# SALARECON connects the Atlantic salmon genome to growth and feed efficiency

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

[1]Norwegian University of Life Sciences (NMBU), Ås, Norway. [2]Wageningen University & Research (WUR), Wageningen, Netherlands.

*These authors contributed equally to this work. †Correspondence: jon.vik@nmbu.no.

## Abstract

**Background** Atlantic salmon (*Salmo salar*) is the most valuable farmed fish globally and there is much interest in optimizing its genetics and conditions for growth and feed efficiency. Also, 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 is what converts nutrients in the feed to energy and biomass, but they 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 reflects expected metabolic (in)capabilities. We show that it can explain observed growth under hypoxia in terms of metabolic fluxes and apply it to aquaculture by simulating growth with commercial feed ingredients. Predicted feed efficiencies and limiting amino acids 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 from the genome. It can explain Atlantic salmon physiology and address key challenges in aquaculture.

## 1. Background

Salmonid aquaculture has grown in volume and economic importance over the past few 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 increased both growth rate and feed efficiency[2].

Growing demand for feed and insufficient marine resources has led to a switch to plant ingredients, reducing production costs and exploitation of fish stocks[3]. However, salmon are not adapted to eating plants, and current plant-based feeds have a negative impact on fish health and environment[4,5].

Plant-based feeds are complex, the ingredient market is fluctuating, and feeding trials are demanding. Thus, finding feeds that minimize cost and environmental impact while providing necessary nutrients to the fish is a key challenge[6].

Metabolic networks convert available nutrients to the energy and building blocks required for growth, and this conversion happens through the enzymatic reactions encoded by the genome. This provides a framework for connecting salmon genomics to metabolism and addressing challenges such as development of novel sustainable feeds[7].

Large databases of metabolic reactions and models[8–10] and methods for metabolic network reconstruction from genome sequences[11,12] have made metabolic models available for organisms ranging from microbes to humans[13]. However, there are hardly any such models of fish available[14–17] and none of salmon or other important aquaculture species.

Here, we present SALARECON: a metabolic model built from the Atlantic salmon genome[18] that predicts metabolic fluxes and growth through flux balance analysis (FBA)[19]. The quality of the model has been evaluated by standardized tests and captures expected metabolic (in)capabilities such as amino acid essentiality. Using growth under oxygen limitation as an example, we show that model predictions can explain salmon physiology in terms of metabolic fluxes that are tied to the genome, and we demonstrate applicaton to aquaculture by predicting growth-limiting amino acids in commercial feed ingredients in agreement with data.

## 2. Results

We built a metabolic model of Atlantic salmon (SALARECON) from its genome[18], metabolic reaction and model databases, and literature (**Fig. 1**). It covers 1,104 genes (2% of all genes and 48% of Atlantic salmon genes mapped to reactions in KEGG[10]), 718 reactions, and 530 metabolites (**Fig. 2a** and **Fig. S1**) divided between five compartments that are connected by transport reactions (**Fig. 2b**). A biomass reaction based on whole-body composition[20] ties growth rate to metabolic fluxes and the genome (**Fig. 2c**).

Comparing our Atlantic salmon model to models of other multicellular eukaryotes bases on presence and absence of reactions, we found that it was closer to zebrafish than to two mammals and a diatom (**Fig. 3a** and **Fig. S2**). It also performed well in standardized Memote tests[21] (**Fig. 3b**) and metabolic tasks[22] (**Fig. 3c** and **Fig. S3**). Notably, predicted growth in the absence of individual amino acids agrees with observed amino acid essentiality[20] (**Fig. 3d**).

We also used SALARECON to predict oxygen-limited growth rates on a minimal feed, using 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, we fitted our predictions to data[23,24] along with a logistic and a Monod model (**Fig. 4b**). All the models fitted the data well, the metabolic and logistic models gave similar estimates of minimal oxygen saturation required for growth, and the metabolic model also allowed estimation of minimal oxygen saturation required for *maximal* growth (**Fig. 4c**). The metabolic and logistic fits agree well with the expected relationship presented by Thorarensen et al.[25].

In contrast to the simple growth models, the metabolic model is mechanistic and allows predictions to be explained in terms of metabolic fluxes (**Fig. 4d**). Across randomly sampled conditions and levels of oxygen limitation, we identified five clusters of reactions with distinct contributions to oxygen-limited growth (**Fig. 4e**). These clusters could also be connected to the genome because SALARECON includes mappings between reactions and genes, allowing us to identify enriched pathways in each cluster (**Fig. 4f**).

We compared limiting amino acids in a fish meal feed to feeds

based on soybean and insect meal (**Table S1**). 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**). According to SALARECON, 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**). Model predictions agree well with expected baseline feed efficiencies[26,27] as well as reports that lysine, methionine, threonine, and arginine are more limiting in plant-based than in marine feeds[28,29].

## 3. 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 knowledge and experimental data[7]. Thus, SALARECON forms a meeting place 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. A metabolic model provides a sharper lens through which to interpret omics data by requiring consistency with flux balances and other known constraints, e.g. from thermodynamics or enzyme usage. 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[11,12]. However, semi-automated annotation and curation combined with automated Memote tests[21] and metabolic tasks[22] allowed faster iteration, and future reconstructions of related species[30] can benefit from our efforts by using SALARECON as a template.

Our work underscores the importance of 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. They 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 salmon genes mapped to reactions in KEGG[10], 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.

A key strength of the model is its extensive annotation of genes, metabolites, and reactions, which facilitates use with existing models, tools, and data. In particular, identifiers from BiGG[9] make it easy to compare and combine SALARECON with state-of-the-art models[31,32], e.g. to predict interactions between Atlantic salmon and its gut microbiota. It also allows direct application of many implemented methods such as evaluation of metabolic tasks[22].

The biomass reaction makes SALARECON a more realistic representation of salmon metabolism than a simple network reconstruction[11]. It enables prediction of growth and related fluxes and is based on organism-specific data[20] rather than copied from a human model as has been done for other eukaryotes such as mouse[33] or even zebrafish[15,16]. As demonstrated for Atlantic cod[17], 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.

By integrating model construction with quality evaluation, we were able to reach a final model that performs very well according to all of our metrics. SALARECON is more similar to the latest zebrafish model[16] than models of other multicellular eukaryotes[34–36], achieves a Memote score of 95%, which is better than all manually curated models in BiGG[9], 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[20]. The Memote score for gene annotation is low compared to reactions and metabolites because salmon genes can be mapped to fewer databases than generic biochemical components.

Our analysis of growth under oxygen limitation shows that phenotypes predicted by SALARECON can be fitted to data, and it produced mechanistic explanations of hypoxic metabolism and growth with implications for fish welfare and productivity in aquaculture. Growth predictions depend on environmental conditions and flux capacities, but SALARECON can be used to account for such uncertainty through random sampling. Average growth predictions fit the available data as well as simple growth models and allowed accurate estimation of critical water oxygen saturations. The predicted metabolic flux distributions could be divided into clusters of reactions with distinct metabolic pathway enrichments and contributions to hypoxic growth.

Predictions contrasting growth-limiting amino acids in three commercial feed ingredients agree well with observations[28,29] and show that SALARECON can be used to evaluate efficiency of sustainable feeds, a key challenge for modern aquaculture. The model predicts baseline feed efficiencies that lie within reported ranges[26,27] and suggests that the baseline 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 potential new feeds, 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[22]. By coupling tissue-specific models to each other and to gut microbiota models, we can make detailed and partially dynamic whole-body models[37]. This would be a major leap from available dynamic models[38] and provide a mechanistic alternative to state-of-the-art bioenergetics models[39], opening up new possibilities for understanding fish physiology and rational engineering of feeds, conditions, and genetics.

## 4. 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. Future work will expand SALARECON and integrate it with omics data to make tissue-specific and partially dynamic whole-body models. SALARECON will 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.

## 5. Methods

### Building the metabolic model

First, we manually built a draft model of Atlantic salmon core metabolism using the genome[18] with KEGG[10] annotations and the software Insilico Discovery (Insilico Biotechnology, Stuttgart, Germany). We used WoLF PSORT[40] and SAPP[41] to assign metabolites and reactions to six different compartments (cytosol, extracellular, mitochondrion, inner mitochondrial membrane, peroxisome, and nucleus). Exchange reactions were added to allow metabolite import (negative flux) and export (positive flux).

Second, we used COBRApy[42] to annotate and curate the draft model. We semi-automatically converted the model to the BiGG[9] namespace and added annotations from MetaNetX[8], BiGG[9], KEGG[10], and UniProt[43]. We also added and removed metabolites, reactions, and genes, mapped genes to reactions using AutoKEGGRec[44], and added a salmon-specific biomass reaction. We assumed that all genes mapped to the same reaction encode isozymes, thus ignoring protein complexes due to lack of knowledge. 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[20]. 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[45].

### Evaluating the quality of the metabolic model

First, we compared the presence and absence of reactions in the model to other models of multicellular eukaryotes available in the BiGG[9] namespace (*Danio rerio*[16], *Homo sapiens*[34], *Cricetulus griseus*[35], and *Phaeodactylum tricornutum*[36]). We computed the Hamming distance between each pair of models by counting reactions found in only one model and dividing by the total number of reactions.

Second, we tested the model's consistency and annotation using the community standard Memote[21] and its metabolic (in)capabilities using tasks defined for mammalian cells[22]. We adapted tasks to Atlantic salmon by moving metabolites from compartments not found in the model to the cytoplasm and modifying the expected outcomes of amino acid synthesis tests to match known essentiality[20].

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. The predicted essentiality was compared to experimental data[20].

### Analyzing oxygen-limited growth

We used parsimonious FBA (pFBA)[46] 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. 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.

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[23,24] by assuming a simple piecewise linear relationship between water oxygen saturation ($x$) and relative oxygen uptake rate:

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

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

$$\frac{\mu}{\mu_{\max}} = f\left(\frac{r}{r_{\max}}\right) \quad (2)$$

where $\mu$ 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,

$$\frac{\mu}{\mu_{\max}} = \frac{2}{1 + e^{k(x_0 - x)}} - 1 \tag{3}$$

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

$$\frac{\mu}{\mu_{\max}} = \frac{x - x_0}{K_s + x - x_0} \tag{4}$$

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

To identify contributions of reactions 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 five clusters to genes and used g:Profiler[47] to identify enriched pathways from KEGG[10]. We used the genes in the model as background, required adjusted $p \leq 0.05$, and discarded pathways outside the model's scope (antibiotics, xenobiotics, and signaling).

### 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[48] (**Table S1**). These ratios were used as coefficients for amino acids in boundary reactions 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[49]. To supplement the feed with the limiting amino acid, we set the bounds of its exchange reaction to only allow import and 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.

## 6. Model and software availability

The model and scripts that reproduce our results can be found at arken.nmbu.no/~jonvi/salarecon.

## 7. Author 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 manuscript.

## 8. Acknowledgements

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

We thank Ingunn Verne Ruud, Pedro Febrer Martínez, and Håvard Molversmyr for helpful contributions to the analysis.

## 9. References

[1] *The State of World Fisheries and Aquaculture: Meeting the Sustainable Development Goals*. Food and Agriculture Organization of the United Nations, 2018.

[2] J. Thodesen et al. "Feed intake, growth and feed utilization of offspring from wild and selected Atlantic salmon (Salmo salar)". *Aquaculture* 180.3-4 (1999).

[3] B. E. Torstensen et al. "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 (2008).

[4] J. Boissy et al. "Environmental impacts of plant-based salmonid diets at feed and farm scales". *Aquaculture* 321.1-2 (2011).

[5] S. Jennings et al. "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 (2016).

[6] T. S. Aas et al. "Utilization of feed resources in the production of Atlantic salmon (Salmo salar) in Norway: An update for 2016". *Aquaculture Reports* 15 (2019).

[7] A. N. Kolodkin et al. "Modeling Approaches in Systems Biology, Including Silicon Cell Models". *Systems Biology and Livestock Science* (2011).

[8] S. Moretti et al. "MetaNetX/MNXref: Unified namespace for metabolites and biochemical reactions in the context of metabolic models". *Nucleic Acids Research* 49.D1 (2021).

[9] C. J. Norsigian et al. "BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree". *Nucleic Acids Research* 1 (2019).

[10] M. Kanehisa et al. "New approach for understanding genome variations in KEGG". *Nucleic Acids Research* 47.D1 (2019).

[11] I. Thiele et al. "A protocol for generating a high-quality genome-scale metabolic reconstruction". *Nature Protocols* 5.1 (2010).

[12] S. N. Mendoza et al. "A systematic assessment of current genome-scale metabolic reconstruction tools". *Genome Biology* 20.1 (2019).

[13] C. Gu et al. "Current status and applications of genome-scale metabolic models". *Genome Biology* 20.1 (2019).

[14] S. Li et al. "Constructing a fish metabolic network model". *Genome Biology* 11.11 (2010).

[15] M. Bekaert. "Reconstruction of Danio rerio Metabolic Model Accounting for Subcellular Compartmentalisation". *PLoS ONE* 7.11 (2012).

[16] L. van Steijn et al. " Predicting Metabolism from Gene Expression in an Improved Whole-Genome Metabolic Network Model of Danio rerio ". *Zebrafish* 16.4 (2019).

[17] E. M. Hanna et al. "ReCodLiver0.9: Overcoming Challenges in Genome-Scale Metabolic Reconstruction of a Non-model Species". *Frontiers in Molecular Biosciences* 7.591406 (2020).

[18] S. Lien et al. "The Atlantic salmon genome provides insights into rediploidization". *Nature* 533.7602 (2016).

[19] J. D. Orth et al. "What is flux balance analysis?" *Nature Biotechnology* 28.3 (2010).

[20] B. Hatlen et al. "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 (2015).

[21] C. Lieven et al. "MEMOTE for standardized genome-scale metabolic model testing". *Nature Biotechnology* 38.3 (2020).

[22] A. Richelle et al. "What does your cell really do? Model-based assessment of mammalian cells metabolic functionalities using omics data". *bioRxiv* (2020).

[23] A. Berg et al. *Oxygen level and fish density Experiments from Hemnskjel June 1989–March 1990*. Norwegian Hydrotechnical Laboratory, 1993.

[24] A. Bergheim et al. *Effects of oxygen deficit on post-smolt salmon. Trial I*. Rogaland Research, 2002.

[25] H. Thorarensen et al. "The biological requirements for post-smolt Atlantic salmon in closed-containment systems". *Aquaculture* 312.1-4 (2011).

[26] K. Kolstad et al. "Family differences in feed efficiency in Atlantic salmon (Salmo salar)". *Aquaculture* 241 (2004).

[27] H. Dvergedal et al. "Selection for feed efficiency in Atlantic salmon using individual indicator traits based on stable isotope profiling". *Genetics Selection Evolution* 51.1 (2019).

[28] *Atlantic salmon - Nutritional requirements*. https://www.fao.org/fishery/affris/species-profiles/atlantic-salmon/nutritional-requirements/en/ (visited on 03/26/2021).

[29] P. Li et al. "New developments in fish amino acid nutrition: Towards functional and environmentally oriented aquafeeds". *Amino Acids* 37 (2009).

[30] D. J. Macqueen et al. "Functional Annotation of All Salmonid Genomes (FAASG): An international initiative supporting future salmonid research, conservation and aquaculture". *BMC Genomics* 18.1 (2017).

[31] D. Machado et al. "Fast automated reconstruction of genome-scale metabolic models for microbial species and communities". *Nucleic Acids Research* 46.15 (2018).

[32] A. Noronha et al. "The Virtual Metabolic Human database: Integrating human and gut microbiome metabolism with nutrition and disease". *Nucleic Acids Research* 47.D1 (2019).

[33] M. I. Sigurdsson et al. "A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1". *BMC Systems Biology* 4 (2010).

[34] E. Brunk et al. "Recon3D enables a three-dimensional view of gene variation in human metabolism". *Nature Biotechnology* 36.3 (2018).

[35] H. Hefzi et al. "A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism". *Cell Systems* 3.5 (2016).

[36] D. Singh et al. "Modelling metabolism of the diatom Phaeodactylum tricornutum". *Biochemical Society Transactions* 43.6 (2015).

[37] I. Thiele et al. "Personalized whole-body models integrate metabolism, physiology, and the gut microbiome". *Molecular Systems Biology* 16.5 (2020).

[38] N. S. Bar et al. "A dynamic model of nutrient pathways, growth, and body composition in fish". *Canadian Journal of Fisheries and Aquatic Sciences* 29.12 (2008).

[39] D. Deslauriers et al. "Fish Bioenergetics 4.0: An R-Based Modeling Application". *Fisheries* 42.11 (2017).

[40] P. Horton et al. "WoLF PSORT: Protein localization predictor". *Nucleic Acids Research* 35.SUPPL.2 (2007).

[41] J. J. Koehorst et al. "SAPP: Functional genome annotation and analysis through a semantic framework using FAIR principles". *Bioinformatics* 34.8 (2018).

[42] A. Ebrahim et al. "COBRApy: COnstraints-Based Reconstruction and Analysis for Python". *BMC Systems Biology* 7.74 (2013).

[43] A. Bateman. "UniProt: A worldwide hub of protein knowledge". *Nucleic Acids Research* 47.D1 (2019).

[44] E. Karlsen et al. "Automated generation of genome-scale metabolic draft reconstructions based on KEGG." *BMC Bioinformatics* 19.1 (2018).

[45] S. M. Keating et al. " SBML Level 3: an extensible format for the exchange and reuse of biological models ". *Molecular Systems Biology* 16.8 (2020).

[46] N. E. Lewis et al. "Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models". *Molecular Systems Biology* 6.390 (2010).

[47] U. Raudvere et al. "g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update)". *Nucleic Acids Research* 47.W1 (2019).

[48] *INRAE-CIRAD-AFZ Feed tables*. https://www.feedtables.com/content/table-dry-matter (visited on 03/26/2021).

[49] Y. Chen et al. "Proteome constraints reveal targets for improving microbial fitness in nutrient-rich environments". *Molecular Systems Biology* 17.4 (2021).

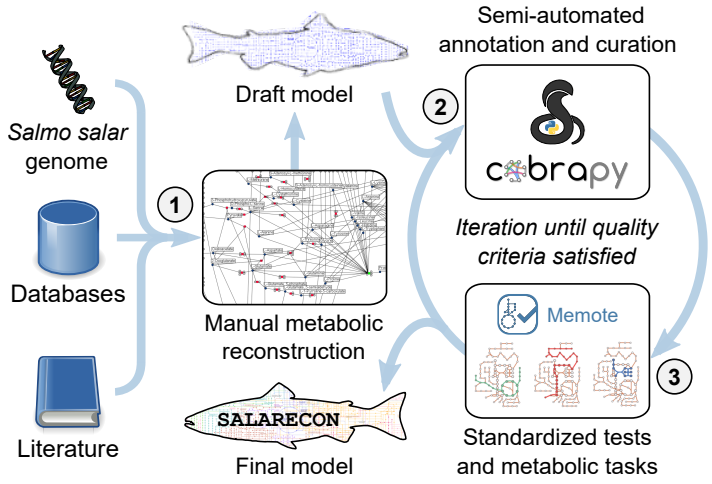


**Figure 1: Model construction.** SALARECON was built from the annotated Atlantic salmon genome, metabolic reaction and model databases, and literature. The procedure involved (1) manual metabolic network reconstruction using Insilico Discovery (Insilico Biotechnology, Stuttgart, Germany), (2) semi-automated annotation and curation using COBRApy[42], and (3) quality evaluation using the standardized metabolic model testing tool Memote[21] and metabolic tasks[22]. Steps 2 and 3 were iterated until quality criteria were satisfied. Illustration of metabolic tasks from Richelle et al.[22].

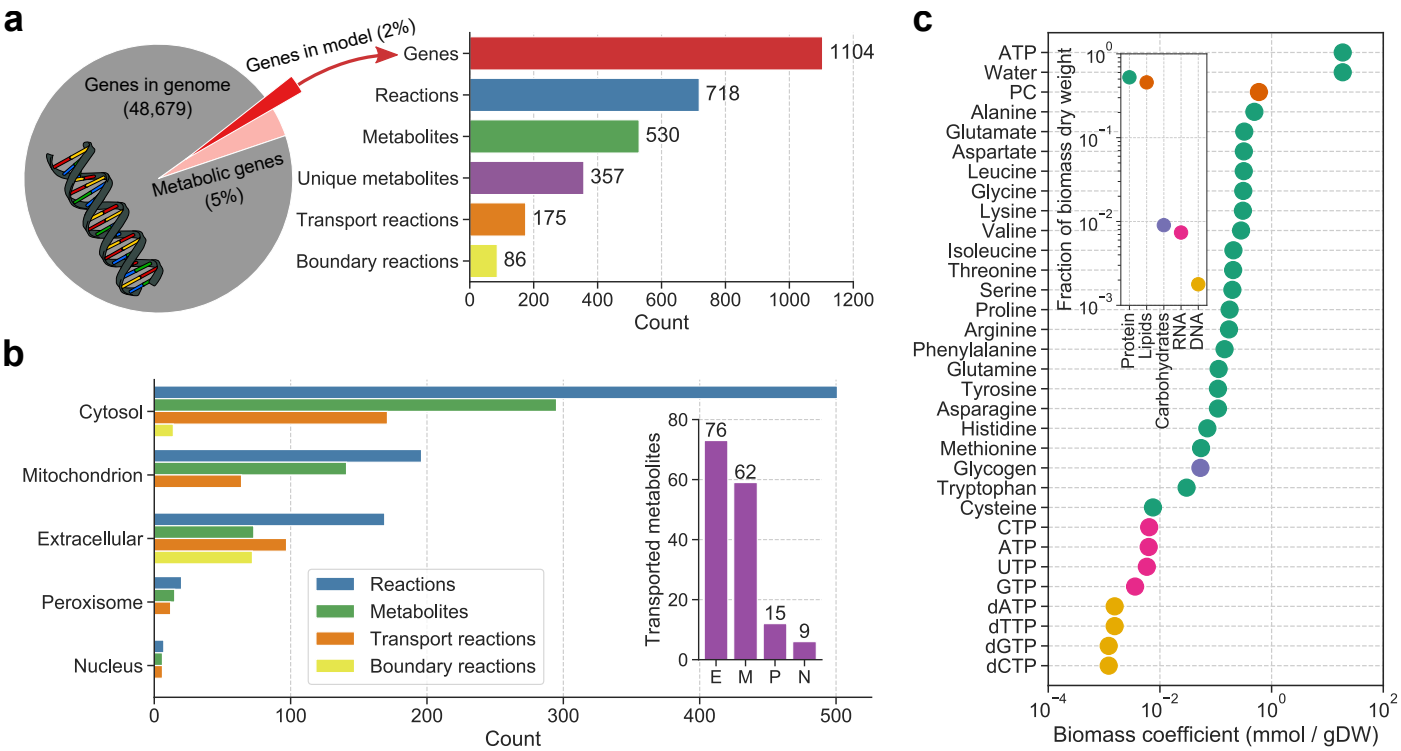


**Figure 2: Model contents.** (**a**) SALARECON contains 1,104 genes (2% of all genes and 48% of Atlantic salmon genes mapped to reactions in KEGG[10]) encoding enzymes that catalyze 718 reactions (175 transporting metabolites between compartments and 86 exchanging metabolites with the extracellular environment) and transform 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[11]. 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[20]. 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.

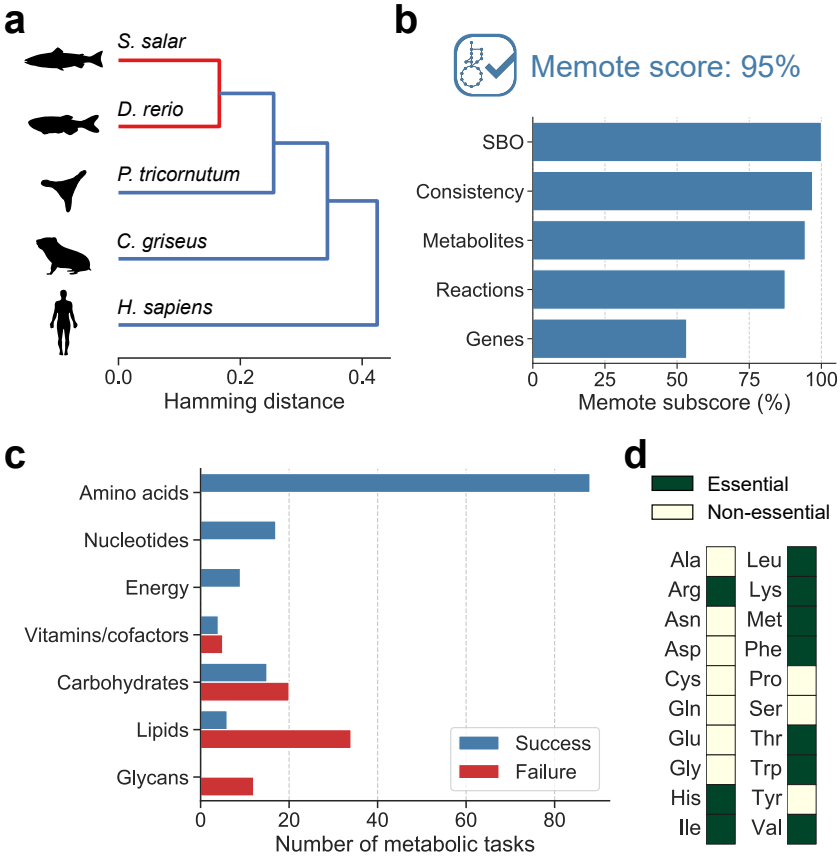


**Figure 3: Model quality evaluation.** (**a**) Comparison of metabolic models of multicellular eukaryotes based on presence and absence of reactions. Atlantic salmon (*Salmo salar*) is closer to zebrafish[16] (*Danio rerio*) than human[34] (*Homo sapiens*), chinese hamster ovary[35] (CHO, *Cricetulus griseus*), and the diatom *Phaeodactylum tricornutum*[36]. (**b**) Model score and subscores from Memote[21]. 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[22]. 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[20].

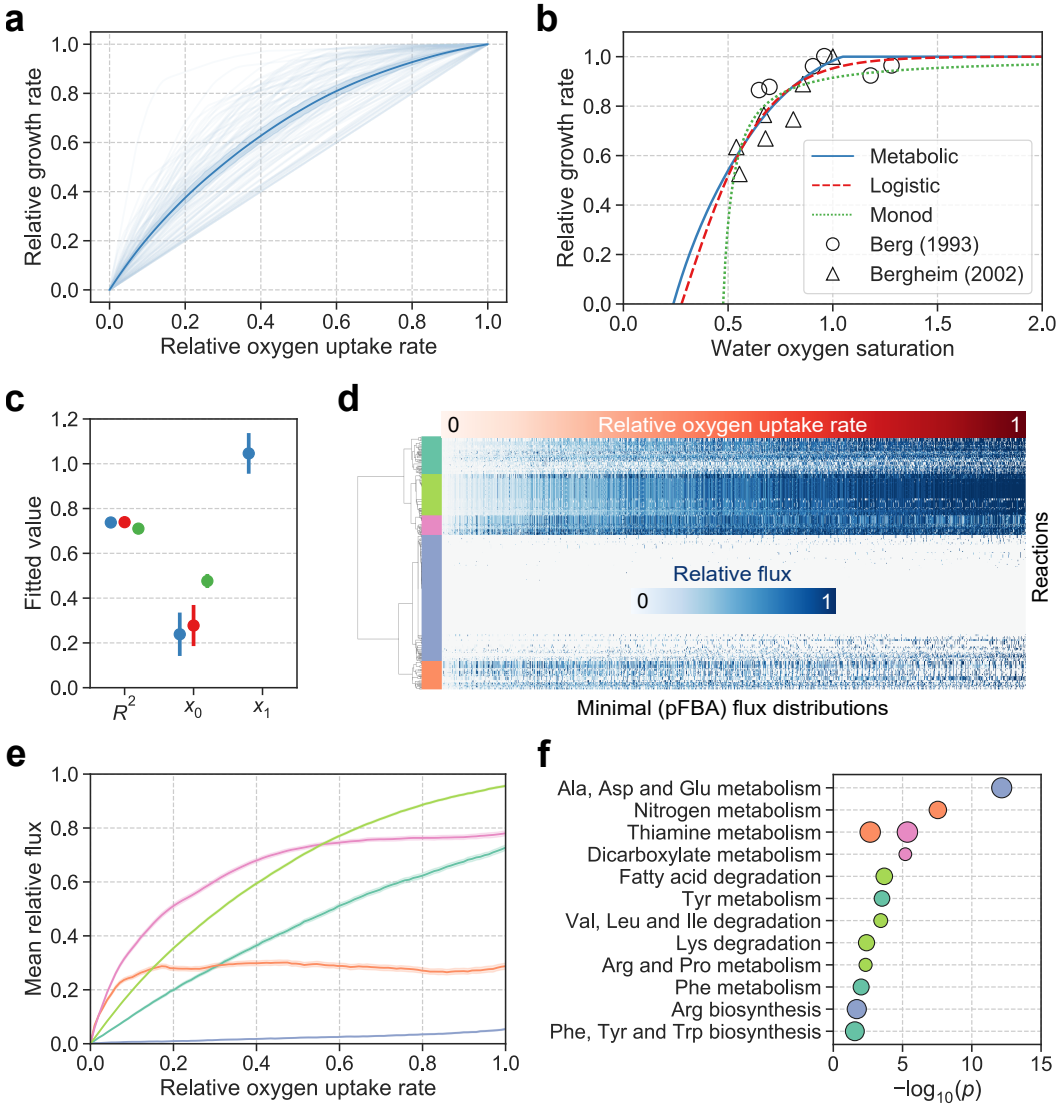


**Figure 4: Oxygen-limited growth analysis.** (**a**) Metabolic model predictions of relative growth rate under oxygen limitation as a function of relative oxygen uptake rate. Feed composition and flux capacities were randomized 100 times (light blue) and the mean across conditions is shown with 95% confidence band from bootstrapping with 1,000 samples (dark blue). (**b**) Metabolic, logistic, and Monod models fitted to experimental data from Berg et al.[23] and Bergheim et al.[24]. The metabolic model 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)[46]. 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 five clusters indicated by colors. (**e**) Mean absolute relative flux with 95% confidence bands from bootstrapping with 1,000 samples for the top five clusters with same colors as in **d**. (**f**) Enrichment of metabolic pathways from KEGG[10] for the top five clusters with same colors as in **d** and **e**.

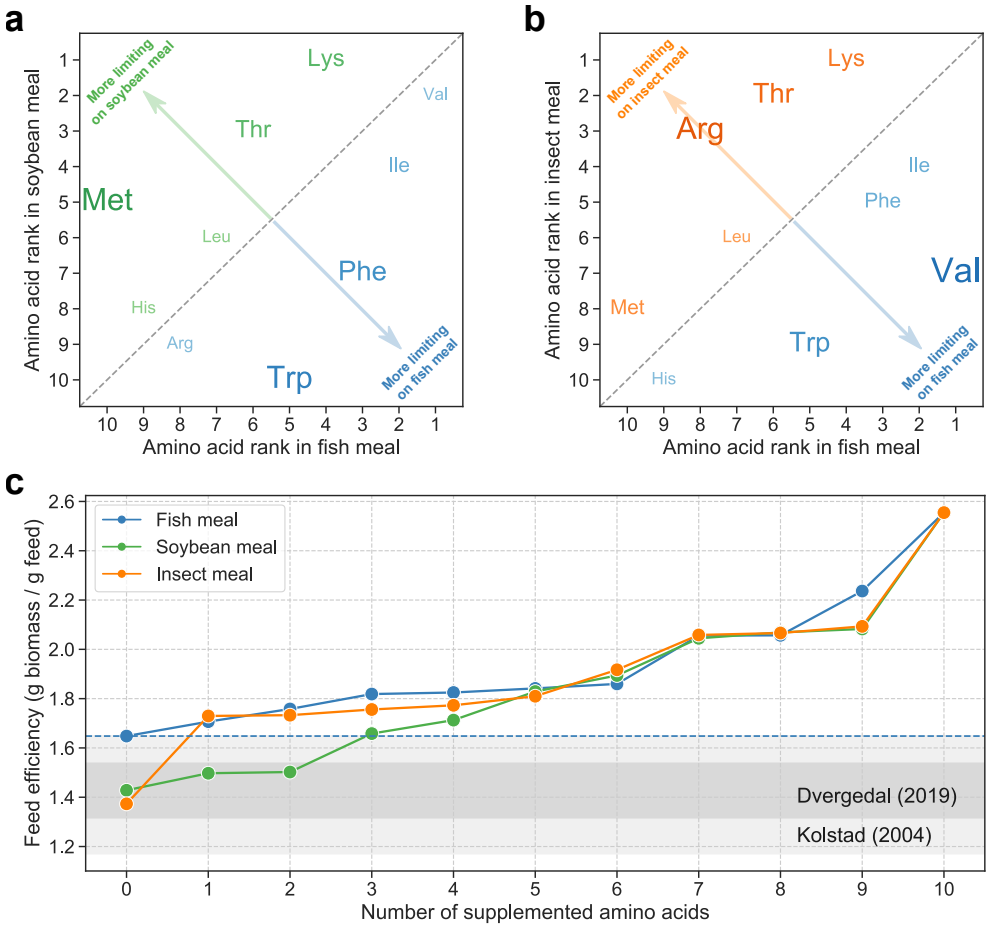


**Figure 5: Growth-limiting amino acids in commercial feed ingredients.** (**a**) Order of amino acid limitations in feed ingredients based on soybean and fish meal. Amino acids that are closer to the top left and bottom right corners are more limiting in soybean meal and fish meal, respectively, as indicated by size and color. (**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 are 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.[26] and Dvergedal et al.[27] are highlighted in gray.

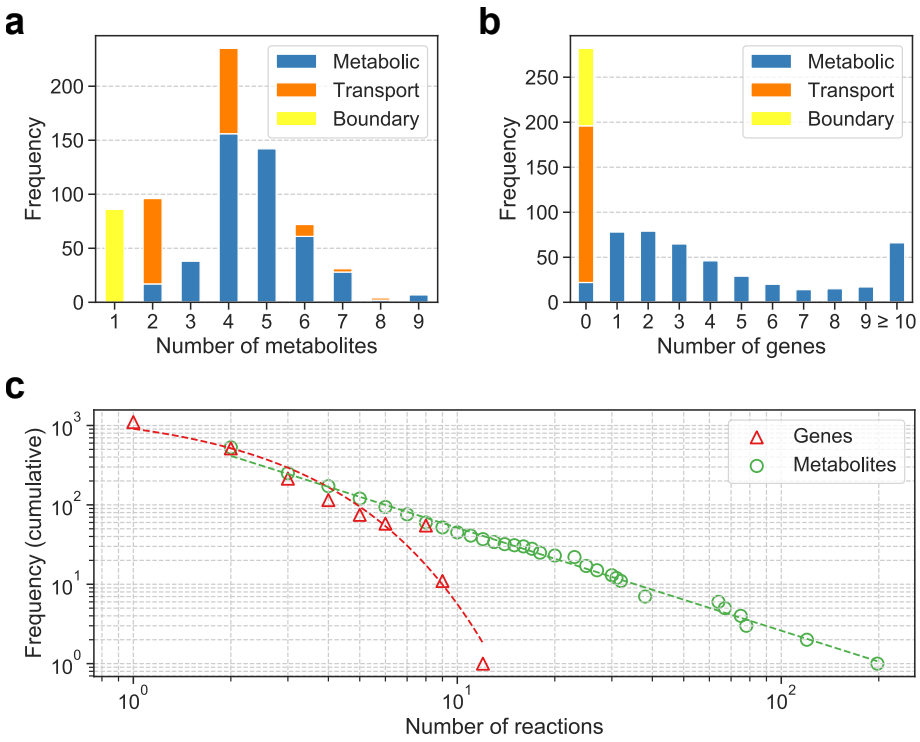


**Figure S1: Model degree distributions.** (**a**) Distribution of number of metabolites converted by reactions. Boundary reactions exchange one metabolite with the extracellular environment and transport reactions usually exchange an even number of metabolites between compartments. (**b**) Distribution 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.

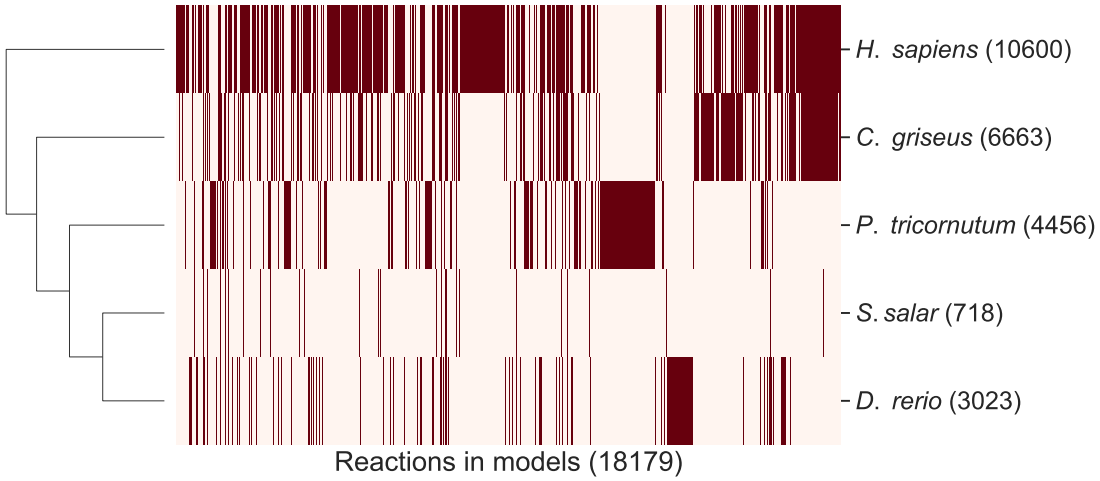


**Figure S2: Reaction contents of models of multicellular eukaryotes.** Clustered heatmap of reaction contents of metabolic models of multicellular eukaryotes. Each row is an organism, each column is a reaction, and a dark cell indicates a reaction that is found in the model of that organism. Rows are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

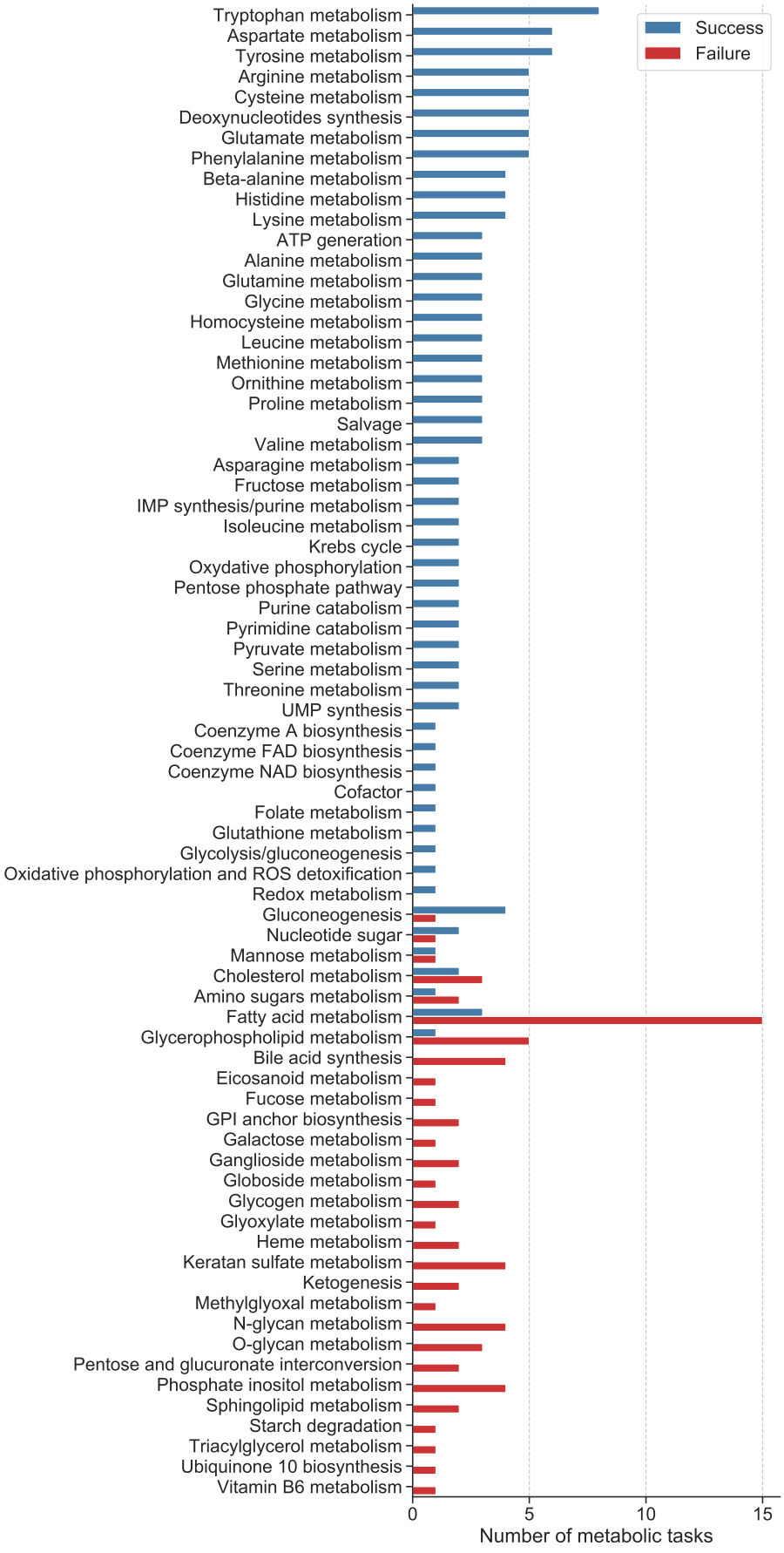


**Figure S3: Metabolic task results by subsystem.** Ability of SALARECON to perform metabolic tasks[22]. Tasks are grouped by metabolic subsystem and classified as successful if model predictions reflected expected metabolic (in)capabilities.

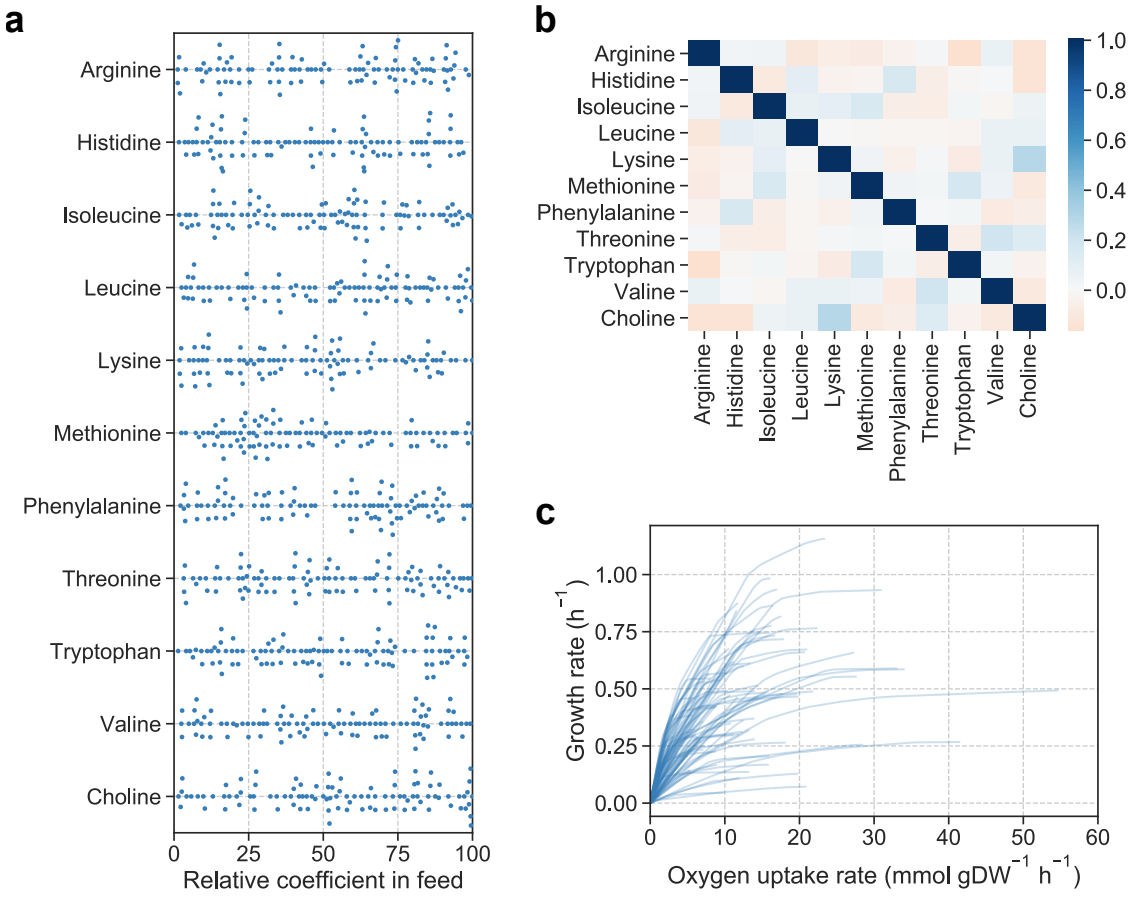


**Figure S4: Oxygen-limited growth analysis.** (**a**) Randomly sampled coefficients of amino acids and choline in minimal feeds used to predict oxygen-limited growth (100 samples). (**b**) Pairwise Pearson correlations of coefficents shown in **a**. (**c**) Predicted absolute growth rates as a function of absolute oxygen uptake rates for the 100 randomly sampled conditions.

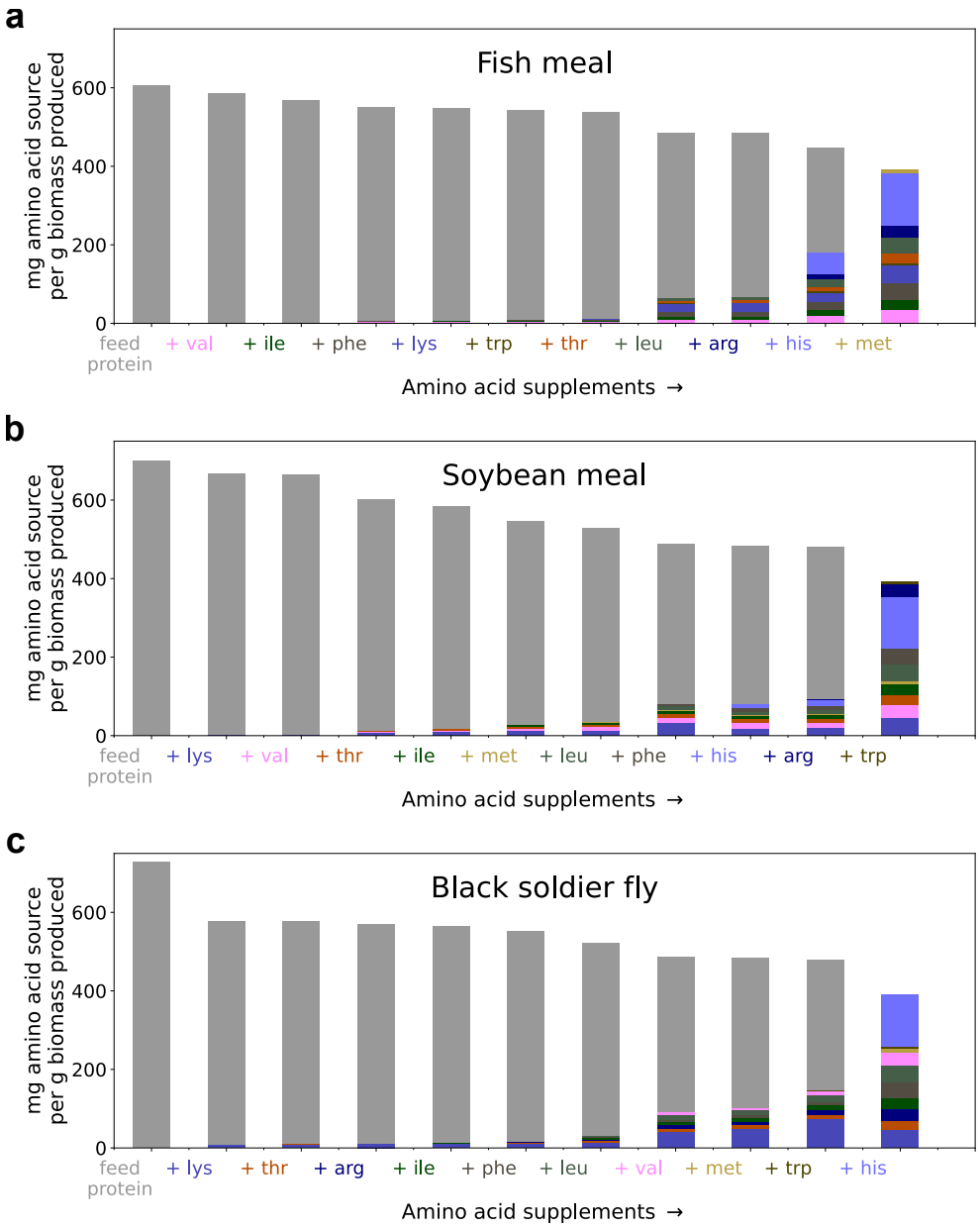


**Figure S5: Growth-limiting amino acids in commercial feed ingredients.** Feed efficiency as a function of number of supplemented amino acids, measured in mg feed ingredient and supplemented amino acids consumed / gDW biomass produced for (**a**) fish meal, (**b**) soybean meal, and (**c**) black soldier 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 S1: 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[48].

| Amino acid | Fish meal | Soybean meal | Black soldier fly larvae 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 |