is 0 and 1, respectively. We estimated x_0 and x_1 by least-squares fitting of

$$\frac{\mu}{\mu_{\max}} = f\left(\frac{r}{r_{\max}}\right), \tag{3}$$

where μ is growth rate, $\mu_{\max}$ is maximal growth rate when oxygen is not limiting, and f is a function that linearly interpolates the metabolic model predictions. We also fitted a logistic model with asymptotes -1 and 1,

$$\frac{\mu}{\mu_{\max}} = \frac{2}{1 + e^{k(x_0 - x)}} - 1, \tag{4}$$

where k is the logistic growth rate, and a Monod model,

$$\frac{\mu}{\mu_{\max}} = \frac{x - x_0}{K_{\mathrm{s}} + x - x_0}, \tag{5}$$

where $K_s + x_0$ is the saturation at which $\mu = \frac{1}{2}\mu_{\max}$.

To identify reaction contributions to oxygen-limited growth, we took the absolute value of the pFBA fluxes, normalized each flux by its maximum value within each condition, and used Ward's minimum variance method to cluster reactions by the resulting absolute relative fluxes. We mapped reactions from the top six clusters to genes and used g:Profiler [51] to identify enriched pathways from KEGG [11]. We used the genes in the model as background, considered pathways with adjusted $p \leq 0.05$ to be enriched, and discarded pathways outside the model's scope (xenobiotics).

Predicting growth-limiting amino acids in feeds

We obtained ratios of amino acids in three commercial feed ingredients: fish, soybean, and black soldier fly larvae meal [52] (**Table 1**). For each feed, these ratios were used as coefficients for amino acids in a boundary reaction representing feed consumption. Mass was divided equally between amino acids that were combined in the feed formulation (Asn/Asp and Gln/Glu). For each feed ingredient, we deactivated import of amino acids via other boundary reactions, fixed the growth rate to 1 h^{-1} (arbitrary, as we were interested in generated biomass relative to consumed feed), and minimized feed uptake flux (normalized to 1 mg gDW^{-1} h^{-1}). To simulate growth limitations from protein synthesis rather than energy generation, we also allowed unlimited uptake of glucose. We multiplied molecular mass with reduced cost in the optimal solution for each amino acid exchange reaction and identified the one with largest negative value as limiting [53]. To supplement the feed with the limiting amino acid, we set the bounds of its exchange reaction to only allow import, and we penalized supplementation by adding the exchange reaction to the objective with coefficient equal to molecular mass. We repeated the steps above until all limiting amino acids had been found for each feed.

Acknowledgements
We thank Ingunn Verne Ruud, Pedro Febrer Martínez, and Håvard Molversmyr for helpful contributions to the analyses and Snorre Sulheim for feedback on the manuscript.

Funding
This work was funded by the Research Council of Norway grant 248792 (DigiSal) with support from grant 248810 (Centre for Digital Life Norway).

Abbreviations
BiGG: Biochemically, Genetically, and Genomically structured. FBA: flux balance analysis. KEGG: Kyoto Encyclopedia of Genes and Genomes. gDW: grams dry weight biomass. GPR: gene-protein-reaction association. MEMOTE: metabolic model tests. pFBA: parsimonious flux balance analysis. SAPP: Semantic Annotation Platform with Provenance. SBML: Systems Biology Markup Language.

Availability of data and materials
The model and scripts that reproduce our results can be found at gitlab.com/digisal/salarecon.

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Authors' contributions
ABG, SRS, VAPMdS, and JOV designed the study. MZ and FR built the draft model. FR, MG, OØ, and JOV built the final model. FR and OØ performed simulations and analyzed the data. MZ, FR, OØ, and JOV wrote the manuscript. JCJvD, MSD, FG, RAH, WvH, JJK, and PJS performed genomics and bioinformatics analyses. YJ and LTM advised on salmon physiology. All authors read and approved the final manuscript.

Author details
[1]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences (NMBU), Ås, Norway. [2]Faculty of Biosciences, Norwegian University of Life Sciences (NMBU), Ås, Norway. [3]Laboratory of Systems and Synthetic Biology, Wageningen University & Research (WUR), Wageningen, The Netherlands.

References
1. The State of World Fisheries and Aquaculture: Meeting the Sustainable Development Goals. Technical report, Food and Agriculture Organization of the United Nations (2018)
2. Thodesen, J., Grisdale-Helland, B., Helland, S.J., Gjerde, B.: Feed intake, growth and feed utilization of offspring from wild and selected Atlantic salmon (Salmo salar). Aquaculture **180**(3-4), 237–246 (1999). doi:10.1016/S0044-8486(99)00204-5
3. Torstensen, B.E., Espe, M., Sanden, M., Stubhaug, I., Waagbø, R., Hemre, G., Fontanillas, R., Nordgarden, U., Hevrøy, E.M., Olsvik, P., Berntssen, M.H.G.: Novel production of Atlantic salmon (Salmo salar) protein based on combined replacement of fish meal and fish oil with plant meal and vegetable oil blends. Aquaculture **285**(1-4), 193–200 (2008). doi:10.1016/j.aquaculture.2008.08.025
4. Boissy, J., Aubin, J., Drissi, A., Werf, H.M.G.V.D., Bell, G.J., Kaushik, S.J.: Environmental impacts of plant-based salmonid diets at feed and farm scales. Aquaculture **321**(1-2), 61–70 (2011). doi:10.1016/j.aquaculture.2011.08.033
5. Jennings, S., Stentiford, G.D., Leocadio, A.M., Jeffery, K.R., Metcalfe, J.D., Katsiadaki, I., Auchterlonie, N.A., Mangi, S.C., Pinnegar, J.K., Ellis, T., Peeler, E.J., Luisetti, T., Baker-Austin, C., Brown, M., Catchpole, T.L., Clyne, F.J., Dye, S.R., Edmonds, N.J., Hyder, K., Lee, J., Lees, D.N., Morgan, O.C., O'Brien, C.M., Oidtmann, B., Posen, P.E., Santos, A.R., Taylor, N.G.H., Turner, A.D., Townhill, B.L., Verner-Jeffreys, D.W.: Aquatic food security: insights into challenges and solutions from an analysis of interactions between fisheries, aquaculture, food safety, human health, fish and human welfare, economy and environment. Fish and Fisheries **17**(4), 893–938 (2016). doi:10.1111/faf.12152
6. Aas, T.S., Ytrestøyl, T., Åsgård, T.: Utilization of feed resources in the production of Atlantic salmon (Salmo salar) in Norway: An update for 2016. Aquaculture Reports **15** (2019). doi:10.1016/j.aqrep.2019.100216
7. Orth, J.D., Thiele, I., Palsson, B.O.: What is flux balance analysis? Nature Biotechnology **28**(3), 245–248 (2010). doi:10.1038/nbt.1614. NIHMS150003
8. Kolodkin, A.N., Boogerd, F.C., Bruggeman, F.J., Westerhoff, H.V.: Modeling Approaches in Systems Biology, Including Silicon Cell Models. Systems Biology and Livestock Science, 31–51 (2011). doi:10.1002/9780470963012.ch2
9. Moretti, S., Tran, V.D.T., Mehl, F., Ibberson, M., Pagni, M.: MetaNetX/MNXref: Unified namespace for metabolites and biochemical reactions in the context of metabolic models. Nucleic Acids Research **49**(D1), 570–574 (2021). doi:10.1093/nar/gkaa992
10. Norsigian, C.J., Pusarla, N., McConn, J.L., Yurkovich, J.T., Dräger, A., Palsson, B.O., King, Z.: BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. Nucleic Acids Research **1**, 1–5 (2019). doi:10.1093/nar/gkz1054
11. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., Tanabe, M.: New approach for understanding genome variations in KEGG. Nucleic Acids Research **47**(D1), 590–595 (2019). doi:10.1093/nar/gky962
12. Thiele, I., Palsson, B.: A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols **5**(1), 93–121 (2010). doi:10.1038/nprot.2009.203
13. Mendoza, S.N., Olivier, B.G., Molenaar, D., Teusink, B.: A systematic assessment of current genome-scale metabolic reconstruction tools. Genome Biology **20**(1), 1–20 (2019). doi:10.1186/s13059-019-1769-1
14. Gu, C., Kim, G.B., Kim, W.J., Kim, H.U., Lee, S.Y.: Current status and applications of genome-scale metabolic models. Genome Biology **20**(1), 1–18 (2019). doi:10.1186/s13059-019-1730-3
15. Li, S., Pozhitkov, A., Ryan, R.A., Manning, C.S., Brown-Peterson, N., Brouwer, M.: Constructing a fish metabolic network model. Genome Biology **11**(11) (2010). doi:10.1186/gb-2010-11-11-r115
16. Bekaert, M.: Reconstruction of Danio rerio Metabolic Model Accounting for Subcellular Compartmentalisation. PLoS ONE **7**(11) (2012). doi:10.1371/journal.pone.0049903
17. van Steijn, L., Verbeek, F.J., Spaink, H.P., Merks, R.M.H.: Predicting Metabolism from Gene Expression in an Improved Whole-Genome Metabolic Network Model of Danio rerio . Zebrafish **16**(4), 348–362 (2019). doi:10.1089/zeb.2018.1712
18. Hanna, E.M., Zhang, X., Eide, M., Fallahi, S., Furmanek, T., Yadetie, F., Zielinski, D.C., Goksøyr, A., Jonassen, I.: ReCodLiver0.9: Overcoming Challenges in Genome-Scale Metabolic Reconstruction of a Non-model Species. Frontiers in Molecular Biosciences **7**(591406), 1–10 (2020). doi:10.3389/fmolb.2020.591406
19. Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Kent, M.P., Nome, T., Hvidsten, T.R., Leong, J.S., Minkley, D.R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A., Walenz, B., Hermansen, R.A., Von Schalburg, K., Rondeau, E.B., Di Genova, A., Samy, J.K.A., Olav Vik, J., Vigeland, M.D., Caler, L., Grimholt, U., Jentoft, S., Inge Våge, D., De Jong, P., Moen, T., Baranski, M., Palti, Y., Smith, D.R., Yorke, J.A., Nederbragt, A.J., Tooming-Klunderud, A., Jakobsen, K.S., Jiang, X., Fan, D., Hu, Y.,

Liberles, D.A., Vidal, R., Iturra, P., Jones, S.J.M., Jonassen, I., Maass, A., Omholt, S.W., Davidson, W.S.: The Atlantic salmon genome provides insights into rediploidization. Nature **533**(7602), 200–205 (2016). doi:10.1038/nature17164

20. Broido, A.D., Clauset, A.: Scale-free networks are rare. Nature Communications **10**(1), 1–10 (2019). doi:10.1038/s41467-019-08746-5. 1801.03400
21. Hatlen, B., Jakobsen, J.V., Crampton, V., Alm, M., Langmyhr, E., Espe, M., Hevrøy, E.M., Torstensen, B.E., Liland, N., Waagbø, R.: Growth, feed utilization and endocrine responses in Atlantic salmon (Salmo salar) fed diets added poultry by-product meal and blood meal in combination with poultry oil. Aquaculture Nutrition **21**(5), 714–725 (2015). doi:10.1111/anu.12194
22. Sigurdsson, M.I., Jamshidi, N., Steingrimsson, E., Thiele, I., Palsson, B.T.: A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. BMC Systems Biology **4** (2010). doi:10.1186/1752-0509-4-140
23. Hefzi, H., Ang, K.S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C.A., Baycin-Hizal, D., Huang, Y., Ley, D., Martinez, V.S., Kyriakopoulos, S., Jiménez, N.E., Zielinski, D.C., Quek, L.E., Wulff, T., Arnsdorf, J., Li, S., Lee, J.S., Paglia, G., Loira, N., Spahn, P.N., Pedersen, L.E., Gutierrez, J.M., King, Z.A., Lund, A.M., Nagarajan, H., Thomas, A., Abdel-Haleem, A.M., Zanghellini, J., Kildegaard, H.F., Voldborg, B.G., Gerdtzen, Z.P., Betenbaugh, M.J., Palsson, B.O., Andersen, M.R., Nielsen, L.K., Borth, N., Lee, D.Y., Lewis, N.E.: A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. Cell Systems **3**(5), 434–4438 (2016). doi:10.1016/j.cels.2016.10.020
24. Brunk, E., Sahoo, S., Zielinski, D.C., Altunkaya, A., Dräger, A., Mih, N., Gatto, F., Nilsson, A., Preciat Gonzalez, G.A., Aurich, M.K., Prlic, A., Sastry, A., Danielsdottir, A.D., Heinken, A., Noronha, A., Rose, P.W., Burley, S.K., Fleming, R.M.T., Nielsen, J., Thiele, I., Palsson, B.O.: Recon3D enables a three-dimensional view of gene variation in human metabolism. Nature Biotechnology **36**(3), 272–281 (2018). doi:10.1038/nbt.4072
25. Singh, D., Carlson, R., Fell, D., Poolman, M.: Modelling metabolism of the diatom Phaeodactylum tricornutum. Biochemical Society Transactions **43**(6), 1182–1186 (2015). doi:10.1042/bst20150152
26. Lieven, C., Beber, M.E., Olivier, B.G., Bergmann, F.T., Ataman, M., Babaei, P., Bartell, J.A., Blank, L.M., Chauhan, S., Correia, K., Diener, C., Dräger, A., Ebert, B.E., Edirisinghe, J.N., Faria, J.P., Feist, A.M., Fengos, G., Fleming, R.M.T., García-Jiménez, B., Hatzimanikatis, V., van Helvoirt, W., Henry, C.S., Hermjakob, H., Herrgård, M.J., Kaafarani, A., Kim, H.U., King, Z., Klamt, S., Klipp, E., Koehorst, J.J., König, M., Lakshmanan, M., Lee, D.Y., Lee, S.Y., Lee, S., Lewis, N.E., Liu, F., Ma, H., Machado, D., Mahadevan, R., Maia, P., Mardinoglu, A., Medlock, G.L., Monk, J.M., Nielsen, J., Nielsen, L.K., Nogales, J., Nookaew, I., Palsson, B.O., Papin, J.A., Patil, K.R., Poolman, M., Price, N.D., Resendis-Antonio, O., Richelle, A., Rocha, I., Sánchez, B.J., Schaap, P.J., Malik Sheriff, R.S., Shoaie, S., Sonnenschein, N., Teusink, B., Vilaça, P., Vik, J.O., Wodke, J.A.H., Xavier, J.C., Yuan, Q., Zakhartsev, M., Zhang, C.: MEMOTE for standardized genome-scale metabolic model testing. Nature Biotechnology **38**(3), 272–276 (2020). doi:10.1038/s41587-020-0446-y
27. Richelle, A., Kellman, B.P., Wenzel, A.T., Chiang, A.W.T., Reagan, T., Gutierrez, J.M., Joshi, C., Li, S., Liu, J.K., Masson, H., Lee, J., Li, Z., Heirendt, L., Trefois, C., Juarez, E.F., Bath, T., Borland, D., Mesirov, J.P., Robasky, K., Lewis, N.E.: What does your cell really do? Model-based assessment of mammalian cells metabolic functionalities using omics data. bioRxiv, 2020–0426057943 (2020). doi:10.1101/2020.04.26.057943
28. Berg, A., Danielsberg, A.: Oxygen level and fish density experiments from hemnskjel june 1989–march 1990. Technical report, Norwegian Hydrotechnical Laboratory (1993)
29. Bergheim, A., Gausen, M., Næss, A., Fjermedal, A.M., Hølland, P.M., Molversmyr, Å.: Effects of oxygen deficit on post-smolt salmon. trial i. Technical report, Rogaland Research (2002)
30. Hosfeld, C.D., Engevik, A., Mollan, T., Lunde, T.M., Waagbø, R., Olsen, A.B., Breck, O., Stefansson, S., Fivelstad, S.: Long-term separate and combined effects of environmental hypercapnia and hyperoxia in Atlantic salmon (Salmo salar L.) smolts. Aquaculture **280**(1-4), 146–153 (2008). doi:10.1016/j.aquaculture.2008.05.009
31. Hosfeld, C.D., Handeland, S.O., Fivelstad, S., Stefansson, S.O.: Physiological effects of normbaric environmental hyperoxia on Atlantic salmon (Salmo salar L.) presmolts. Aquaculture **308**(1-2), 28–33 (2010). doi:10.1016/j.aquaculture.2010.08.003
32. Thorarensen, H., Farrell, A.P.: The biological requirements for post-smolt Atlantic salmon in closed-containment systems. Aquaculture **312**(1-4), 1–14 (2011). doi:10.1016/j.aquaculture.2010.11.043
33. Lewis, N.E., Hixson, K.K., Conrad, T.M., Lerman, J.A., Charusanti, P., Polpitiya, A.D., Adkins, J.N., Schramm, G., Purvine, S.O., Lopez-Ferrer, D., Weitz, K.K., Eils, R., König, R., Smith, R.D., Palsson, B.: Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Molecular Systems Biology **6**(390) (2010). doi:10.1038/msb.2010.47
34. Kolstad, K., Grisdale-Helland, B., Gjerde, B.: Family differences in feed efficiency in Atlantic salmon (Salmo salar). Aquaculture **241**, 169–177 (2004). doi:10.1016/j.aquaculture.2004.09.001
35. 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. Genetics Selection Evolution **51**(1), 1–14 (2019). doi:10.1186/s12711-019-0455-9
36. Atlantic Salmon - Nutritional Requirements. `https://www.fao.org/fishery/affris/species-profiles/atlantic-salmon/nutritional-requirements/en/` Accessed 2021-03-26
37. Li, P., Mai, K., Trushenski, J., Wu, G.: New developments in fish amino acid nutrition: Towards functional and environmentally oriented aquafeeds. Amino Acids **37**, 43–53 (2009). doi:10.1007/s00726-008-0171-1
38. Macqueen, D.J., Primmer, C.R., Houston, R.D., Nowak, B.F., Bernatchez, L., Bergseth, S., Davidson, W.S., Gallardo-Escárate, C., Goldammer, T., Guiguen, Y., Iturra, P., Kijas, J.W., Koop, B.F., Lien, S., Maass, A., Martin, S.A.M., McGinnity, P., Montecino, M., Naish, K.A., Nichols, K.M., Ólafsson, K.,

Omholt, S.W., Palti, Y., Plastow, G.S., Rexroad, C.E., Rise, M.L., Ritchie, R.J., Sandve, S.R., Schulte, P.M., Tello, A., Vidal, R., Vik, J.O., Wargelius, A., Yáñez, J.M.: Functional Annotation of All Salmonid Genomes (FAASG): An international initiative supporting future salmonid research, conservation and aquaculture. BMC Genomics **18**(1), 1–9 (2017). doi:10.1186/s12864-017-3862-8

39. Machado, D., Andrejev, S., Tramontano, M., Patil, K.R.: Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Research **46**(15), 7542–7553 (2018). doi:10.1093/nar/gky537
40. Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., Daníelsdóttir, A.D., Krecke, M., Merten, D., Haraldsdóttir, H.S., Heinken, A., Heirendt, L., Magnúsdóttir, S., Ravcheev, D.A., Sahoo, S., Gawron, P., Friscioni, L., Garcia, B., Prendergast, M., Puente, A., Rodrigues, M., Roy, A., Rouquaya, M., Wiltgen, L., Žagare, A., John, E., Krueger, M., Kuperstein, I., Zinovyev, A., Schneider, R., Fleming, R.M.T., Thiele, I.: The Virtual Metabolic Human database: Integrating human and gut microbiome metabolism with nutrition and disease. Nucleic Acids Research **47**(D1), 614–624 (2019). doi:10.1093/nar/gky992
41. Bergheim, a., Gausen, M., Næss, a., Fjermedal, A., Holland, P.: Effects of Oxygen Deficit on Post-smolt Salmon. Trial I, p. 25 (2002)
42. Thiele, I., Sahoo, S., Heinken, A., Hertel, J., Heirendt, L., Aurich, M.K., Fleming, R.M.: Personalized whole-body models integrate metabolism, physiology, and the gut microbiome. Molecular Systems Biology **16**(5), 1–24 (2020). doi:10.15252/msb.20198982
43. Bar, N.S., Sigholt, T., Shearer, K.D., Krogdahl, Å.: A dynamic model of nutrient pathways, growth, and body composition in fish. Canadian Journal of Fisheries and Aquatic Sciences **29**(12), 1669–1682 (2008). doi:10.1139/F07-127
44. Deslauriers, D., Chipps, S.R., Breck, J.E., Rice, J.A., Madenjian, C.P.: Fish Bioenergetics 4.0: An R-Based Modeling Application. Fisheries **42**(11), 586–596 (2017). doi:10.1080/03632415.2017.1377558
45. Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., Nakai, K.: WoLF PSORT: Protein localization predictor. Nucleic Acids Research **35**(SUPPL.2), 585–587 (2007). doi:10.1093/nar/gkm259
46. Koehorst, J.J., Van Dam, J.C.J., Saccenti, E., Martins Dos Santos, V.A.P., Suarez-Diez, M., Schaap, P.J.: SAPP: Functional genome annotation and analysis through a semantic framework using FAIR principles. Bioinformatics **34**(8), 1401–1403 (2018). doi:10.1093/bioinformatics/btx767
47. Ebrahim, A., Lerman, J.A., Palsson, B.O., Hyduke, D.R.: COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Systems Biology **7**(74), 10–118617520509774 (2013). doi:10.1186/1752-0509-7-74
48. Bateman, A.: UniProt: A worldwide hub of protein knowledge. Nucleic Acids Research **47**(D1), 506–515 (2019). doi:10.1093/nar/gky1049
49. Karlsen, E., Schulz, C., Almaas, E.: Automated generation of genome-scale metabolic draft reconstructions based on KEGG. BMC Bioinformatics **19**(1), 467 (2018). doi:10.1186/s12859-018-2472-z
50. Keating, S.M., Waltemath, D., König, M., Zhang, F., Dräger, A., Chaouiya, C., Bergmann, F.T., Finney, A., Gillespie, C.S., Helikar, T., Hoops, S., Malik-Sheriff, R.S., Moodie, S.L., Moraru, I.I., Myers, C.J., Naldi, A., Olivier, B.G., Sahle, S., Schaff, J.C., Smith, L.P., Swat, M.J., Thieffry, D., Watanabe, L., Wilkinson, D.J., Blinov, M.L., Begley, K., Faeder, J.R., Gómez, H.F., Hamm, T.M., Inagaki, Y., Liebermeister, W., Lister, A.L., Lucio, D., Mjolsness, E., Proctor, C.J., Raman, K., Rodriguez, N., Shaffer, C.A., Shapiro, B.E., Stelling, J., Swainston, N., Tanimura, N., Wagner, J., Meier-Schellersheim, M., Sauro, H.M., Palsson, B., Bolouri, H., Kitano, H., Funahashi, A., Hermjakob, H., Doyle, J.C., Hucka, M., Adams, R.R., Allen, N.A., Angermann, B.R., Antoniotti, M., Bader, G.D., Červený, J., Courtot, M., Cox, C.D., Dalle Pezze, P., Demir, E., Denney, W.S., Dharuri, H., Dorier, J., Drasdo, D., Ebrahim, A., Eichner, J., Elf, J., Endler, L., Evelo, C.T., Flamm, C., Fleming, R.M., Fröhlich, M., Glont, M., Gonçalves, E., Golebiewski, M., Grabski, H., Gutteridge, A., Hachmeister, D., Harris, L.A., Heavner, B.D., Henkel, R., Hlavacek, W.S., Hu, B., Hyduke, D.R., Jong, H., Juty, N., Karp, P.D., Karr, J.R., Kell, D.B., Keller, R., Kiselev, I., Klamt, S., Klipp, E., Knüpfer, C., Kolpakov, F., Krause, F., Kutmon, M., Laibe, C., Lawless, C., Li, L., Loew, L.M., Machne, R., Matsuoka, Y., Mendes, P., Mi, H., Mittag, F., Monteiro, P.T., Natarajan, K.N., Nielsen, P.M., Nguyen, T., Palmisano, A., Pettit, J., Pfau, T., Phair, R.D., Radivoyevitch, T., Rohwer, J.M., Ruebenacker, O.A., Saez-Rodriguez, J., Scharm, M., Schmidt, H., Schreiber, F., Schubert, M., Schulte, R., Sealfon, S.C., Smallbone, K., Soliman, S., Stefan, M.I., Sullivan, D.P., Takahashi, K., Teusink, B., Tolnay, D., Vazirabad, I., Kamp, A., Wittig, U., Wrzodek, C., Wrzodek, F., Xenarios, I., Zhukova, A., Zucker, J.: SBML Level 3: an extensible format for the exchange and reuse of biological models . Molecular Systems Biology **16**(8), 1–21 (2020). doi:10.15252/msb.20199110
51. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., Vilo, J.: g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Research **47**(W1), 191–198 (2019). doi:10.1093/nar/gkz369
52. INRAE-CIRAD-AFZ Feed Tables. `https://www.feedtables.com/content/table-dry-matter` Accessed 2021-03-26
53. Chen, Y., Pelt-KleinJan, E., Olst, B., Douwenga, S., Boeren, S., Bachmann, H., Molenaar, D., Nielsen, J., Teusink, B.: Proteome constraints reveal targets for improving microbial fitness in nutrient-rich environments. Molecular Systems Biology **17**(4), 1–13 (2021). doi:10.15252/msb.202010093

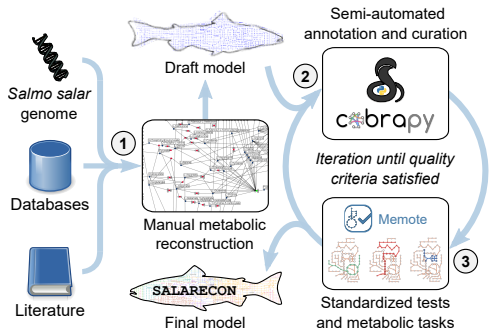

Figure 1 Model construction. SALARECON was built from the annotated Atlantic salmon genome, metabolic reaction and model databases, and literature. The procedure involved (1) manual metabolic network reconstruction using Insilico Discovery (Insilico Biotechnology, Stuttgart, Germany), (2) semi-automated annotation and curation using COBRApy [47], and (3) quality evaluation using the standardized metabolic model testing tool MEMOTE [26] and metabolic tasks [27]. Steps 2 and 3 were iterated until quality criteria were satisfied. Illustration of metabolic tasks from Richelle et al. [27].

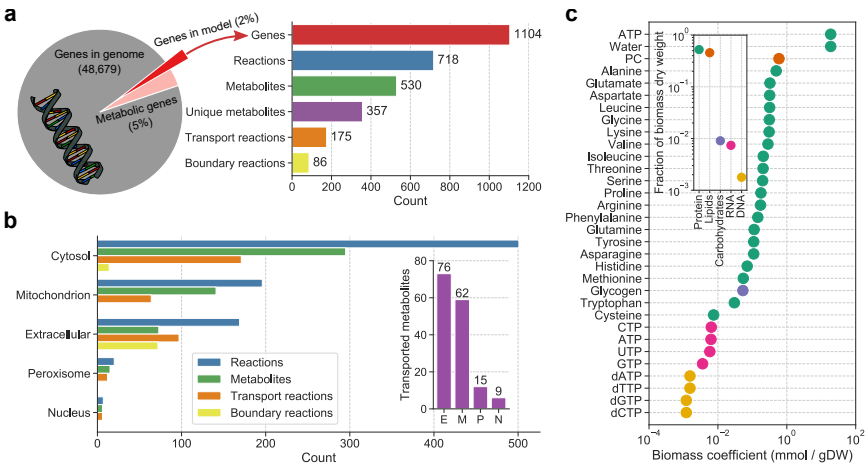

Figure 2 Model contents. (**a**) SALARECON contains 1,104 genes (2% of all genes and 48% of Atlantic salmon genes mapped to reactions in KEGG [11]), 718 reactions (175 transporting metabolites between compartments and 86 exchanging metabolites with the extracellular environment), and 530 metabolites (357 when metabolites occuring in multiple compartments are only counted once). (**b**) Metabolites and reactions are divided between five compartments (mitochondrion includes the inner mitochondrial membrane). Transport reactions are counted multiple times (once for each compartment of exhanged metabolites). Boundary reactions in cytosol are sink or demand reactions [12]. The inset shows how many unique metabolites can be transported between the cytosol and the other compartments (indicated by their initials). (**c**) Biomass composition of Atlantic salmon estimated from measured whole-body composition [21]. The inset summarizes each class of macromolecules. Carbohydrates and lipids are represented by glycogen and phosphatidylcholine (PC), respectively. ATP serves both as energy for protein synthesis and as a building block in RNA synthesis.

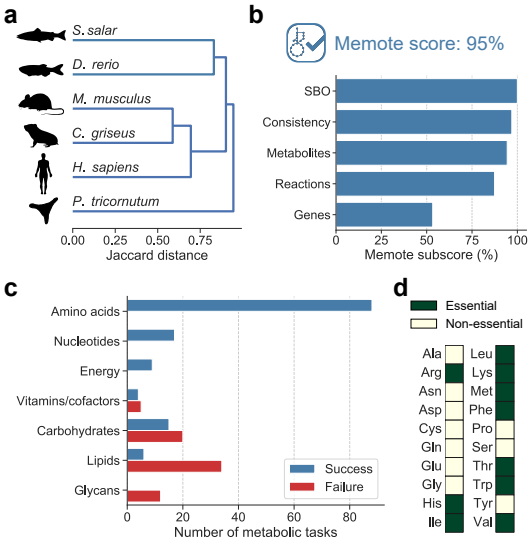

Figure 3 Model quality evaluation. (**a**) Clustering of SALARECON and metabolic models of other multicellular eukaryotes based on Jaccard distance between reaction contents. Atlantic salmon (*Salmo salar*) is closer to zebrafish (*Danio rerio*) [17] than mouse (*Mus musculus*) [22], chinese hamster ovary (CHO, *Cricetulus griseus*) [23], human (*Homo sapiens*) [24], and the diatom *Phaeodactylum tricornutum* [25]. (**b**) Model score and subscores from MEMOTE [26]. Subscores evaluate Systems Biology Ontology (SBO) annotation, model consistency, and database mappings for metabolites, reactions, and genes. (**c**) Ability of SALARECON to perform metabolic tasks [27]. Tasks are grouped by metabolic system and classified as successful if model predictions reflected expected metabolic (in)capabilities. (**d**) Essential amino acids predicted by SALARECON match data [21].

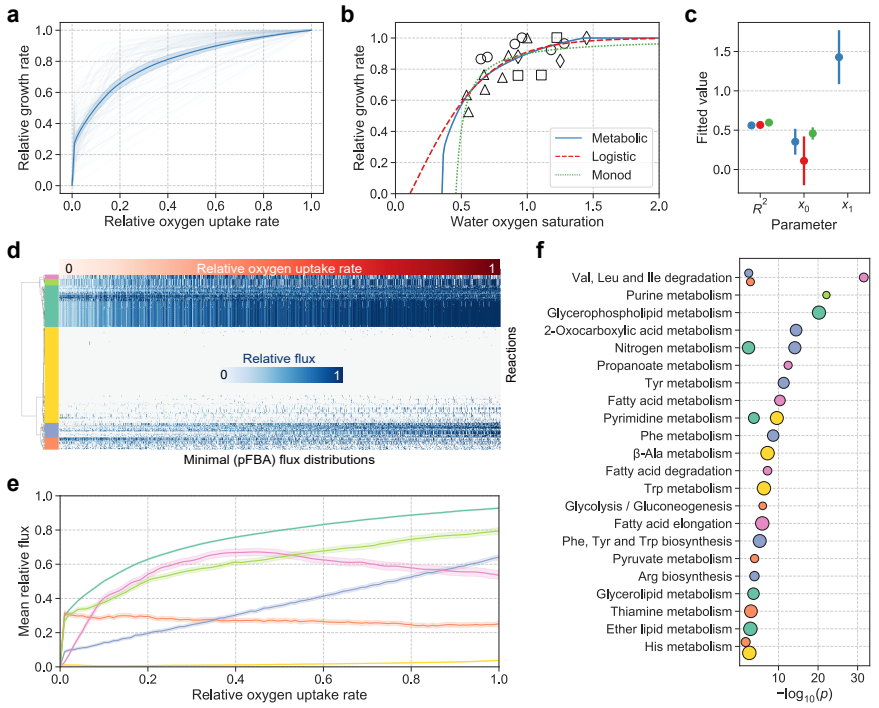

Figure 4 Oxygen-limited growth analysis. (**a**) SALARECON predictions of relative growth rate under oxygen limitation as a function of relative oxygen uptake rate. Feed composition and flux capacities were randomized 100 times (light blue) and the mean across conditions is shown with 95% confidence band from bootstrapping with 1,000 samples (dark blue). (**b**) Metabolic, logistic, and Monod model fits to experimental data from Berg and Danielsberg [28] (circles), Bergheim et al. [29] (triangles), Hosfeld et al. [30] (squares), and Hosfeld et al. [31] (diamonds). SALARECON predictions were fitted by assuming a linear relationship between relative oxygen uptake rate and water oxygen saturation. (**c**) Coefficient of determination (R^2), minimal oxygen saturation required for growth (x_0), and minimal oxygen saturation required for *maximal* growth (x_1) from fitted models with same colors as in **b**. Error bars indicate two standard errors of the estimates. (**d**) Minimal flux distributions for metabolic model predictions shown in **a** from parsimonious flux balance analysis (pFBA) [33]. Rows are reactions, columns are flux distributions sorted by relative oxygen uptake rate, and each cell shows absolute flux normalized by maximum value for each condition. Rows are clustered by Ward's minimum variance method and divided into six clusters indicated by colors. (**e**) Mean absolute relative flux with 95% confidence bands from bootstrapping with 1,000 samples for the six clusters with same colors as in **d**. (**f**) Enrichment of metabolic pathways from KEGG [11] for the six clusters with same colors as in **d** and **e** and size reflecting the fraction of genes in each pathway that are found in a cluster.

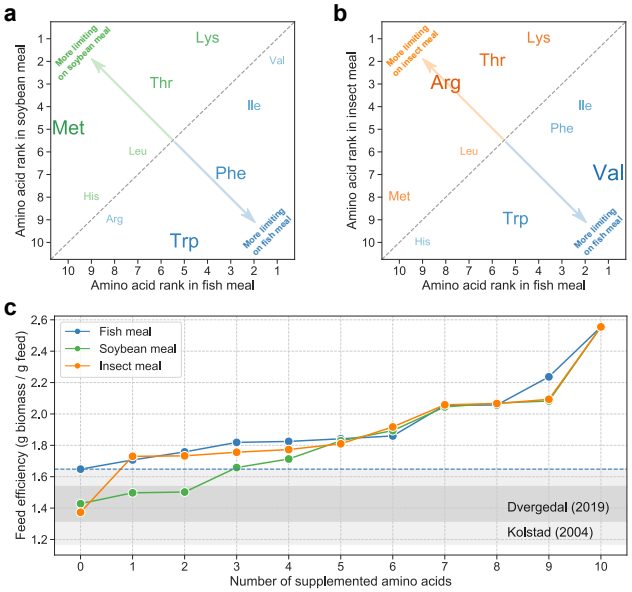

Figure 5 Growth-limiting amino acids in commercial feed ingredients. (**a**) Order of amino acid limitations in feed ingredients based on soybean and fish meal. Amino acids that are closer to the top left and bottom right corners were more limiting in soybean meal and fish meal, respectively, as indicated by size. (**b**) Order of amino acid limitations in feed ingredients based on insect and fish meal. Amino acids that are closer to the top left and bottom right corners were more limiting in insect meal and fish meal, respectively, as indicated by size and color. (**c**) Feed efficiency after successive supplementation of the most limiting amino acid for fish, soybean, and insect meal. The baseline feed efficiency of fish meal is indicated by a dashed blue line, and ranges observed by Kolstad et al. [34] and Dvergedal et al. [35] are highlighted in gray.

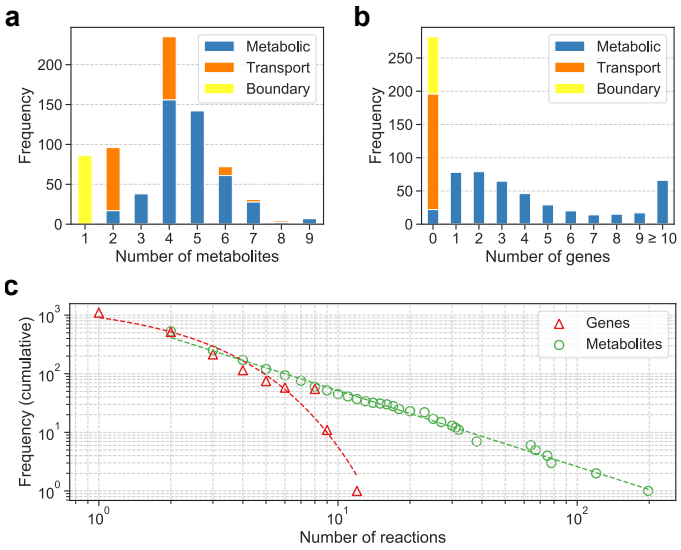

Figure S1 Model degree distributions. (**a**) Distribution of number of metabolites converted by reactions. Boundary reactions exchange one metabolite with the extracellular environment and transport reactions usually exchange an even number of metabolites between compartments. (**b**) Distribution of number of genes associated with reactions. Transport and boundary reactions lack annotation and are not associated with any genes. Most metabolic reactions (95%) are associated with one or more genes. (**c**) Cumulative distribution of number of reactions associated with genes and metabolites (number of genes or metabolites associated with k or more reactions for all k). Most genes and metabolites are associated with a few reactions but some metabolites are highly connected hubs. Exponential and power law fits are shown for genes and metabolites, respectively.

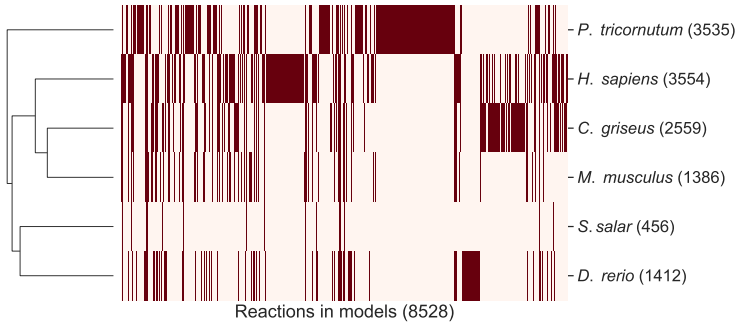

Figure S2 Reaction contents of models of multicellular eukaryotes. Clustered heatmap of reaction contents of metabolic models of multicellular eukaryotes. Each row is an organism, each column is a reaction, and a dark cell indicates a reaction that is found in the model of that organism. Rows are clustered by Jaccard distance with number of non-boundary reactions given.

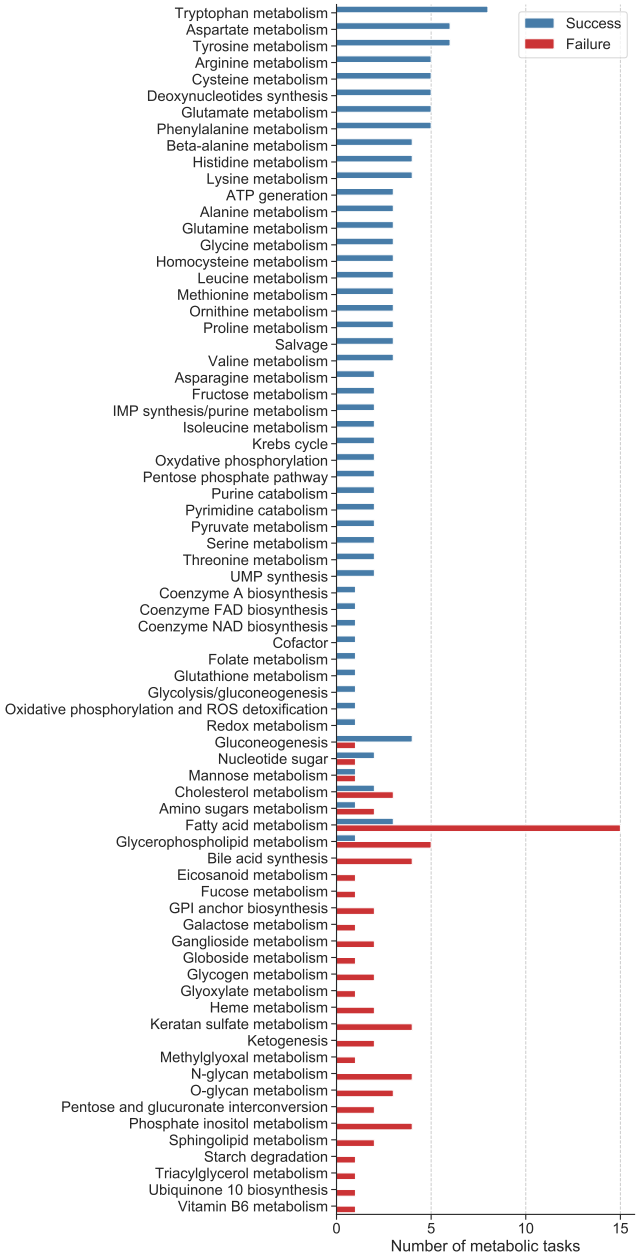

Figure S3 Metabolic task results by subsystem. Ability of SALARECON to perform metabolic tasks [27]. Tasks are grouped by metabolic subsystem and classified as successful if model predictions reflected expected metabolic (in)capabilities.

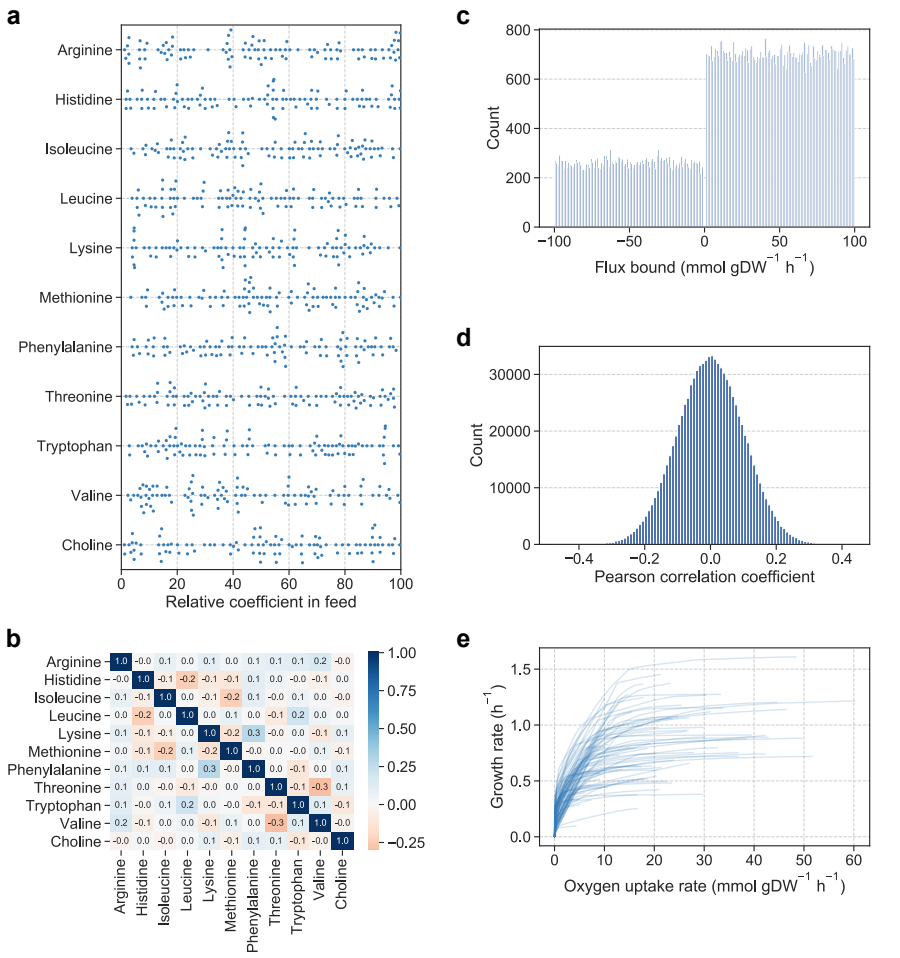

Figure S4 Oxygen-limited growth analysis. (**a**) Uniformy randomly sampled feed coefficients of amino acids and choline in conditions used to predict oxygen-limited growth. (**b**) Pairwise Pearson correlations between metabolites of feed coefficents shown in **a**. (**c**) Uniformly randomly sampled flux bounds in conditions used to predict oxygen-limited growth (100 samples). (**d**) Pairwise Pearson correlations between reactions of flux bounds shown in **a**. (**e**) Predicted absolute growth rates as a function of absolute oxygen uptake rates for the 100 uniformly randomly sampled conditions.

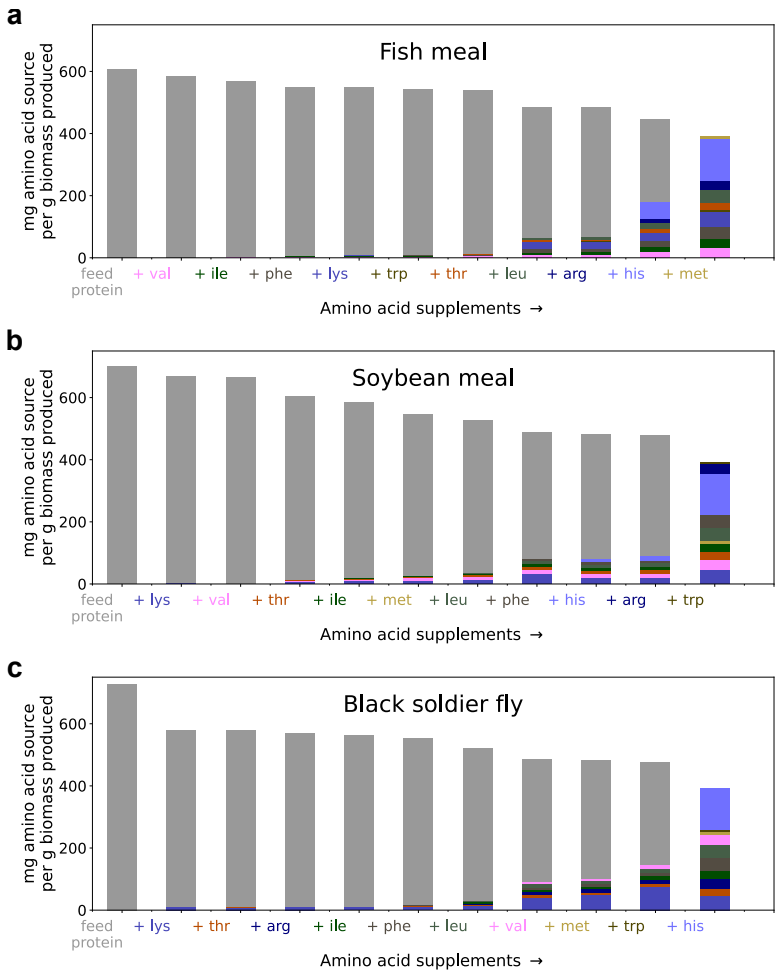

Figure S5 Growth-limiting amino acids in commercial feed ingredients. Feed efficiency as a function of number of supplemented amino acids, measured in mg feed ingredient and supplemented amino acids consumed / gDW biomass produced for (**a**) fish meal, (**b**) soybean meal, and (**c**) black soldier fly larvae meal. Amino acids are indicated by color and ordered from most limiting (left) to least limiting (right). Each bar represents the fed amount of amino acid sources, with one amino acid supplemented per step towards the right. Limiting amino acids were supplemented until all feed protein had been replaced.

Table 1 Amino acid compositions of feed ingredients. Mass percentage of each amino acid relative to total mass of amino acids in feed ingredients used in simulations [52].

Amino acid	Fish meal	Soybean meal	Insect meal
Ala	6.82	4.43	7.05
Arg	7.19	7.54	5.34
Asn/Asp	10.02	11.87	10.07
Cys	0.93	1.74	0.62
Gln/Glu	13.98	18.74	11.12
Gly	6.88	4.19	6.67
His	2.62	2.69	3.32
Ile	4.64	4.61	4.86
Leu	7.91	8.02	7.76
Lys	8.31	6.44	6.19
Met	3.07	1.45	2.06
Phe	4.29	5.22	4.31
Pro	4.45	5.08	6.39
Ser	4.29	4.13	4.71
Thr	4.57	3.67	4.29
Trp	1.13	1.58	1.61
Tyr	3.40	3.60	6.85
Val	5.48	5.00	6.79