Norwegian University
of Life Sciences

Master's Thesis 2021 60 ECTS
Faculty of Chemistry, Biotechnology and Food Science

Model-based integration of omics data for context-specific analysis of Atlantic salmon metabolism

Håvard Molversmyr
Bioinformatics and Applied Statistics

Acknowledgements

First and foremost, I would like to thank my main supervisor Jon Olav Vik for giving me the opportunity to write this thesis in collaboration with him and his research team on the DigiSal project. I feel honoured to have been included even before starting the thesis itself and privileged that I've been able to work with such incredibly professional, knowledgeable and amazing people. In addition to my main supervisor, I would also like to offer my sincerest thanks to my co-supervisor Ove Øyås and PhD candidate Filip Rotnes. Their guidance, advice, support and patience have been invaluable in the process of completing this thesis. Although busy with their work, they took the time to aid me with any questions I had or challenges I faced, sacrificing countless hours over the course of this work, showcasing their dedication to me and my work. For that, I am forever grateful. Furthermore, I would like to thank the rest of the DigiSal research team and the BIAS group at the university for inspiration, guidance and advice in our meetings when needed. Lastly, I would also like to thank my friends and family for their support and encouragement throughout this process, and for making my time at the Norwegian University of Life Sciences meaningful, unforgettable and indubitably a highlight of my life. I sincerely thank you from the bottom of my heart.

Ås, June, 2021

Håvard Molversmyr

Summary

Metabolism is the set of biochemical reactions that occur within a living organism in order to maintain life and grow. Most of these reactions are catalysed by enzymes which are coded for by genes. Using existing biochemical, genetic and genomic knowledge, one can link reactions together into pathways and further into metabolic networks, accounting for all enzyme-coding genes and which reactions they catalyse. Thus, metabolic networks can be made for entire organisms from their sequenced and annotated genome. As a means to predict network functionality and phenotypes, they are converted into genome-scale metabolic models (GEMs). GEMs are increasingly used to study the physiology of various organisms, ranging from microbes to complex multicellular eukaryotes, in order to understand and possibly benefit from their metabolic activities. However, there is increasing evidence that only a subset of metabolic reactions in a network is active in any given context, making GEMs superfluous when specific conditions are investigated. Therefore, several methods have been developed to extract context-specific metabolic models by integrating omics data with GEMs. Although context-specific models are assumed to yield more accurate predictions of phenotypes in a particular context, their accuracy regarding metabolic functionality has not yet been sufficiently tested. To overcome this, I here assess the capability of six model extraction methods (MEMs) to create functionally accurate context-specific models, using an Atlantic salmon GEM and hepatic transcriptomic data. To this end, I extend current methods for predicting sample-specific activity states of metabolic tasks to overcome the particular challenge of not having an objective truth to benchmark against in MEM comparisons. Context-specific models outperformed the GEM from which they were built, indicating that context-specific modelling captures realistic representations of metabolism in a given context and thus yield practical and biologically meaningful predictions. These results support current evidence that context-specific models are advantageous when studying the metabolic behaviour of organisms, especially when investigating specific contexts of interest. The findings of this study contribute to the current knowledge regarding context-specific metabolic modelling and may facilitate further research. Consequently, this may potentially be beneficial for both academic and industrial purposes.

Sammendrag

Stoffskiftet, eller metabolisme, er alle biokjemiske reaksjoner som forekommer i en levende organisme for å kunne leve og vokse. Disse reaksjonene katalyseres av enzymer som kodes for av gener. Ved å bruke eksisterende biokjemisk, genetisk og genomisk kunnskap kan man koble sammen reaksjoner til veier og videre til større nettverk, og gjøre rede for alle enzymkodende gener og hvilke reaksjoner de katalyserer. Dermed kan metabolske nettverk lages for hele organismer fra deres kartlagte genom. Som et middel til å forutsi nettverksfunksjonalitet og fenotyper, blir de konvertert til genomskala metabolske modeller (GEM). GEM-er brukes i økende grad til å studere fysiologien til forskjellige organismer, fra mikrober til komplekse flercellede eukaryoter, både for å få mer kunnskap og muligens utnytte deres metabolske aktiviteter. Imidlertid øker mengden bevis for at bare en delmengde av biokjemiske reaksjoner i et nettverk er aktivt i en gitt kontekst, noe som gjør GEM-er overflødige når spesifikke forhold undersøkes. Derfor har flere metoder blitt utviklet for å trekke ut kontekstspesifikke metabolske modeller ved å integrere ulike data som genereres med funksjonell genomikk ("omics"-data) med GEM-er. Selv om kontekstspesifikke modeller antas å mer nøyaktige kunne estimere fenotyper under bestemte forhold, har deres nøyaktighet med hensyn til metabolsk funksjonalitet ennå ikke blitt tilstrekkelig testet. I denne oppgaven vurderer jeg seks modellekstraksjonsmetoders (MEM) evne til å lage funksjonelt nøyaktige kontekstspesifikke modeller, ved å integrere hepatisk genuttrykksdata fra Atlanterhavslaks på en GEM av denne arten. Jeg bruker eksisterende metoder for å forutsi prøvespesifikke aktivitetstilstander for metabolske oppgaver for å løse den nåværende utfordringen å ikke ha en objektiv sannhet å teste mot i MEM-sammenligninger. De kontekstspesifikke modellene utkonkurrerte GEM-en de ble laget fra, noe som indikerer at kontekstspesifikk modellering evner å realistisk representere metabolismen til en organisme i en gitt kontekst og dermed kunne gi praktiske og biologisk meningsfulle spådommer. Resultatene støtter eksisterende bevis for at kontekstspesifikke modeller er fordelaktige når man studerer organismenes metabolske atferd, spesielt når man undersøker spesifikke forhold. Funnene av denne studien bidrar til å øke den nåværende kunnskapen om kontekstspesifikk metabolsk modellering og kan legge til rette for videre forskning. Derfor kan dette potensielt være gunstig for både akademiske og industrielle formål.

Table of Contents

List of Figures

List of Tables

List of Acronyms

COBRA	Constraint-based reconstruction and analysis
FBA	Flux balance analysis
GEM	Genome-scale metabolic model
GPR	Gene–protein–reaction rule
LP	Linear programming
MEM	Model extraction method
MILP	Mixed-integer linear programming
MT	Metabolic task
PCA	Prinicpal component analysis
pFBA	Parsimonious flux balance analysis
PLS-DA	Partial least squares discriminant analysis
RAL	Reaction activity level

1. Introduction

Within a cell, a multitude of biochemical reactions converts available nutrients into energy and the building blocks required to maintain vital processes and to grow. Metabolism is this vast network of reactions and, although very complicated and highly interconnected, it is often studied in individual parts (Sulheim, 2020). Although this has led to extensive knowledge about single reactions and metabolic pathways, it provides limited insights into the function of the network as a whole. By piecing together individual reactions and pathways, the structure and dynamics of entire systems can be scrutinised to uncover the function of the individual parts in interaction. Hence, by approaching networks from a systems level, one can start studying the entire metabolism of organisms to understand and possibly benefit from the given organism's metabolic activities.

1.1 What is systems biology?

Systems biology is an interdisciplinary field built upon the understanding that molecular components form the networks representing the entirety of living organisms, and that these components are intimately interconnected and explicable only by reference to the whole (Palsson, 2015). The relationship between cell components, with their genetic bases, and their physiological functions are interpreted holistically, thus linking genotype to phenotype (Palsson, 2015). Systems biologists take the step from network reconstructions, i.e. maps of the individual parts and the links between them, to mathematical models fit for simulations, monitoring component interactions and the resulting function and behaviour of the biological system. This practice was made possible by the emergence of the first full genome sequences and high-throughput sequencing technologies in the mid-1990s, enabling scientists to study organism-specific metabolism as systems (Palsson, 2015). By identifying and functionally annotating genes in genomes using bioinformatics tools and connecting gene-protein-reaction relationships through reaction databases, an organism's metabolism could now be outlined as a genome-scale metabolic model (GEM) (Sulheim, 2020).

1.2 Metabolic modelling

The concept of metabolic modelling is based on metabolites and reactions, forming networks when connected. Although different graph representations have been discussed (Montañez et al., 2010; Cazzaniga et al., 2014), metabolites and reactions are often represented by nodes and links, respectively. In this case, reactions that share metabolites as either reactants or products may be linked together to form reaction maps. For instance, a metabolite **a** may be taken up from the environment and transported into the intracellular space, before it is used in a reaction to produce a second metabolite **b**. This metabolite may in turn be a part of separate reactions, thus extending the reaction map. This network would be represented by a node for each of the metabolites and a link between the metabolites involved in the given reaction (Fig. 1.1). However, network complexity quickly increases with the number of reactions, revealing the need for a mathematical format to analyse GEMs.

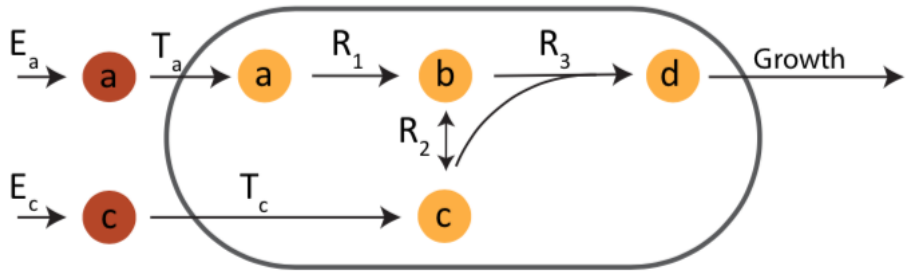


Figure 1.1: Minimal reaction network. Reactions and metabolites are represented by arrows and circles, respectively. Exchange reactions (E) allow metabolites to enter or leave the system, transport reactions (T) transport metabolites across membranes, while metabolic reactions (R) take place within cellular compartments. The objective function is represented by an artificial reaction (growth). Metabolites may be either intracellular (yellow) or extracellular (red). Figure from Sulheim (2020).

The mathematical format used for GEM analyses is based on the stoichiometric matrix, **S**. This matrix comprises the stoichiometric coefficients of the metabolites in each reaction in the network. The columns of **S** represent reactions while the rows represent the metabolites, with the mass-balanced coefficients of the metabolites in each reaction as entries in the matrix. For every metabolite consumed and produced in any given reaction, there is a negative and positive stoichiometric coefficient, respectively, and for every metabolite that does not take part in a particular reaction, a coefficient of zero is used. In addition to the intracellular reactions, a GEM also includes transport and boundary reactions. Furthermore, it is common to add a reaction representing a specific biological functionality of the system. The mathematical representations of such reactions are referred to as objective functions (Palsson, 2015). Biomass production is often used as the primary metabolic objective. However, it may be any combination of the reactions in the model.

The turnover rate or flow of metabolites through a network is called fluxes (Bordbar et al., 2014), commonly measured in mmol produced metabolites per gram dry weight per unit time, usually hours (mmol gDW^{-1} h^{-1}) (Orth et al., 2010b). The flux through all n reactions can be represented mathematically as a flux vector,

$$\mathbf{v} = (v_1, v_2, \ldots, v_n)$$

which when multiplied with the stoichiometric matrix results in a vector of changes in metabolite concentrations ($\mathbf{x}$) over time:

$$\mathbf{Sv} = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} \tag{1.1}$$

Metabolite turnover rates are documented to be on the order of seconds (Buchholz et al., 2002). This is much faster than those of other biochemical processes such as transcription, translation and protein degradation, which are on the order of minutes to hours (Maier et al., 2011; Shamir et al., 2016). Thus, metabolite concentrations are assumed to be constant and a steady-state constraint is imposed on the fluxes (Orth et al., 2010b). Consequently, equation (1.1) reduces to

$$\mathbf{Sv} = \mathbf{0} \tag{1.2}$$

Since the stoichiometric matrix usually comprises more reactions than metabolites, there are infinite solutions to equation (1.2), which spans a volume known as the null space of $\mathbf{S}$ (Sulheim, 2020). Every point in the null space of $\mathbf{S}$, i.e. every solution, is a flux distribution. In order to further constrain the allowable steady-state flux distributions of the network, upper and lower reaction flux bounds are set, specifying the maximum and minimum allowable fluxes, respectively (Orth et al., 2010b; Palsson, 2015). All solutions that satisfy these additional constraints define the allowable solution space of the network under the given constraints (Fig. 1.2).

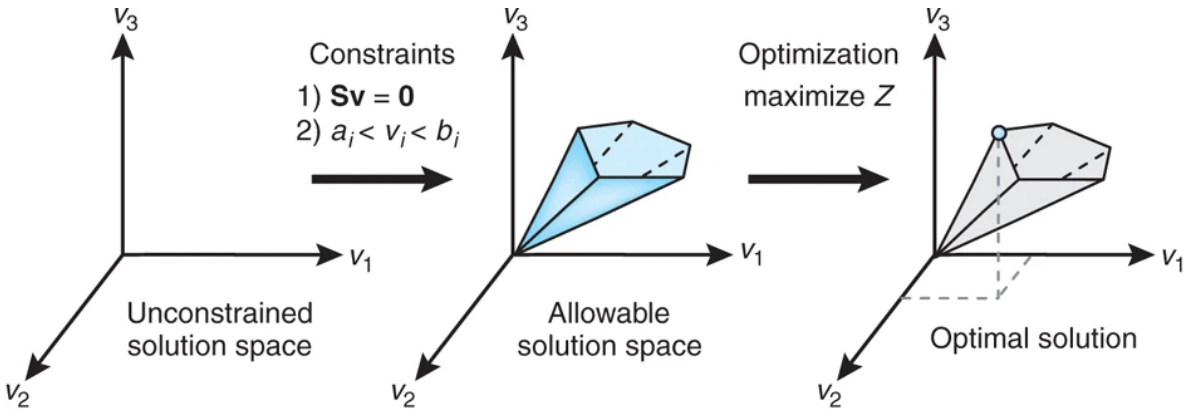


Figure 1.2: The solution space. The solution space defines the allowable fluxes of an organism given the current constraints. Flux balance analysis can be used to find optimal solutions in the solution space. Figure from Orth et al. (2010b).

The solution space demarcates an organism's achievable cellular functions, and thus phenotypes, in a particular environment (Price et al., 2004; Palsson, 2015). The expressed functionality of an organism is determined by its evolutionary past and present, as organisms are subject to environment-specific selection pressure based on e.g. nutrient availability. Therefore, pathways will adapt to optimise metabolic objectives (Nam et al., 2011; Schuetz et al., 2012) and the resulting state is in some sense optimal for the organism for the specific conditions that have driven its evolution (Palsson, 2015). Consequently, the functional states and capabilities of a network given a set of constraints can be predicted using optimisation methods (Palsson, 2015).

Numerous optimisation methods exist (Banga, 2008) and can be categorised as either linear or non-linear depending on the linearity of the objective function and the constraints (Zomorrodi et al., 2012). In conjunction with GEMs, these methods can be used to address relevant biological questions, which may be outlined through the chosen objective function. To date, the most commonly used optimisation method for large-scale systems has been linear programming (LP) (Beard and Qian, 2005). This method optimises a linear objective function, finding extreme solutions of interest by maximising or minimising as appropriate. Thus, the solutions determine the optimal use of the network to meet the specified objective, enabling the study of specific capabilities of a metabolic network (Palsson, 2015). Optimisation methods are therefore beneficial for both academia and the industry, enabling exploration of physiologically meaningful objectives and objectives tailored for commercial applications.

The most commonly used approach for studying the biological capabilities of GEMs is flux balance analysis (FBA). FBA is a method that relies on the stoichiometry, growth requirements and optimality principles of a metabolic network (Varma and Palsson, 1994). The method aims to find one extreme solution in the solution space of allowable physiological states for a chosen objective through the means of LP (Orth et al., 2010b). All returned solutions of the FBA algorithm will consequently be on the edges of the solution space (Fig. 1.2). Given a set of constraints, the solutions need not be one unique point in the solution space. Rather, it may be a set of equal solutions, which is often the case for GEMs (Reed and Palsson, 2004). In FBA, the following linear program is solved:

$$\begin{aligned} &\text{maximise (or minimise) } Z = \mathbf{c}^T\mathbf{v} \\ &\text{subject to:} \\ &\mathbf{Sv} = \mathbf{0} \\ &\mathbf{v_{min}} \leq \mathbf{v} \leq \mathbf{v_{max}} \end{aligned} \tag{1.3}$$

where $\mathbf{v}$ is a vector of the flux through all reactions in the network and $\mathbf{c}$ is a vector of weights indicating the contribution of each reaction to the objective function Z, such that $Z = \mathbf{c}^T\mathbf{v} = c_1 v_1 + c_2 v_2 + \ldots, c_n v_n$. v_{min} and v_{max} defines the lower and upper flux bounds, respectively.

A commonly used objective for GEMs is the maximisation of growth rate. The definition of this objective depends on knowledge of the cell's chemical composition and its energetic requirements (Feist and Palsson, 2010) (Fig. 1.3). From experimental data, one can determine the requirements for making one gram of biomass, and convert this into a reaction that drains precursor metabolites and produces by-product metabolites (Orth et al., 2010a). Additionally, the energetic cost of biosynthesis and maintaining biomass needs to be taken into account for realistic growth simulations (Feist and Palsson, 2010; Palsson, 2015). To be able to use the reaction for modelling purposes, the multiple components and precursors of biomass have to be quantified in relative proportions, such that the reaction drains and produces metabolites at appropriate ratios. The biomass reaction is then scaled such that its flux represents a specific growth rate per unit time, usually per hour (Feist and Palsson, 2010).

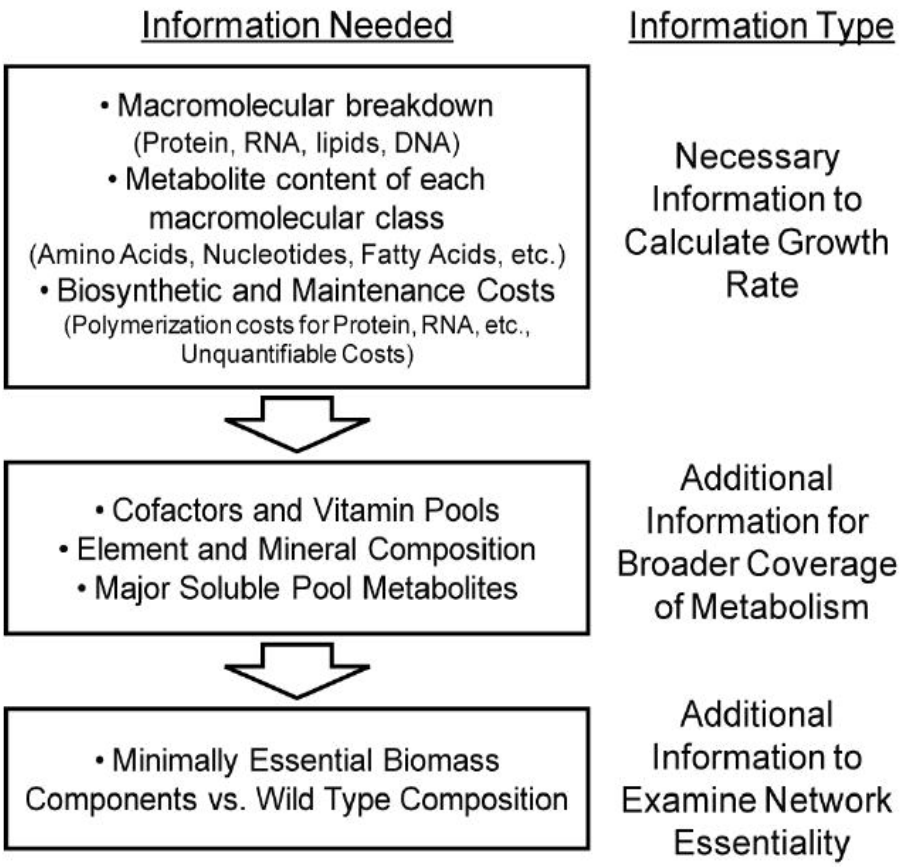


Figure 1.3: Formulation of the biomass objective function. Generating a biomass objective function is dependent on different types of information. Figure from Feist and Palsson (2010).

Although maximising growth rate is often used as an objective function in GEMs, this may not always be the obvious choice (Feist and Palsson, 2016). For microorganisms, cellular growth is beneficial to secure the continuation of its genetic material and is thus an evolutionarily plausible choice of a metabolic objective. Reproduction, however, is not the biological goal of all cells of multicellular organisms (Fouladiha and Marashi, 2017). Different organs, tissues and cell types are specialised in various metabolic functions, leading to different objectives (e.g. Chang et al., 2010; Gille et al., 2010; Toroghi et al., 2016), which may also depend on interactions with each other (Lewis et al., 2010a; Robaina Estévez and Nikoloski, 2014). Furthermore, maximisation or minimisation of specific metabolite production rates may be more valuable for industrial applications, where growth may be a constraint. Thus, finding an appropriate objective function is challenging, and may depend on the size and scope of the model as well as the tissue or cell type.

1.3 Context-specific metabolic modelling

Over the past few decades, the scope of GEMs has greatly increased. Genome sequencing techniques have evolved, leading to increased and improved omics data availability and analyses (Shendure and Aiden, 2012; Land et al., 2015; Zhang et al., 2020). Thanks to methods for reconstructing metabolic networks from genome sequences (Thiele and Palsson, 2010; Mendoza et al., 2019) along with databases of metabolic reactions and models (Ganter et al., 2013; King et al., 2016; Kanehisa et al., 2019), GEMs have become increasingly available. Consequently, GEMs have been reconstructed for numerous organisms, ranging from microbes to multicellular organisms (Gu et al., 2019).

Although GEMs include all metabolites and reactions that comprise the cells of an organism, the metabolic behaviour and activity level of each tissue or cell type differ (Shlomi et al., 2008; Uhlén et al., 2015). The activity profile of cells' and tissues' molecular functions may be altered at any moment in time to adapt to its surrounding environment (Richelle et al., 2020). Hence, a complete genome-scale metabolic model may be too flexible and superfluous in certain conditions, which may lead to inaccurate predictions (Pacheco et al., 2015). Therefore, several methods have been developed to integrate omics data with GEMs to extract context-specific models which represent a specific condition from which the omics data has been sampled (Machado and Herrgård, 2014; Robaina Estévez and Nikoloski, 2014). As context-specific models are a representation of the metabolism of a single cell, tissue or condition, their ability for accurate predictions in the given context is assumed to be greater than that of complete GEMs (Opdam et al., 2017).

There are myriad applications of context-specific models. For instance, context-specific models of Recon 1, the first genome-scale reconstruction of human metabolism, have been used for studying host-pathogen interactions (Bordbar et al., 2010) and brain meta-

bolism (Lewis et al., 2010a), as well as for drug target discoveries in cancer (Frezza et al., 2011). Moreover, non-human models have been used to study e.g. the flux distribution in mesophyll and bundle sheet cells of C_4 grasses during photosynthesis (Dal'Molin et al., 2010), the metabolic behaviour of organs related to production, storage and consumption of sugars during the generative phase of barley (*Hordeum vulgare*) (Grafahrend-Belau et al., 2013), or more recently the stress responses to drought for thale cress (*Arabidopsis thaliana*) (Siriwach et al., 2020). Needless to say, the potential that lies within GEMs and in their extension context-specific models for studying metabolic behaviour is enormous, thus enabling them to address and contribute to solving key challenges.

Several methods have been developed to extract subsets of GEMs representing a particular condition, cell or tissue (Machado and Herrgård, 2014; Robaina Estévez and Nikoloski, 2014; Opdam et al., 2017). These methods, hereafter model extraction methods (MEMs), deploy diverse strategies to find the best reduction of the full model to create context-specific models (Robaina Estévez and Nikoloski, 2014; Opdam et al., 2017; Correia et al., 2018). Numerous MEMs already exist, and new methods are being created rapidly. Some commonly used MEMs include FASTCORE (Vlassis et al., 2014), GIMME (Becker and Palsson, 2008), iMAT (Shlomi et al., 2008; Zur et al., 2010), INIT (Agren et al., 2012), MBA (Jerby et al., 2010) and mCADRE (Wang et al., 2012), which can be categorised into three families based on their mathematical and algorithmic formulation, namely the GIMME-like, iMAT-like and MBA-like (Robaina Estévez and Nikoloski, 2014) (Table 1.1).

Table 1.1: Overview of implemented MEMs. Concise description of each model extraction method (MEM) and their required inputs in the COBRA Toolbox 3.0 (Heirendt et al., 2019). Information of each parameter's contents when implementing the methods on SALARECON is also included. Note that the biomass reaction was handled manually in the utilisation of each MEM to ensure its inclusion in the extracted model.

Method		**Description (Opdam et al., 2017)**	**Required inputs (Richelle et al., 2019b)**	**Parameter contents**
MBA-like	**MBA (Jerby et al., 2010)**	Two sets of core reactions are defined, with high and medium probability to be active in a given context. The algorithm then reconstructs a context-specific model containing all of the high confidence reactions, as many medium confidence reactions as possible, and a minimal set of other generic model reactions that are required for obtaining overall model consistency.	Two sets of core reactions; one with high confidence and one with medium confidence of being active in the given context.	The 75^{th} percentile of the distribution of all gene scores and $5\ln(2)$ were used as thresholds to divide all reactions into the reaction sets. All reactions associated with a gene score above the 75^{th} percentile were added to the high confidence set, while all remaining reactions with a score above $5\ln(2)$ were added to the medium confidence reaction set. The biomass reaction was manually added to the high confidence set.

Table 1.1 continued from previous page

Method		Description (Opdam et al., 2017)	Required inputs (Richelle et al., 2019b)	Parameter contents
MBA-like	**mCADRE (Wang et al., 2012)**	Using a defined set of core reactions, all non-core reactions are pruned based on expression level, connectivity to the core, in addition to a confidence score. Superfluous reactions, i.e. reactions not needed to support the core or defined functionalities, are removed. In cases where a core reaction is supported by a certain number of zero-expression reactions, it is removed.	Two sets of reaction scores; a quantification of how often a gene is expressed across samples in the dataset, the ubiquity scores, and literature-based evidence scores, the confidence scores.	As the expression distribution of genes is used in the calculation for the gene scores, the gene scores were used as the ubiquity scores. The biomass reaction was given a confidence score of 3. All other reactions were given a score of 1 if they were associated with at least one gene or 0 otherwise.
	FASTCORE (Vlassis et al., 2014)	A set of core reactions guaranteed to be active in a certain context of interest is defined, and the algorithm tries to find the minimal number of reactions possible to support the core.	Single set of core reactions.	All reactions with a gene score superior to $5\ln(2)$ was added to the core reaction set, as well as the biomass reaction, which was manually added.
iMAT-like	**iMAT (Shlomi et al., 2008; Zur et al., 2010)**	Maximises the number of matches between a reaction's minimum flux value, and the group it belongs to, i.e. either highly or lowly expressed. Thus, it finds an optimal trade-off between including highly expressed reactions and removing lowly expressed reactions.	Two threshold values defining non-expressed and expressed genes, in addition to the gene expression values themselves.	Gene scores were used as expression values, while the upper and lower threshold values were both set to $5\ln(2)$. The biomass reaction was manually attributed a score of $10\ln(2)$.
	INIT (Agren et al., 2012)	Finds an optimal trade-off between including and removing reactions based on their attributed weights.	An attributed weight to each reaction, which is positive or negative for highly or lowly expressed reactions, respectively.	Reactions with a gene score below $5\ln(2)$ were attributed a weight of -8, while gene score divided by $5\ln(2)$ was defined as the weight for the remaining reactions. The weight of the biomass reaction was put to the maximum of obtained reaction weights.
GIMME-like	**GIMME (Becker and Palsson, 2008)**	Removes reactions associated with an expression level below a user-defined threshold value. Subsequently, reactions will be reinserted to achieve a required metabolic function (i.e. the objective function).	A gene expression dataset and a metabolic objective function.	Gene scores were used as expression values, while the threshold value was set to $5\ln(2)$. The biomass reaction was manually attributed a score of $10\ln(2)$.

Of the above-mentioned MEMs, the GIMME-like family comprises only GIMME, which assumes a correlation between transcriptomics data and reaction fluxes. It first optimises an objective function with FBA, before removing inactive reactions, i.e. reactions associated with transcript levels below a user-defined threshold. If the flux through the objective function is not above a set fraction of the full model, a minimum number of inactive reactions are added back into the system. This is done by minimising the difference between their expression levels and flux values.

The iMAT-like family encompasses iMAT and INIT, which aims to find a trade-off between removing reactions associated with low gene expression and keeping reactions whose genes are highly expressed. iMAT applies a user-defined threshold to categorise reactions as either highly or lowly expressed based on their associated expression values, before finding a flux distribution that maximises the number of matches between reaction activity and gene expression state. For the INIT algorithm, reactions are given positive or negative weights based on whether they are associated with high or low transcript levels. It then tries to find a trade-off between including reactions with positive weights and removing reactions with negative weights.

The MBA-like family consist of MBA, mCADRE and FASTCORE, all of which utilise reaction sets to determine which reactions should be included in the extracted model, the core set, and which reactions can be removed, the non-core set. MBA requires the user to categorise core reactions into a high- and medium-confidence set, based on their likelihood to be present in the extracted model and then prunes all non-core reactions in an iterative process. A randomly selected non-core reaction is removed from the model, which is then checked for consistency (i.e. no blocked reactions), and if all high-confidence and a user-defined fraction (defaults to 0.5) of medium-confidence reactions remain unblocked, the selected non-core reaction and potential other blocked medium-confidence and non-core reactions are removed from the final model. The process continues until the non-core reaction set is empty. mCADRE ranks reactions according to expression-, connectivity-, and confidence level-based scores. Based on the expression-based scores, all reactions associated with a score above a set threshold value are defined as core reactions, while the rest are defined as non-core reactions. The non-core reaction set is then ranked according to connectivity- and confidence level-based scores, before it is pruned sequentially, removing reactions not necessary to support the core. FASTCORE also aims to keep all core reactions, and finds the minimum number of non-core reactions to support the core.

The overall goal of MEMs is to extract a model with a functional subset of reactions from a larger model, though their approach may differ. Unlike the other MEMs, GIMME requires *a priori* knowledge of a metabolic objective, which has been suggested to yield more accurate models and flux distributions (Machado and Herrgård, 2014; Robaina Estévez and

Nikoloski, 2014). However, for organisms with no clear biological objective such as multicellular organisms, this may not be advantageous and other MEMs may be more easily applied (Blazier and Papin, 2012). Although MEMs like iMAT and INIT, for this reason, may be better suited for extracting context-specific models of multicellular organisms, they do so through the means of computationally demanding mixed-integer linear programming (MILP) problems, as opposed to the LP problems utilised by the other MEMs. Therefore, these methods generally require longer computation time, but a trade-off between searching for multiple model optima and computational time may be considered and defined by the user in the implementation of these MEMs. Similarly to iMAT and INIT, the methods that comprise the MBA-like family are independent of a metabolic objective. In contrast to the other MEMs, however, this family allows the user to integrate different data types with biochemical knowledge, which may ultimately increase the certainty of reaction presence in a certain context (Robaina Estévez and Nikoloski, 2014). Furthermore, the definition of the core set for mCADRE can be fully automated, caused by its expression-based determination of threshold values. MBA and mCADRE may on the other hand be quite time-consuming due to their pruning process. However, provided that a defined core set is given, FASTCORE outperforms these methods with regard to computational time by two or three orders of magnitude (Vlassis et al., 2014), and may thus be the method of choice when time is limited.

1.4 Metabolic modelling for aquaculture

As a result of a growing human population and diminishing returns of terrestrial agriculture, there has been a change towards aquatic environments for sustenance (Tibbetts, 2018). For the past few decades, the growth of aquatic farming (i.e. fish and shellfish) has greatly surpassed that of terrestrial farming, and is currently responsible for approximately half of the international fish consumption, with expectations of further growth (Troell et al., 2014). Salmonid aquaculture has experienced similar growth within the aquaculture industry and Atlantic salmon (*Salmo salar*) has become the most valued fish commodity in the world (FAO, 2018).

Aquaculture has become a major consumer of finite marine resources and feeds, which has led to a shift in feed composition with higher proportions of alternative proteins, grains and oils (Tibbetts, 2018). However, salmon are carnivorous by nature and not adapted to eating plants. Current plant-based feeds are complex and have been reported to have negative impacts on several factors, including fish welfare, growth, feed utilisation, the environment, as well as fatty acid composition (Mundheim et al., 2004; Gatlin III et al., 2007; Turchini et al., 2009; Boissy et al., 2011; Jennings et al., 2016).

Fatty acid composition is important both for the fish itself and for the consumer. For many fish, and especially fatty fish like the Atlantic salmon, lipids and their constituent fatty acids are a major component, reflecting their importance as a role in metabolic energy expenditure in fish, including growth (Tocher, 2003). Moreover, unlike fatty meat products, fish lipids comprise copious quantities of omega-3 long-chain polyunsaturated fatty acids (LC omega-3), which have been well documented to have beneficial effects for humans (Domingo et al., 2007). Consequently, certain levels of LC omega-3 are sought after in salmon farming. However, the vegetable oils used in aquafeeds usually contain large amounts of omega-6 fatty acids, which are incorporated into the tissues of farmed fish (Jennings et al., 2016), and causes changes in the fatty acid composition of the fish. For Atlantic salmon, both total LC omega-3 content and the omega-3/omega-6 ratio have decreased (Nichols et al., 2014), which have been indicated to reduce consumption benefits (Ramsden et al., 2010). Thus, an understanding of lipid and fatty acid metabolism is crucial for salmon farming.

A fully sequenced and mapped genome of the Atlantic salmon was published in 2016 (Lien et al., 2016), making a GEM reconstruction of this species a reality in 2021 (Zakhartsev et al., 2021), hereafter referred to as SALARECON. As a result, it is now possible to study how molecules get modified from what the fish eats until it makes the flesh that we eat, thus providing a framework for connecting the genome and metabolism of salmon and addressing challenges such as the prediction of novel feeds. Although lipid metabolism is not well represented in most GEMs today (Sánchez et al., 2019), SALARECON included, the liver is important to understand lipid metabolism (Henderson, 1996). Hence, being able to create accurate context-specific models from this tissue, and establishing empirical evidence that they better represent its metabolism than the GEM, would have potential benefit for the salmon farming industry concerning economic growth and development, while concurrent considerations of fish welfare and other factors can be accounted for.

From the emergence of commercial-scale salmon farming in the 1960s to its global economic importance today, a large body of research on the Atlantic salmon has been established, ranging from effects of domestication and selective breeding to genome-wide studies of function and regulation (Macqueen et al., 2017; Houston and Macqueen, 2019). Large quantities of omics data are thus available for this species, all of which contain information on metabolic network changes across different conditions between individuals. This information can be used to capture differences between conditions, cell types and tissues using context-specific modelling, and derive non-trivial hypotheses through *in silico* flux analysis.

1.5 Aims of the thesis

Using transcriptomics data from Atlantic salmon liver, I here test whether context-specific metabolic models created with the above-mentioned MEMs can outperform the full salmon GEM, and thus potentially be beneficial for the aquaculture industry in the future. To this end, I test the aptness of these methods to create liver-specific metabolic models, by comparing their ability to successfully perform a list of metabolic tasks (Fig. 1.4) curated and standardised by Richelle et al. (2020) with that of the full model. It is acknowledged that this list was originally intended for model-based assessment of mammalian cells. However, the context-specific models in this thesis were extracted from and compared to only one GEM. Thus, only relative metabolic functionalities are compared.

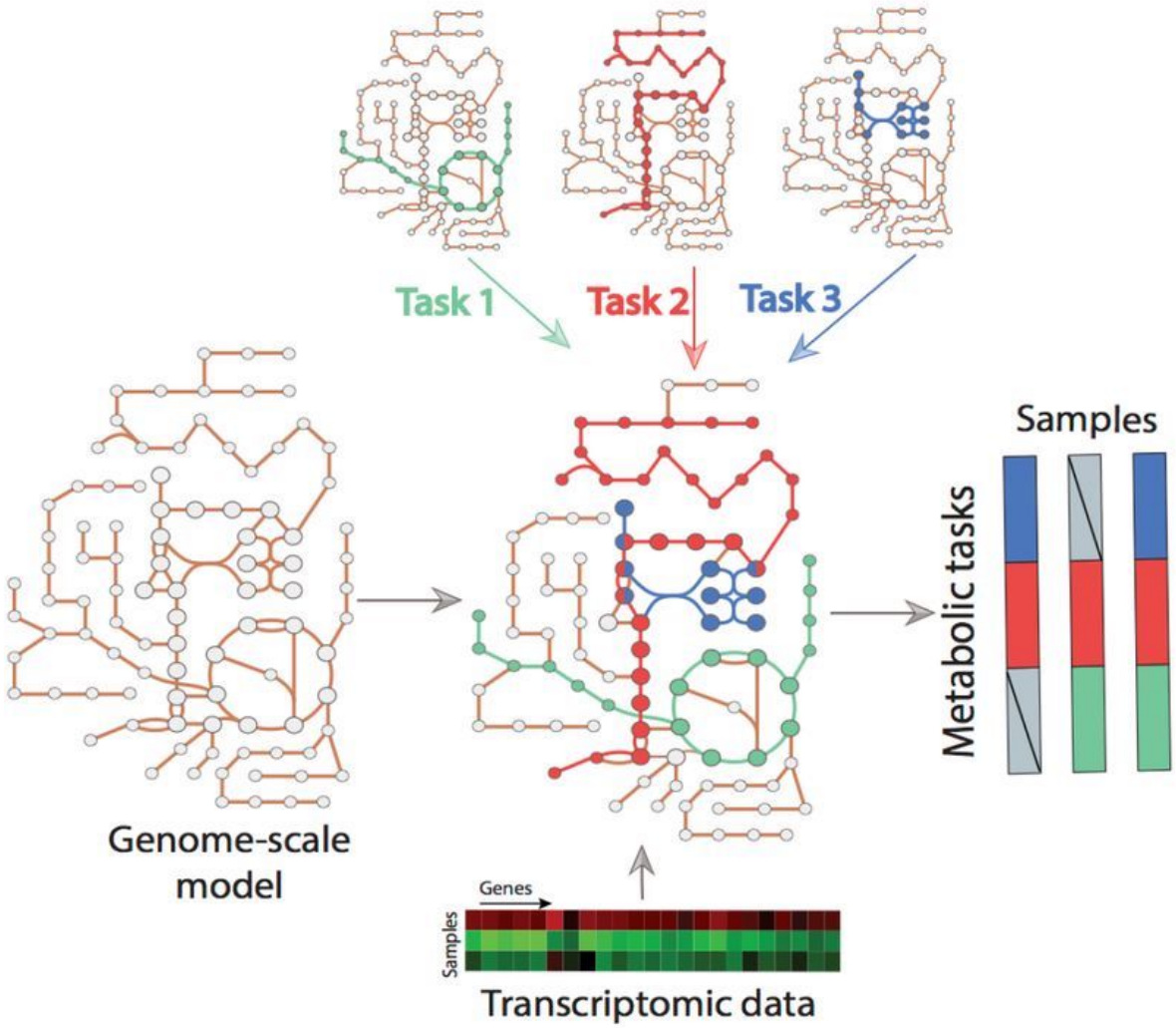


Figure 1.4: Metabolic tasks. These are sets of reactions required to transform input metabolites into defined products. A list of metabolic tasks can be used to infer the activity of metabolic functions in a metabolic model from transcriptomic data. Figure from Richelle et al. (2020).

2. Methods

2.1 Data collection

Transcriptomic data from Atlantic salmon liver were provided by researchers involved with the GenoSysFat project (NFR 244164), and downloaded from the project's page on FAIRDOMHub, along with corresponding sample metadata. A detailed description of the feeding trial can be found in Gillard et al. (2018). Briefly, Atlantic salmon fry was reared in freshwater tanks and continuously fed on one of two diets, based on either vegetable oil (VO), a combination of linseed oil and palm oil, or marine oil (MA), containing only North Atlantic fish oil. After 26 weeks, a feed switch took place, with sampling occurring at specific time points. Subsequently, smoltification was triggered and a feed switch was repeated in saltwater. A subset of each sample was used for RNA extraction and sequencing.

2.2 The Atlantic salmon GEM

The reconstructed Atlantic salmon GEM (SALARECON) (Zakhartsev et al., 2021) contains 718 reactions, 530 metabolites and 1,104 genes, which corresponds to 2% of the genes in its genome and 48% of Atlantic salmong metabolic genes from KEGG (Kanehisa et al., 2019). The compounds are divided into five compartments that are connected by transport reactions. To allow prediction of growth, a biomass reaction based on whole-body composition (Hatlen et al., 2015) is included.

2.3 Pre-processing of transcriptomic data

The raw gene expression data was first reduced to only contain genes that were also existent in the model. Subsequently, a gene expression threshold was set to determine gene activity in any of the samples, and any gene with an activity score above this threshold was defined as active. Each gene was given an individual threshold equal to the 90^{th} percentile of its expression value across all samples in the dataset, as this has been documented to yield better models than lower threshold values (Opdam et al., 2017). The 25^{th} percentile

of the overall gene expression value distribution (i.e. all genes in all samples) was used to deal with lowly expressed genes. To this end, it was set as the threshold for any gene with a threshold lower than this percentile. The gene score was then computed as follows (Richelle et al., 2019b):

$$\text{Gene score} = 5\ln\left(1 + \frac{\text{Expression level}}{\text{Threshold}}\right) \tag{2.1}$$

By parsing the gene-protein-reaction rules (GPRs) associated with each reaction, the gene scores were mapped to SALARECON. As this model only contains genes associated with isozymes (OR rules), the maximum score of all genes involved in a reaction was selected as the reaction's activity level (RAL).

2.4 Extracting context-specific models

Six different model extraction methods were used to extract sample-specific context-specific models from SALARECON, specifically FASTCORE (Vlassis et al., 2014), GIMME (Becker and Palsson, 2008), iMAT (Shlomi et al., 2008; Zur et al., 2010), INIT (Agren et al., 2012), MBA (Jerby et al., 2010) and mCADRE (Wang et al., 2012). The algorithms of these MEMs are implemented in the COBRA Toolbox 3.0, a MATLAB software suite for constraint-based modelling (Heirendt et al., 2019). Using the function *createTissueSpecificModel*, sample-specific models were extracted from SALARECON. The values of the required parameters needed to execute the different MEMs were set equal to the values described by Richelle et al. (2019b) (Table 1.1). As the biomass reaction is not directly associated with any genes, it was handled manually in the utilisation of each MEM to ensure its inclusion in the extracted model. To this end, the lower bound of the biomass reaction flux was set to one. Additionally, the reaction was manually added to the core reaction set parameters of FASTCORE and MBA, assigned a gene score greater than the threshold for GIMME and iMAT, as well as assigned a specific weight for INIT.

The implementation of the mCADRE algorithm did not perform as expected, as it stopped with an error if removing a reaction led to an infeasible solution, instead of continuing its model reduction process. To overcome this, I took advantage of the fact that the COBRA Toolbox is version-controlled using Git; a free and open-source distributed version control system that tracks changes in computer files. The MATLAB code was manually edited before raising an issue on the COBRA Toolbox's repository page. The contribution was later merged into the development branch (commit be52c22) and subsequently into the master branch of the repository (commit 6c1ba69).

2.5 Metabolic task activity

A curated and standardised list of 210 metabolic tasks (MTs) covering seven major metabolic activities of a cell (energy generation, nucleotide, carbohydrate, amino acid, lipid, vitamin and cofactor and glycan metabolism) was downloaded from Richelle et al. (2019b). Subsequently, their framework for predicting task activity from transcriptomics data was used. This framework utilises the GPRs of the model to compute which reactions are responsible for performing each metabolic task, and further which genes are involved in the specific reactions, ultimately being able to link the metabolic tasks to transcriptomics data. Hence, reactions and associated genes responsible for executing each metabolic task were determined using parsimonious flux balance analysis (pFBA) as described by Lewis et al. (2010b) and the GPRs of the model. The enzyme production and abundance of these genes are naturally dependent on their expression levels, intuitively making some genes more predominant than others for a specific reaction. Consequently, one single gene may be selected as the most important for a particular reaction. Thus, each reaction was also attributed an activity level, based on the expression level of the associated main determinant gene. As the expression levels of the genes are the underlying information of which gene activity scores were computed, the RAL was defined to correspond to the main determinant gene's activity score. Ultimately, the MT scores were calculated as the mean of the activity levels of the reactions involved in the specific task:

$$\text{MT score} = \frac{\sum(\text{RAL})}{\text{Number of reactions involved in the task}} \tag{2.2}$$

To be consistent with the gene score computation that both the context-specific models and the MT scores are based on, a threshold of $5\ln(2)$ was set to define whether or not a metabolic task was active in a particular sample (Richelle et al., 2019b).

2.6 Analysing MEM performance

Using each of the six different MEMs and the gene expression data, context-specific models were extracted from SALARECON. In this regard, each sample of the transcriptomic dataset was considered as a context. Subsequently, the context-specific models' ability to perform the metabolic tasks were tested and compared. To this end, the template GEM's task performance was first tested, removing the infeasible tasks from the list, thus only testing the extracted models' task performance on the tasks that they potentially could be able to perform. From the reduced task list, a binary matrix containing information on task feasibility was constructed, in which each row represents a model and each column represents a task, with entries of 1 (feasible) or 0 (infeasible). In addition, a binary MT score

matrix was computed from the reduced task list and the transcriptomic data, defining the expected functionality of each sample in the data. Ultimately, the MT scores were used to assess the MEMs capabilities to create models with a metabolic functionality that concurred with the transcriptomic data, overcoming this previously challenging part of MEM comparisons. To this end, the Hamming distance was calculated between two binary vectors, one representing a particular model's ability to perform the metabolic tasks, the other the MT scores of the corresponding sample from which the model was extracted. Furthermore, principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) was performed to assess the impacts of MEM, feed type and life stage. For this, models not able to perform any of the tasks, i.e. rows of all zeros, were removed from the matrix, which subsequently was centred to have zero mean within each row. Additionally, the relative contributions of each factor level to the first three principal components were calculated and scaled to percentages. PCA and PLS-DA were then repeated on a binary matrix containing information on reaction presence, in which rows represents models and columns represents reactions, with entries of 1 (present) or 0 (absent).

3. Results

3.1 Model extraction and contents

Using the transcriptomic dataset, each MEM was used to extract 208 context-specific models (one for each sample in the dataset) from SALARECON, resulting in a total of 1,248 models. Four mCADRE models were non-functional, i.e. they contained too few genes, reactions and metabolites to be able to perform any tasks. The remaining 1,244 models varied greatly in contents and functionality, both between and within the used extraction methods. The number of retained reactions in the extracted models varied from 162 as a minimum to 714 as a maximum (Fig. 3.1). Investigating presence of specific reactions, genes and metabolites, models extracted using the same MEM were more similar to each other than to the other models (Figs. A.1-A.3).

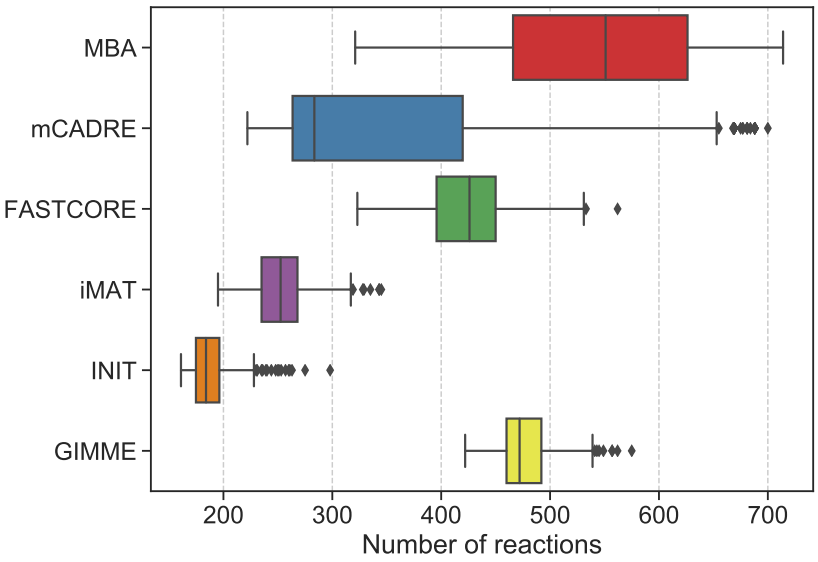


Figure 3.1: Number of reactions in extracted models. The number of retained reactions in a model varies between extraction methods, but there is also large variability within MEMs. The MEMs are sorted by family, from MBA-like to iMAT-like and ending with GIMME. Non-functional models are not included in the plot.

3.2 PCA and PLS-DA

The impacts of the three factors MEM, feed type and life stage on reaction inclusion and model functionality differed. For the PCA of reaction inclusion in the extracted models, explained variability of the first and second principal components was 14.8% and 8.9%, respectively (Fig. 3.2). GIMME explained most of the variation within the first component, while INIT explained the majority of the variance in the second component (Fig. 3.2a,b). For the other factors, there was no clear separation between the levels, i.e. between the four feed types and between the two life stages, and there was not much difference in explained variability between them within the first three principal components (Fig. 3.2c-f). The second principal component explained much of the variation seen in model inclusion of certain reactions, while the first two components explained approximately half of this variation for several other reactions (Fig. 3.2g). 200 components were required to explain approximately 90% of the variability in the reaction inclusion of the extracted models (Fig. 3.2h).

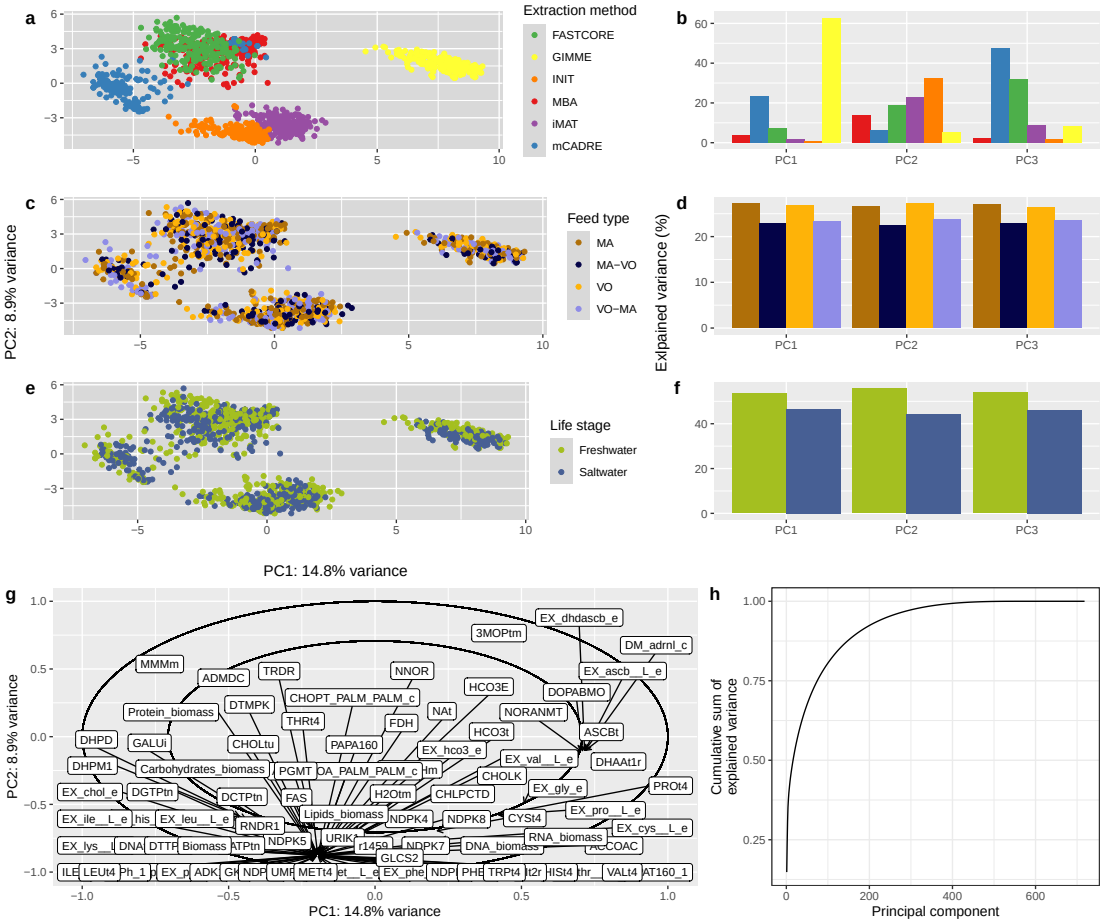


Figure 3.2: PCA on reaction presence. (**a and b**) Extraction method influences both PCs, with GIMME being very important for the first PC. (**c–f**) Feed type and life stage influence on the PCs is more variable. (**g**) A correlation loadings plot shows that the first two PCs explain much of the variance in several reactions. (**h**) Approx. 90% of the variation is explained by 200 PCs.

The variation in reaction inclusion between the context-specific models was well separated by MEM, with the first and second PLS components explaining 14.6% and 13.1% of it, respectively (Fig. 3.3a). Models extracted using the same method were clustered together, albeit with some variability. Similarly to the PCA, between 50% and 100% of the variability in reaction inclusion of several reactions was explained by these PLS components (Fig. 3.3b). Although the models were in some form separated, the levels of the other factors were not separated into distinct clusters (Fig. 3.3c,e). However, PLS components of life stage separation better explained the observed variance in reaction inclusion between the extracted models for some reactions (Fig. 3.3d,f). Additionally, the overall explained variance was lower for these factors, with more variation being explained by the second PLS component than the first.

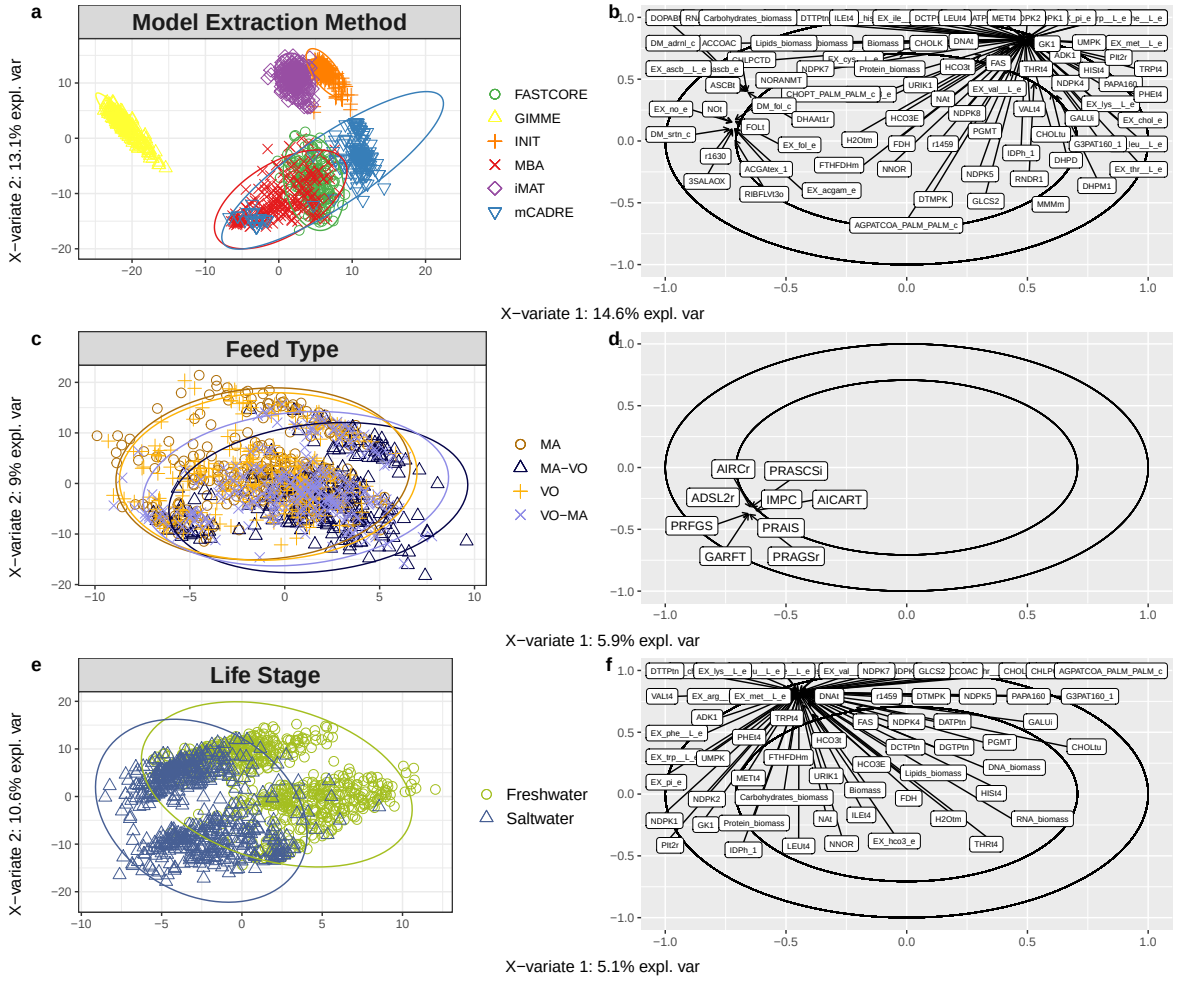


Figure 3.3: PLS-DA on reaction presence. (**a and b**) The models are well clustered together by MEM, although the separation of mCADRE models is less compact. The PLS-DA separates the models into the correct MEM families. (**b**) The variance in several reactions are almost fully explained by the first two PLS components for MEM separation. (**c–f**) Although highly overlapping, reaction presence in the models can be somewhat separated by life stage, but not as well by feed type. Variation in some reactions is better explained by the life stage PLS components.

For the PCA of model functionality of the extracted models, the explained variation in task feasibility between the context-specific models was 9.4% and 8.7% for the first and second principal components, respectively (Fig. 3.4). Within these components, the majority of the variance was explained by FASTCORE and GIMME (Fig. 3.4a,b). As for the reaction inclusion PCA, there was no clear separation between the levels in the other factors, with only small differences in explained variability between them within the first three principal components (Fig. 3.4c-f). Less than 50% of task-specific variance in feasibility between context-specific models was explained by the first two principal components for the majority of the metabolic tasks (Fig. 3.4g). 65 components were required to explain approximately 90% of the variance in task feasibility between the extracted models (Fig. 3.4h).

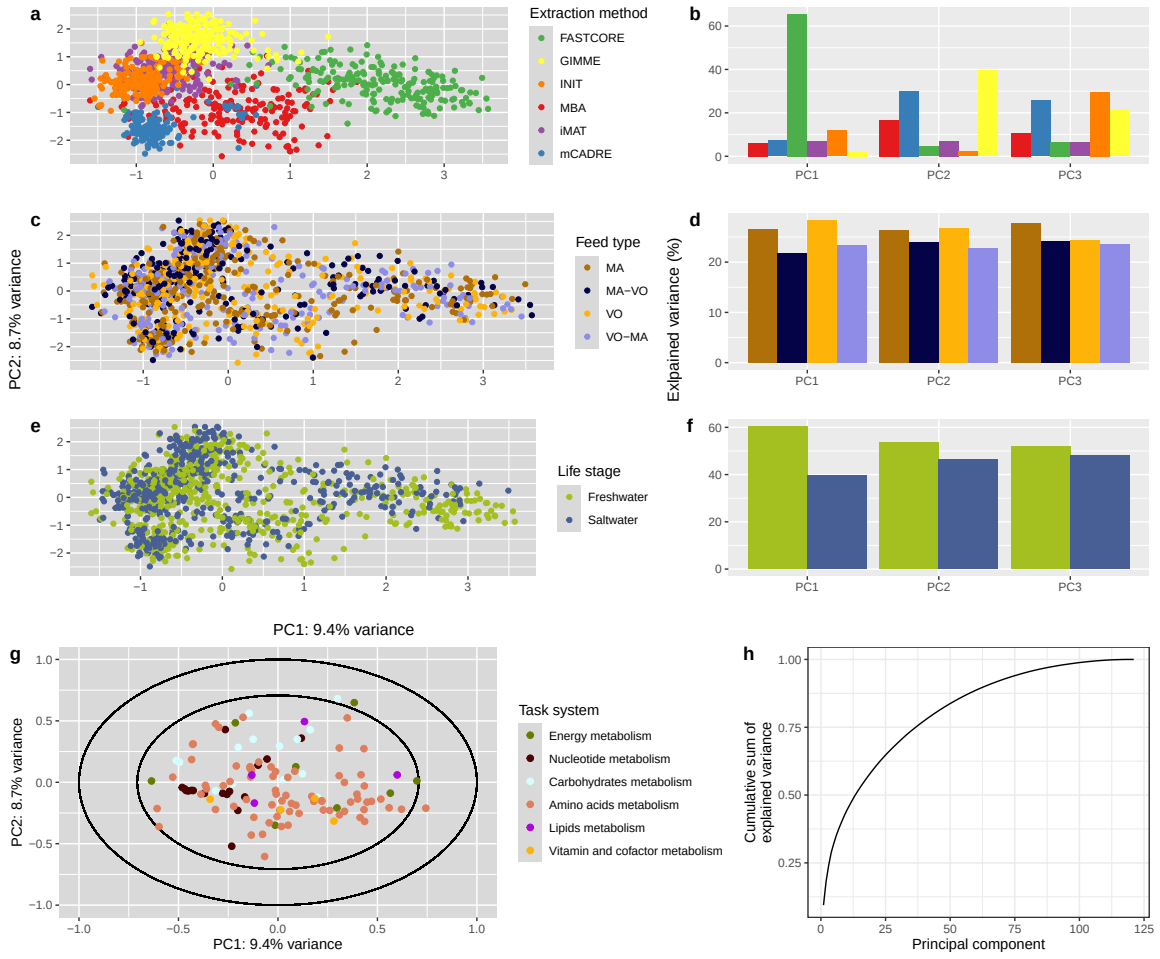


Figure 3.4: PCA on task feasibility. (**a and b**) The influence of extraction method is clear in the first PC, with FASTCORE being the method contributing most to the PC. (**c–f**) Feed type and life stage influence on the PCs is more variable. (**g**) The first two PCs explain >50% of a few tasks involved in energy, amino acids and carbohydrate metabolism. (**h**) Approx. 90% of the variation is explained by about 60-65 PCs.

Task feasibility variance in the context-specific models was better separated by MEM than feed type or life stage, similarly to reaction inclusion variance (Fig. 3.5). For the extraction method factor, the first two PLS components explained 10.5% and 6.8% of the variability, and although there was some overlap, the models were clustered together into MEM used (Fig. 3.5a). These two components explained more than or equal to 50% of the variability in the feasibility of only two tasks, one involved in amino acids metabolism, the other in carbohydrates metabolism (Fig. 3.5b). Likewise to the reaction inclusion matrix, the levels of the other factors were not as well separated into distinct clusters. However, life stages were better separated than feed types (Fig. 3.5c,e). Moreover, these factors explained less variance than the analysis on MEM, with more variation being explained by the second PLS component than the first. Furthermore, several tasks positively correlated with the first component and negatively with the second for life stage separation, while the opposite was observed for feed type separation (Fig. 3.5d,f).

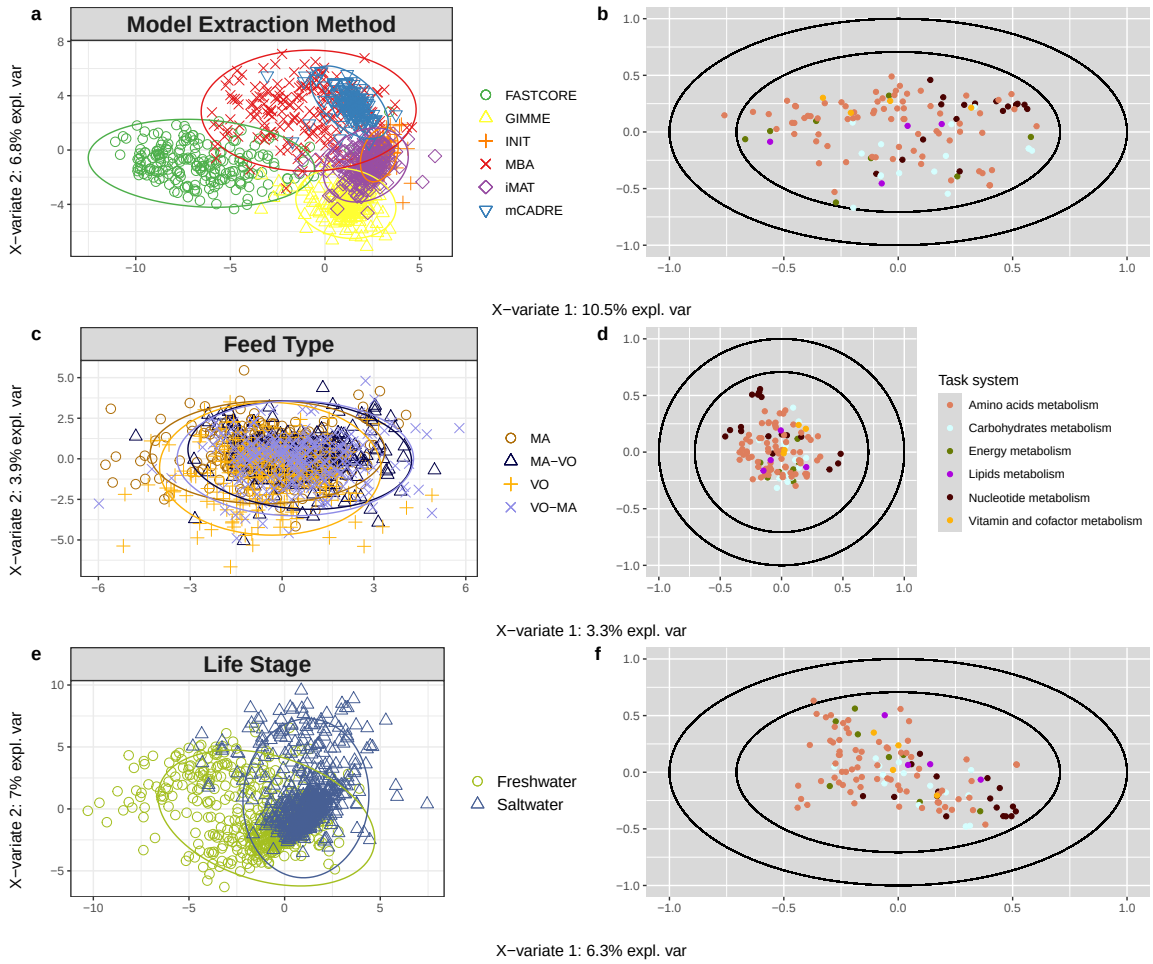


Figure 3.5: PLS-DA on task feasibility. (**a and b**) The models are well clustered together by MEM and separated well by family. (**b**) About 50% of feasibility variance of some tasks is explained by two PLS components for MEM separation. (**c–f**) The discriminant analysis separates life stage better than feed type. Variation in some tasks is better explained by the life stage PLS components.

3.3 Metabolic functionality of context-specific models

121 of the 210 metabolic tasks passed in SALARECON, comprising six task systems, i.e. metabolic activities. These tasks are hereafter referred to as all tasks. For the context-specific models, metabolic functionality correlated with number of retained reactions (Fig. 3.6). The extracted models were also more similar within than between MEMs, except for some MBA and mCADRE models in which a higher percentage of the tasks were feasible than for the rest of the models (Fig. 3.7). Within the six task systems, the number of context-specific models that were capable to perform the system-specific tasks differed between MEMs and varied between systems (Fig. 3.8). Comparing MEM families, the iMAT-like family were more consistent than other families in creating models with similar capabilities for most of the task systems, i.e. the variability in these models' metabolic functionality was smaller.

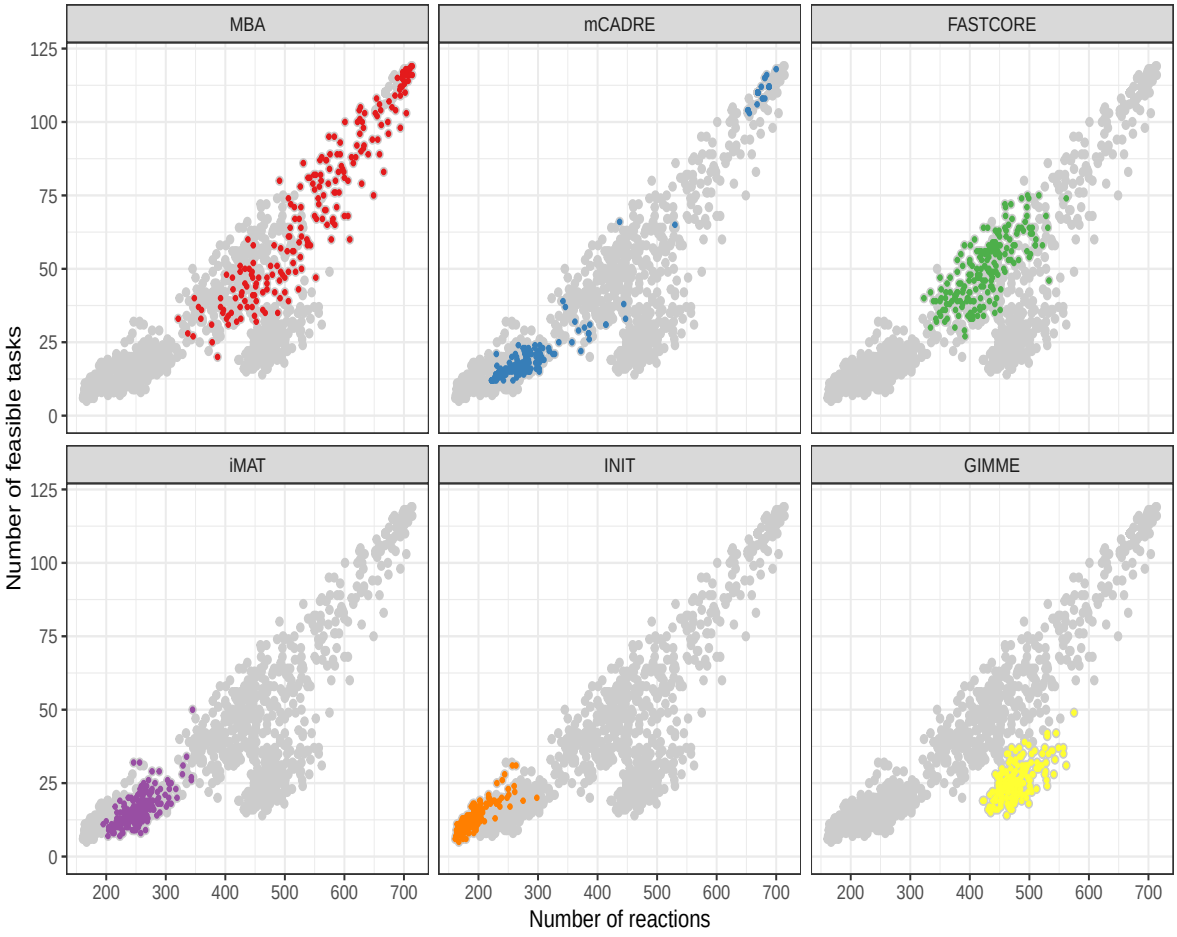


Figure 3.6: Task performance correlates with number of reactions. Metabolic functionality increases with number of retained reactions. The variation in reaction inclusion differs between extraction methods. Non-functional models are not included in the plot.

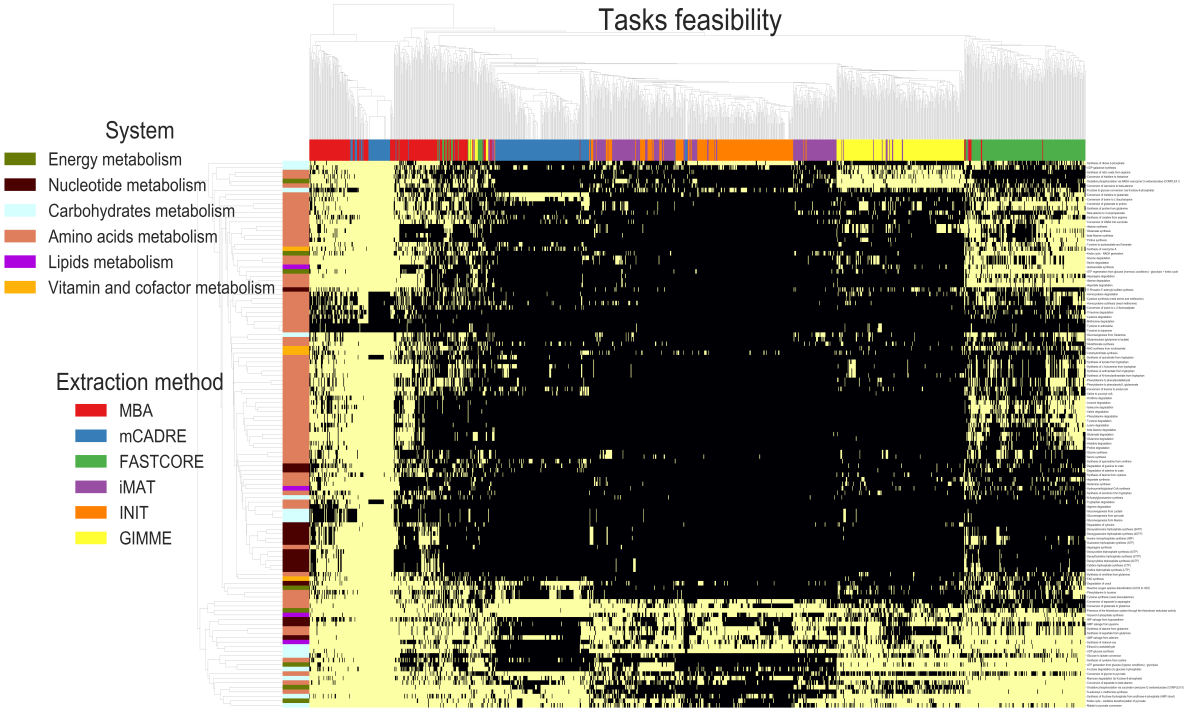


Figure 3.7: MBA models have a higher fraction of feasible tasks than the other models. Some tasks are commonly or seldomly feasible independent of the extraction method. MEM is overall well clustered together, at least within families.

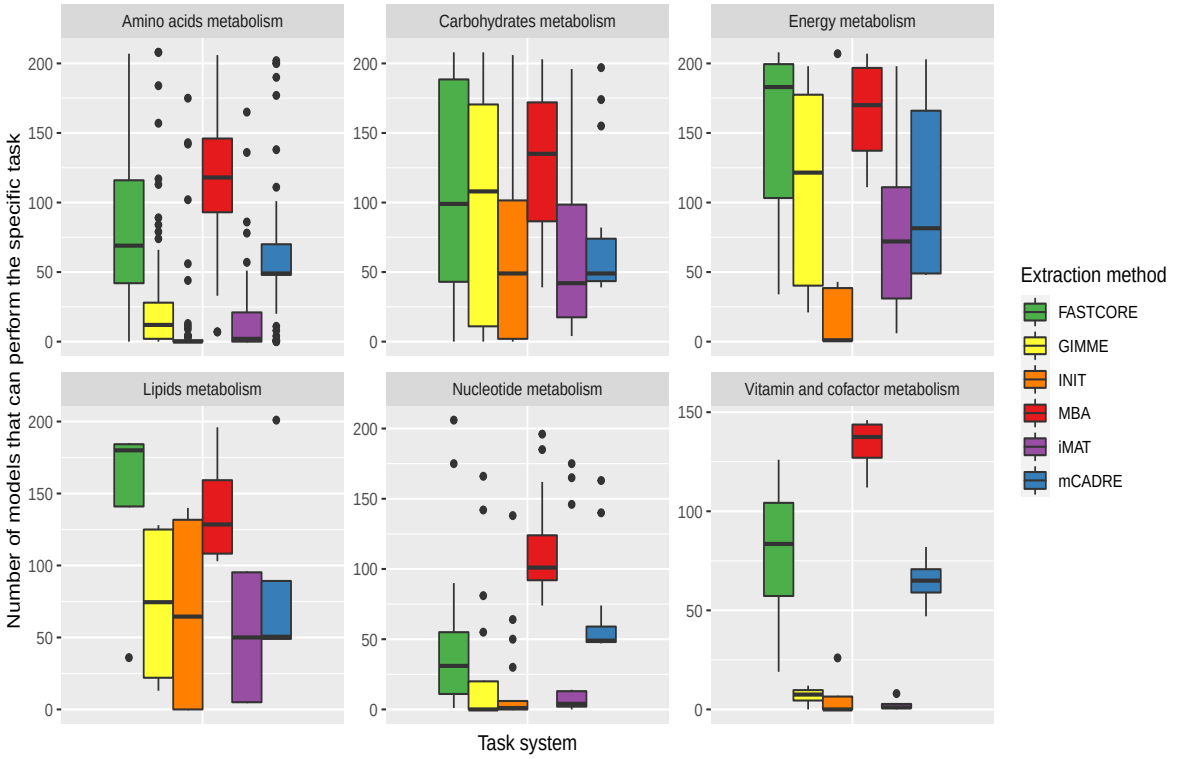


Figure 3.8: Number of models in which tasks pass by system. There is large variability in the number of models that can perform tasks in each system, both within and between MEMs.

Model functionality was compared with expected functionality based on the transcriptomic data. The Hamming distance between passed tasks in a model and the computed binary MT scores for the corresponding sample was generally shorter for the context-specific models than for SALARECON (Fig. 3.9a). The shortest distance was observed for iMAT, INIT and GIMME models, of which the iMAT models had the highest proportion of models with Hamming distance less than 0.2. However, some of the models extracted with these methods also overlapped with the distance of several samples of GEM assessment. A cluster analysis revealed that the overlap was indeed the same samples and that these lead to approximately equal Hamming distances across all MEMs and SALARECON (Fig. 3.9b). Comparing other samples across all MEMs, there was some variability in distance between them. However, two additional clusters, albeit comprising fewer samples, produced models with similar Hamming distances. Although MEMs were clustered together with other methods in the same family based on the distance, the agreement between expected and model functionality varied between samples, especially true for MBA and mCADRE models. FASTCORE models were the most consistent with each other, but their distance was longer than those of the majority of iMAT, INIT and GIMME models. The other factors, i.e. feed type and life stage, were not well clustered together in the samples.

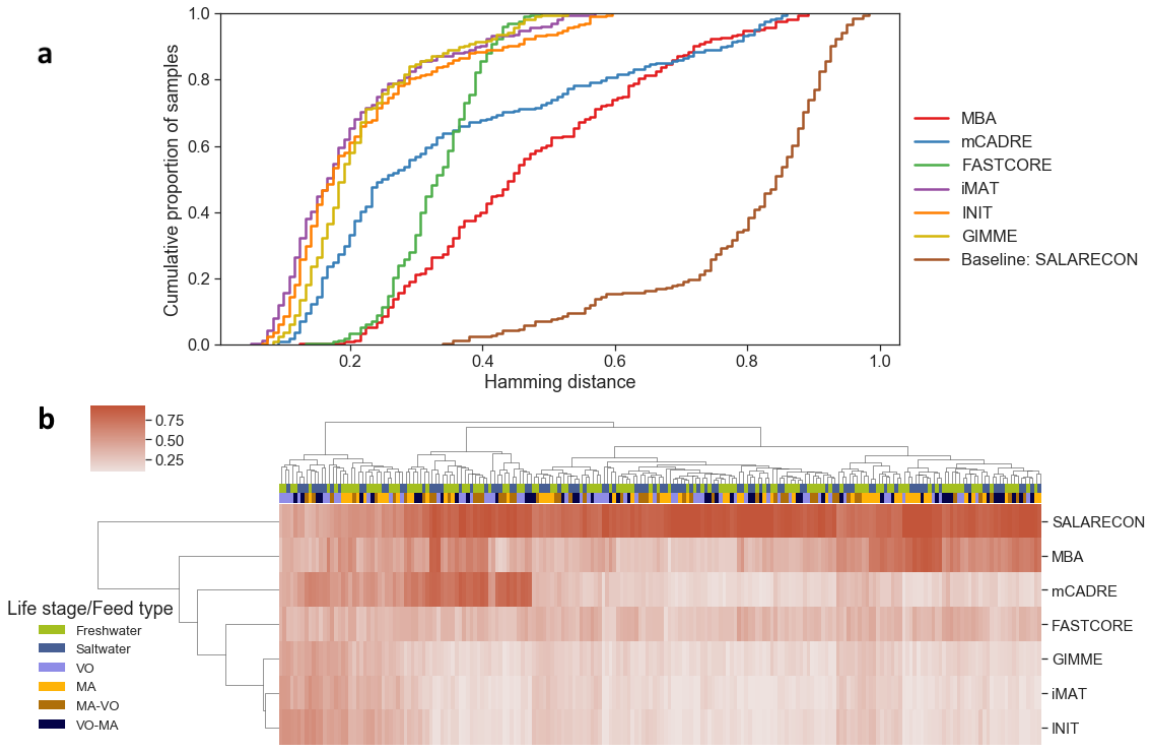


Figure 3.9: Hamming distance between passed tasks and computed binary MT scores. (**a**) Context-specific model functionality corresponds better with expected functionality from transcriptomic data than SALARECON. (**b**) Expected task feasibility is more accurate for models extracted with all MEMs. The models are also clustering correctly within their families. However, the factors life stage and feed type are not well clustered together. The Hamming distance is approximately equal in some samples for all MEMs and SALARECON.

4. Discussion

In this study, I assessed the capability of six model extraction methods to create accurate context-specific models using a hepatic transcriptomic dataset. Each context-specific models' ability to perform metabolic tasks was compared with the expected task activity of the particular sample it was extracted from. This method of comparing the MEMs is based on the newest literature, extrapolating current methods for predicting sample-specific task activities (Richelle et al., 2019b, 2020) to overcome the particular challenge of not having an objective truth to benchmark against in MEM comparisons. The cluster analysis showed that all the evaluated MEMs generally extract models with a metabolic functionality that concurs better with that of the liver than SALARECON (Fig. 3.9b). This indicates that context-specific models outperform full GEMs with regard to capturing realistic representations of metabolism in a given context and thus yield practical and biologically meaningful predictions, suggesting their advantage when studying the metabolic behaviour of organisms for both academic and industrial purposes.

Context-specific models seem to outperform GEMs due to differences in enzyme activity between contexts, e.g. tissues or conditions such as environmental changes or developmental stages. These differences change network structure and inactivate certain reactions, making GEMs superfluous (Robaina Estévez and Nikoloski, 2014; Pacheco et al., 2015). This is especially the case for multicellular organisms, where metabolic functionality and objective differs between organs, tissues and cell types and interact with each other (Lewis et al., 2010a; Robaina Estévez and Nikoloski, 2014). Consequently, the use of GEMs for context-specific analysis may lead to inferior predictions compared with submodels where inactive reactions have been removed (Opdam et al., 2017). Therefore, methods to reduce the scope of GEMs to be concordant with omics data are proving useful, especially considering the accessibility of transcriptomic data. In practice, context-specific models have been reported to yield more accurate predictions of genotype-phenotype relationships of four human cell lines, as they were able to accurately predict known metabolic functionalities and better predict gene-essentiality than the template GEM (Opdam et al., 2017). Additionally, confining GEMs to specific tissues or to be in accordance with a given context have been demonstrated to generate predictions more consistent with ex-

perimental data than the unconstrained model and be able to identify important aspects of the organism's metabolism for various applications (Chang et al., 2010; Gille et al., 2010; Bordbar et al., 2012; Shen et al., 2016; Fouladiha and Marashi, 2017; Siriwach et al., 2020). These findings indicate MEM superiority over GEMs, especially for multicellular organisms, corroborating the results of this study.

Between the implemented MEMs, iMAT had the highest proportion of models with the shortest Hamming distance between passed tasks and the computed MT scores and therefore was the best performing method, closely followed by INIT and GIMME (Fig. 3.9a). These results suggest that the iMAT and GIMME-like families may be better suited for analysing context-specific metabolism of multicellular organisms like the Atlantic salmon when transcriptomic data is available. iMAT was developed with only the integration of transcript profiles in mind, whereas INIT allows for the integration of additional proteomic and metabolomic data. As only transcriptomic data was used in the analyses of this study, these MEM attributes may explain why iMAT was the slightly better performer of the two. INIT may, however, have performed better if other data types had been included in the analyses. Additionally, iMAT has performed better when only one expression threshold was used to identify highly and lowly expressed reactions, removing moderately expressed states (Jamialahmadi et al., 2019), supporting the discoveries of this study where such an approach was also used. A reason why the median distance for GIMME was somewhat higher than the other two methods may be its inherent demand for defining a metabolic objective, as demonstrated previously (Machado and Herrgård, 2014). For multicellular organisms, the metabolic objectives of tissues may depend on the surrounding environment (Robaina Estévez and Nikoloski, 2014), i.e. the context, thus complicating the definition of an objective function for use with GIMME. Here, maximum growth rate was used as the objective, which may be reasonable for the rearing salmon from which the transcriptomic data was sampled, and may explain the relatively low Hamming distance of the GIMME models. Nevertheless, the definition of a tissue-specific objective could possibly have improved the models further, as has been suggested for humans (Chang et al., 2010; Gille et al., 2010; Toroghi et al., 2016).

The poor performance of the MBA-like family seen in this study may be explained by only integrating one data type in the extraction process. The full utility of these MEMs lies in the possibility of including complimentary pieces of experimental evidence to define the core sets, and ultimately increase the confidence of reaction presence in the context of interest (Robaina Estévez and Nikoloski, 2014). Consequently, optimal results require incorporating multiple data types, and as only transcriptomic data was used in this study to define the input parameters, it may explain the observed outcome. The median values of Hamming distances amongst these MEMs are quite dissimilar, with the MBA method performing much more poorly than the others. One property of its algorithm is the pruning

of non-core reactions in random order, inevitably affecting the topology of the final model (Jerby et al., 2010). Therefore, the algorithm should be run many times to get a population of context-specific models from which a consensus model can be extracted, e.g. 1000 times as in the implementation of Jerby et al. (2010). However, this is both very time-consuming and cumbersome. Consequently, no consensus population of models were created with MBA and only one model was extracted from each sample in the present study.

The choice of MEM had a larger impact on model contents and functionality than both feed type and water salinity, i.e. life stage. This is evident by the observed variability between the extracted models, which was highest for choice of MEM (Fig. 3.2 and Fig. 3.4). Similar tendencies have been reported in prior studies (Ferreira et al., 2017; Opdam et al., 2017; Richelle et al., 2019b). Well-separated clusters of each MEM are formed, caused by similarities and differences of algorithms between and within MEM families, indicating that MEM selection, or at least choice of MEM family, is the most important of the considered factors with regard to model content and functionality. In addition to the PCA and PLS-DA plots, this is also evident in Fig. 3.7, Fig. 3.8 and Figs. A.1-A.3. Not unexpectedly, the impact of MEM selection is most apparent in reaction content. The algorithms employ different strategies to remove redundant reactions, but do not consider task feasibility information in this process. This is also apparent in Fig. 3.2g and Fig. 3.4g , where >50% of observed variation in only three tasks (2.5%) can be explained by the first two PCs, while this number is a lot higher for reactions (11.7%). Additionally, a higher proportion of principal components is needed to explain the same amount of variation in task feasibility than reaction inclusion (Fig. 3.2h and Fig. 3.4h). Interestingly, GIMME tremendously impacts PC1 in reaction content, and as seen from the correlation loadings plot, variation in the inclusion of a lot of exchange reactions are explained by the PCs. Upon further investigation, all GIMME models contained all exchange reactions, explaining the mentioned pattern. The reason for this is unclear. However, GIMME removes reactions associated with expression levels below a user-defined threshold, provided that the flux through the objective function stays above a set fraction of that of the template GEM (Becker and Palsson, 2008). In this case, exchange reactions are necessary to supply the metabolic network with metabolites needed to grow. Thus, the algorithmic strategy of GIMME, requiring the formulation and maintenance of a metabolic objective, might cause the observed inclusion of exchange reactions in the GIMME models.

The supervised separation analysis of the extracted models supports the findings that MEM choice has the most impact on the extracted models compared to the other investigated factors (Fig. 3.3 and Fig. 3.5). Comparably to the PCA, the PLS components explain a lot more of the intra-reaction variance than intra-task variation for all tested factors, as task

feasibility information is not a part of the extraction process. This is further shown in the cluster separation of the factors, especially the separation of MEMs. The MEM families form decently separated clusters, but the FASTCORE models stand out, forming their own cluster. In contrast to its other family members, i.e. MBA and mCADRE, the algorithm of FASTCORE does not eliminate non-core reactions and check for consistency at each step, but rather solves two LP problems. To this end, it uses the reaction cardinalities, i.e. the number of non-zero fluxes in the reaction flux vectors. In an alternate fashion, FASTCORE maximises the cardinality of the core reaction set, before minimising the cardinality of the non-core set. This is repeated until all reactions in the core set are able to carry flux, supported by a minimum number of non-core reactions. The approach taken by FASTCORE may lead to different reactions being removed, ultimately impacting which tasks pass in the resulting model. For feed types and life stages, the separation is moderate at best. However, variation in reaction presence of models still results in the formation of clusters. As the models used to separate these factors are extracted using six different MEMs, differences in algorithmic strategies between them lead to differences in reaction presence of the models. Furthermore, the context-specific analyses of this study are based on transcriptomic data from a study that focused on different contexts affecting lipid metabolism. However, due to the poor representation of lipids in SALARECON, it may not capture any potential variability in reaction activity between the investigated feed types. Therefore, the impact of feed type on model content and functionality is limited in this study. In reality, however, feed type may affect these model attributes more than what is observed in this study.

Model functionality seems to correlate with number of retained reactions, albeit with large variability between MEMs (Fig. 3.6). Intuitively, the more reactions a model retains, the more alike it is to the template GEM, facilitating task performance. Similar patterns has also been documented in former research (Correia et al., 2018; Richelle et al., 2019b). This further highlights the impact of algorithm choice on model functionality. Most within-MEM variability was seen in the MBA-like family (MBA, mCADRE and FASTCORE), suggesting that their shared feature, i.e. the definition of core reaction sets, might be the causing factor. While some mCADRE models can be categorised as outliers, there is considerable variance between MBA models. MBA groups the core reactions into a high and a medium confidence set, thus adding an extra factor the user has to account for when defining the parameters. In this study, sample-specific thresholds were used for adding reactions into these reaction sets, which may have contributed to the observed variation. Moreover, the upper threshold was set to the 75^{th} percentile of the sample-specific gene scores. Opdam et al. (2017), however, registered more accurate models with more stringent thresholds, albeit for gene-essentiality predictions. Hence, less variation might be expected from setting a more strict threshold than was used here. As previously mentioned,

all exchange reactions were included in the GIMME models, yet they are not able to perform as many tasks as other models with the same number of reactions, suggesting that exchange reactions are not crucial for performing the tasks of the applied task list.

Although there is increasing evidence that only a subset of metabolic reactions in a network is active in any given context (Robaina Estévez and Nikoloski, 2014), context-specific modelling is challenging and MEM applicability and use have been questioned on several occasions. Despite being shown to be able to provide accurate insights into the metabolic state of organisms in specific contexts (Shen et al., 2016; Fouladiha and Marashi, 2017; Jamialahmadi et al., 2019), it is yet unclear how to tackle each individual reconstruction problem to ensure optimal results, a question subject to numerous factors.

The number of methods for creating context-specific models from omics data has increased in the last couple of decades (Machado and Herrgård, 2014; Jamialahmadi et al., 2019) and, realistically, more methods will be developed in the coming years. Individual advantages and limitations of these methods make selecting one for each specific extraction problem a challenging task (Blazier and Papin, 2012; Robaina Estévez and Nikoloski, 2014). As observed in this study, the choice of MEM has an immense impact on both the functionality and contents of the extracted models. Model variability between MEMs has similarly been documented previously. Correia and Rocha (2015) reported large differences in model functionality between three MEMs in different data sources. Furthermore, huge variability in reaction content between models that had been created using the same input has been observed (Pacheco et al., 2015). In another study, context-specific models of healthy liver and hepatocellular carcinoma (HCC) cells were generated using four different MEMs and analysed functionally and structurally (Ferreira et al., 2017). The authors observed large differences in models between the methods and concluded that although they can provide accurate insights into tissue- or cell-specific metabolism, none of the examined algorithms was ideal. This supports the results of Machado and Herrgård (2014), showing that none of the investigated methods outperforms the others in terms of flux predictions compared to experimental data. Several additional studies aiming to benchmark algorithms for creating context-specific models have recently been published. Pacheco et al. (2015) proposed a benchmarking method to aid in identifying weaknesses of both old and new MEMs in order to increase the quality of their algorithms and created models. In a different study, the impacts of algorithm choice, expression thresholds and metabolic constraints were systematically evaluated on the predictive abilities of cancer cell-specific models (Opdam et al., 2017). More recently, Jamialahmadi et al. (2019) analysed the functional and structural properties of several MEMs and identified which considerations to take into account to improve their functional performance. Furthermore, the authors developed a new method for generating context-specific models using their results, demonstrated its predictive capabilities, and provided guidelines for considering both weaknesses and strengths of exist-

ing methods when adopting their approach. Additionally, Robaina Estévez and Nikoloski (2014) categorised MEMs based on the mathematical formulation of their algorithms and provided a flowchart for choosing an appropriate method when facing a context-specific reconstruction problem.

Subsequent to choosing a suitable MEM when facing a context-specific reconstruction problem, the algorithm's input parameters need to be defined. Intuitively, the choice of parameter values affects the contents and functionality of the final model. Although the user may opt to leave all optional parameters to their default values, as done in this study, this may not always be optimal for each individual context that is investigated (Pacheco et al., 2015). Therefore, careful consideration has to be taken to determine how to best assign parameter values. Most important of the parameters is arguably the threshold defining the activity state of genes in the case of transcriptomic data. The choice of threshold has previously been documented to have the most impact of tested parameters on model contents and functionality and should therefore also be adjusted to the specific goals of the individual reconstruction problem (Opdam et al., 2017; Richelle et al., 2019a). However, systematically testing parameter settings for an optimum, through e.g. cross-validation, is cumbersome and computationally onerous, and may even pose overfitting problems (Pacheco et al., 2015). Indubitably, selecting both an appropriate MEM and its optimum parameter values given a particular context-specific reconstruction problem is a tremendously challenging task subject to multiple factors, and although a suggestion of what an ideal algorithm-parameter combination of methods have been proposed (Kim and Lun, 2014), no such MEMs exist. Therefore, thorough contemplation is required before choosing a method and defining its parameters for context-specific analyses.

In the present work, six MEMs were used to generate context-specific models from SALARECON using transcriptomic data. However, other alternatives may have yielded better results. It has been suggested that algorithm choice is less important for creating accurate context-specific models than data source (Correia et al., 2018). Omics data integration have been shown to lead to spaces of alternative optima, inducing ambiguous model predictions (Robaina Estévez and Nikoloski, 2017), which may be one explanation. Furthermore, many algorithms assume that the level of mRNA transcripts strongly correlates with protein activity, determining the activity of reactions based on transcript levels. However, the soundness of this assumption has been questioned on several occasions (Greenbaum et al., 2003; Rossell et al., 2006; Daran-Lapujade et al., 2007). Spearman's rank correlation coefficients ranging from 0.21 to 0.61 have been reported in earlier work (Ideker et al., 2001; Griffin et al., 2002; Washburn et al., 2003; Moxley et al., 2009), exhibiting only a modest relationship between transcript and protein levels. The observed discrepancy may be explained by factors such as post-translational modification, post-transcriptional regulation of protein synthesis and enzyme kinetics (Zhang et al., 2010). Moreover, the activity

level of transcriptional regulation on protein expression has been demonstrated to vary between pathways as a result of an evolutionary trade-off between being able to quickly react to changes in environmental conditions and minimizing the protein cost to achieve the appropriate response (Wessely et al., 2011). Protein levels should be more concordant with metabolic fluxes than gene expression levels (Robaina Estévez and Nikoloski, 2014). Nevertheless, context-specific analysis by using proteomic instead of transcriptomic data by Machado and Herrgård (2014) did not result in improved predictions. Last in the hierarchical cascade that regulates metabolic fluxes is metabolite levels, partaking in metabolic reactions, and should therefore provide more accurate snapshots of the metabolic state of a network (Robaina Estévez and Nikoloski, 2014). Despite current methods for measuring metabolite levels may not offer enough metabolome coverage to portray the metabolic state of the entire network (Fernie, 2007), integrating metabolomics data has been proven successful in context-specific analysis, especially in combination with other types of omics data (Yizhak et al., 2010; Kleessen et al., 2012). Consequently, integrating transcriptomic data alone may not be sufficient to create context-specific models that are functionally correct and of high quality (Pacheco et al., 2015; Correia et al., 2018). Thus, the full potential of the compared MEMs in extracting context-specific models may not have been reached in this study.

Above all, the reliability of a context-specific model predominantly depends on the quality of the GEM it is extracted from. Hence, shortcomings in GEM reconstructions is a huge limitation for context-specific modelling. Although automatic reconstruction tools exist, there are several problems with genome annotations and databases (**LU_2018**; Salzberg, 2019). Additionally, data availability varies for different organisms and complex organism-specific characteristics require manual assessment, leading to large variability in the quality of GEMs (Thiele and Palsson, 2010). Consequently, GEMs often include errors such as incorrect directionality definition (i.e. establishing the reversibility of reactions) and metabolic gaps (Di Filippo et al., 2016; Xu et al., 2020). Coupled with a lack of probabilistic factor representation and imperfect knowledge of biology and biochemistry, it is troublesome to expect absolute consistency between omics data and a network model (Hyduke et al., 2013). Additionally, algorithms generating context-specific models using omics data heavily relies on the GPR rules of the GEM. However, the translation of the GPRs in the reconstruction process has been shown to lead to errors in gap-filling and the exclusion of the molecular context used to create the model itself (Ponce-de-León et al., 2020). Despite this, context-specific modelling is still useful for hypothesis generation (Hyduke et al., 2013), and have proven successful in numerous applications (Fouladiha and Marashi, 2017).

5. Conclusions

Here, the metabolic functionality of the Atlantic salmon liver was derived from transcriptomic data and compared to the functionality of context-specific models extracted by six different MEMs. Through a new way of evaluating model quality, context-specific models were shown to more accurately capture metabolic functionality than the GEM, with the iMAT- and GIMME-like MEMs performing the best. The findings of this study corroborate current evidence that context-specific models better represent the actual metabolism of organisms in specific contexts and therefore should provide more practical and biologically meaningful predictions of an organism's metabolic behaviour. Thus, context-specific models should be advantageous when studying the metabolic behaviour of organisms, especially when investigating specific contexts of interest, and potentially be beneficial for both academic and industrial purposes. As the scope of GEMs has rapidly grown from studying microbial behaviour to biomedical and industrial applications, future work should aim to further develop and create new MEMs and identify the metabolic functionality of tissues in relevant organisms. In this regard, the panoply of omics data enables the advancement of algorithms, directing their focus to the integration of multiple omics data sources with GEMs to create improved context-specific models. Through this approach, context-specific models of higher quality will be generated, ultimately leading to more accurate and deeper insights into the metabolic behaviour of organisms in particular contexts.

References

Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I. and Nielsen, J. (2012). Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLOS Computational Biology* 8 (5): 1–9. DOI: 10.1371/journal.pcbi.1002518.

Banga, J. R. (2008). Optimization in computational systems biology. *BMC Systems Biology* 2: 47. DOI: 10.1186/1752-0509-2-47.

Beard, D. A. and Qian, H. (2005). Thermodynamic-based computational profiling of cellular regulatory control in hepatocyte metabolism. *American Journal of Physiology-Endocrinology and Metabolism* 288 (3): E633–44. DOI: 10.1152/ajpendo.00239.2004.

Becker, S. A. and Palsson, B. O. (2008). Context-specific metabolic networks are consistent with experiments. *PLOS Computational Biology* 4 (5): e1000082. DOI: 10.1371/journal.pcbi.1000082.

Blazier, A. S. and Papin, J. A. (2012). Integration of expression data in genome-scale metabolic network reconstructions. *Frontiers in Physiololgy* 3: 299. DOI: 10.3389/fphys.2012.00299.

Boissy, J., Aubin, J., Drissi, A., Werf, H. M. G. V. D., Bell, G. J. and Kaushik, S. J. (2011). Environmental impacts of plant-based salmonid diets at feed and farm scales. *Aquaculture* 321 (1-2): 61–70. DOI: 10.1016/j.aquaculture.2011.08.033.

Bordbar, A., Lewis, N. E., Schellenberger, J., Palsson, B. Ø. and Jamshidi, N. (2010). Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. *Molecular Systems Biology* 6 (1): 422. DOI: 10.1038/msb.2010.68.

Bordbar, A., Mo, M. L., Nakayasu, E. S., Schrimpe-Rutledge, A. C., Kim, Y.-M., Metz, T. O., Jones, M. B., Frank, B. C., Smith, R. D., Peterson, S. N., Hyduke, D. R., Adkins, J. N. and Palsson, B. O. (2012). Model-driven multi-omic data analysis elucidates metabolic immunomodulators of macrophage activation. *Molecular Systems Biology* 8: 558. DOI: 10.1038/msb.2012.21.

Bordbar, A., Monk, J. M., King, Z. A. and Palsson, B. Ø. (2014). Constraint-based models predict metabolic and associated cellular functions. *Nature Reviews Genetics* 15 (2): 107–120. DOI: 10.1038/nrg3643.

Buchholz, A., Hurlebaus, J., Wandrey, C. and Takors, R. (2002). Metabolomics: quantification of intracellular metabolite dynamics. *Biomolecular Engineering* 19 (1): 5–15. DOI: 10.1016/s1389-0344(02)00003-5.

Cazzaniga, P., Damiani, C., Besozzi, D., Colombo, R., Nobile, M. S., Gaglio, D., Pescini, D., Molinari, S., Mauri, G., Alberghina, L. and Vanoni, M. (2014). Computational strategies for a system-level understanding of metabolism. *Metabolites* 4 (4): 1034–87. DOI: 10.3390/metabo4041034.

Chang, R. L., Xie, L., Xie, L., Bourne, P. E. and Palsson, B. Ø. (2010). Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLOS Computational Biology* 6 (9): e1000938. DOI: 10.1371/journal.pcbi.1000938.

Correia, S. and Rocha, M. (2015). A critical evaluation of methods for the reconstruction of tissue-specific models. In: *Progress in artificial intelligence*. Ed. by F. Pereira, P. Machado, E. Costa and A. Cardoso. Cham: Springer International Publishing: 340–352. DOI: 10.1007/978-3-319-23485-4_35.

Correia, S., Costa, B. and Rocha, M. (2018). Reconstruction of consensus tissue-specific metabolic models. *bioRxiv*: 327262. DOI: 10.1101/327262.

Dal'Molin, C. G. d. O., Quek, L.-E., Palfreyman, R. W., Brumbley, S. M. and Nielsen, L. K. (2010). C4GEM, a genome-scale metabolic model to study C_4 plant metabolism. *Plant Physiology* 154 (4): 1871–85. DOI: 10.1104/pp.110.166488.

Daran-Lapujade, P., Rossell, S., van Gulik, W. M., Luttik, M. A. H., de Groot, M. J. L., Slijper, M., Heck, A. J. R., Daran, J.-M., de Winde, J. H., Westerhoff, H. V., Pronk, J. T. and Bakker, B. M. (2007). The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. *Proceedings of the National Academy of Sciences of the United States of America* 104 (40): 15753–8. DOI: 10.1073/pnas.0707476104.

Di Filippo, M., Colombo, R., Damiani, C., Pescini, D., Gaglio, D., Vanoni, M., Alberghina, L. and Mauri, G. (2016). Zooming-in on cancer metabolic rewiring with tissue specific constraint-based models. *Computational Biology and Chemistry* 62: 60–9. DOI: 10.1016/j.compbiolchem.2016.03.002.

Domingo, J. L., Bocio, A., Falcó, G. and Llobet, J. M. (2007). Benefits and risks of fish consumption Part I. A quantitative analysis of the intake of omega-3 fatty acids and chemical contaminants. *Toxicology* 230 (2-3): 219–26. DOI: 10.1016/j.tox.2006.11.054.

FAO (2018). *The state of world fisheries and aquaculture: meeting the sustainable development goals*. Tech. rep. Rome: Food and Agriculture Organization of the United Nations.

Feist, A. M. and Palsson, B. O. (2010). The biomass objective function. *Current Opinion in Microbiology* 13 (3): 344–9. DOI: 10.1016/j.mib.2010.03.003.

Feist, A. M. and Palsson, B. O. (2016). What do cells actually want? *Genome Biology* 17 (1): 110. DOI: 10.1186/s13059-016-0983-3.

Fernie, A. R. (2007). The future of metabolic phytochemistry: larger numbers of metabolites, higher resolution, greater understanding. *Phytochemistry* 68 (22-24): 2861–80. DOI: 10.1016/j.phytochem.2007.07.010.

Ferreira, J., Correia, S. and Rocha, M. (2017). Analysing algorithms and data sources for the tissue-specific reconstruction of liver healthy and cancer cells. *Interdisciplinary Sciences–Computational Life Sciences* 9 (1): 36–45. DOI: 10.1007/s12539-017-0214-y.

Fouladiha, H. and Marashi, S.-A. (2017). Biomedical applications of cell- and tissue-specific metabolic network models. *Journal of Biomedical Informatics* 68: 35–49. DOI: 10.1016/j.jbi.2017.02.014.

Frezza, C., Zheng, L., Folger, O., Rajagopalan, K. N., MacKenzie, E. D., Jerby, L., Micaroni, M., Chaneton, B., Adam, J., Hedley, A., Kalna, G., Tomlinson, I. P. M., Pollard, P. J., Watson, D. G., Deberardinis, R. J., Shlomi, T., Ruppin, E. and Gottlieb, E. (2011). Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature* 477 (7363): 225–228. DOI: 10.1038/nature10363.

Ganter, M., Bernard, T., Moretti, S., Stelling, J. and Pagni, M. (2013). MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics* 29 (6): 815–816. DOI: 10.1093/bioinformatics/btt036.

Gatlin III, D. M., Barrows, F. T., Brown, P., Dabrowski, K., Gaylord, T. G., Hardy, R. W., Herman, E., Hu, G., Krogdahl, Å., Nelson, R., Overturf, K., Rust, M., Sealey, W., Skonberg, D., J Souza, E., Stone, D., Wilson, R. and Wurtele, E. (2007). Expanding the utilization of sustainable plant products in

aquafeeds: a review. *Aquaculture Research* 38 (6): 551–579. DOI: 10.1111/j.1365-2109.2007.01704.x.

Gillard, G., Harvey, T. N., Gjuvsland, A., Jin, Y., Thomassen, M., Lien, S., Leaver, M., Torgersen, J. S., Hvidsten, T. R., Vik, J. O. and Sandve, S. R. (2018). Life-stage-associated remodelling of lipid metabolism regulation in Atlantic salmon. *Molecular Ecology* 27 (5): 1200–1213. DOI: 10.1111/mec.14533.

Gille, C., Bölling, C., Hoppe, A., Bulik, S., Hoffmann, S., Hübner, K., Karlstädt, A., Ganeshan, R., König, M., Rother, K., Weidlich, M., Behre, J. and Holzhütter, H.-G. (2010). HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Molecular Systems Biology* 6: 411. DOI: 10.1038/msb.2010.62.

Grafahrend-Belau, E., Junker, A., Eschenröder, A., Müller, J., Schreiber, F. and Junker, B. H. (2013). Multiscale metabolic modeling: dynamic flux balance analysis on a whole-plant scale. *Plant Physiology* 163 (2): 637–47. DOI: 10.1104/pp.113.224006.

Greenbaum, D., Colangelo, C., Williams, K. and Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology* 4 (9): 117. DOI: 10.1186/gb-2003-4-9-117.

Griffin, T. J., Gygi, S. P., Ideker, T., Rist, B., Eng, J., Hood, L. and Aebersold, R. (2002). Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Molecular & Cellular Proteomics* 1 (4): 323–33. DOI: 10.1074/mcp.m200001-mcp200.

Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. and Lee, S. Y. (2019). Current status and applications of genome-scale metabolic models. *Genome Biology* 20 (1): 1–18. DOI: 10.1186/s13059-019-1730-3.

Hatlen, B., Jakobsen, J. V., Crampton, V., Alm, M., Langmyhr, E., Espe, M., Hevrøy, E. M., Torstensen, B. E., Liland, N. and Waagbø, R. (2015). Growth, feed utilization and endocrine responses in Atlantic salmon (S*almo salar*) fed diets added poultry by-product meal and blood meal in combination with poultry oil. *Aquaculture Nutrition* 21 (5): 714–725. DOI: 10.1111/anu.12194.

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdóttir, H. S., Wachowiak, J., Keating, S. M., Vlasov, V., Magnusdóttir, S., Ng, C. Y., Preciat, G., Žagare, A., Chan, S. H. J., Aurich, M. K., Clancy, C. M., Modamio, J., Sauls, J. T., Noronha, A., Bordbar, A., Cousins, B., El Assal, D. C., Valcarcel, L. V., Apaolaza, I., Ghaderi, S., Ahookhosh, M., Ben Guebila, M., Kostromins, A., Sompairac, N., Le, H. M., Ma, D., Sun, Y., Wang, L., Yurkovich, J. T., Oliveira, M. A. P., Vuong, P. T., El Assal, L. P., Kuperstein, I., Zinovyev, A., Hinton, H. S., Bryant, W. A., Aragón Artacho, F. J., Planes, F. J., Stalidzans, E., Maass, A., Vempala, S., Hucka, M., Saunders, M. A., Maranas, C. D., Lewis, N. E., Sauter, T., Palsson, B. Ø., Thiele, I. and Fleming, R. M. T. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols* 14 (3): 639–702. DOI: 10.1038/s41596-018-0098-2.

Henderson, R. J. (1996). Fatty acid metabolism in freshwater fish with particular reference to polyunsaturated fatty acids. *Archiv für Tierernahrung* 49 (1): 5–22. DOI: 10.1080/17450399609381859.

Houston, R. D. and Macqueen, D. J. (2019). Atlantic salmon (S*almo salar* L.) genetics in the 21st century: taking leaps forward in aquaculture and biological understanding. *Animal Genetics* 50 (1): 3–14. DOI: 10.1111/age.12748.

Hyduke, D. R., Lewis, N. E. and Palsson, B. Ø. (2013). Analysis of omics data with genome-scale models of metabolism. *Molecular BioSystems* 9 (2): 167–74. DOI: 10.1039/c2mb25453k.

Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R. and Hood, L. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292 (5518): 929–34. DOI: 10.1126/science.292.5518.929.

Jamialahmadi, O., Hashemi-Najafabadi, S., Motamedian, E., Romeo, S. and Bagheri, F. (2019). A benchmark-driven approach to reconstruct metabolic networks for studying cancer metabolism. *PLOS Computational Biology* 15 (4): e1006936. DOI: 10.1371/journal.pcbi.1006936.

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., Turner, A. D., Townhill, B. L. and Verner-Jeffreys, D. W. (2016). 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. DOI: 10.1111/faf.12152.

Jerby, L., Shlomi, T. and Ruppin, E. (2010). Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Molecular Systems Biology* 6 (1): 401. DOI: 10.1038/msb.2010.56.

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. and Tanabe, M. (2019). New approach for understanding genome variations in KEGG. *Nucleic Acids Research* 47 (D1): D590–D595. DOI: 10.1093/nar/gky962.

Kim, M. K. and Lun, D. S. (2014). Methods for integration of transcriptomic data in genome-scale metabolic models. *Computational and Structural Biotechnology Journal* 11 (18): 59–65. DOI: 10.1016/j.csbj.2014.08.009.

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O. and Lewis, N. E. (2016). BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research* 44 (D1): D515–D522. DOI: 10.1093/nar/gkv1049.

Kleessen, S., Araújo, W. L., Fernie, A. R. and Nikoloski, Z. (2012). Model-based confirmation of alternative substrates of mitochondrial electron transport chain. *Journal of Biological Chemistry* 287 (14): 11122–31. DOI: 10.1074/jbc.M111.310383.

Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M. R., Ahn, T.-H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S. and Ussery, D. W. (2015). Insights from 20 years of bacterial genome sequencing. *Functional & Integrative Genomics* 15 (2): 141–61. DOI: 10.1007/s10142-015-0433-4.

Lewis, N. E., Schramm, G., Bordbar, A., Schellenberger, J., Andersen, M. P., Cheng, J. K., Patel, N., Yee, A., Lewis, R. A., Eils, R., König, R. and Palsson, B. Ø. (2010a). Large-scale *in silico* modeling of metabolic interactions between cell types in the human brain. *Nature Biotechnology* 28 (12): 1279–1285. DOI: 10.1038/nbt.1711.

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. and Palsson, B. Ø. (2010b). Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology* 6: 390. DOI: 10.1038/msb.2010.47.

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., 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., Jonassen, I., Maass, A., Omholt, S. W. and Davidson, W. S. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* 533 (7602): 200–205. DOI: 10.1038/nature17164.

Machado, D. and Herrgård, M. (2014). Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLOS Computational Biology* 10 (4): e1003580. DOI: 10.1371/journal.pcbi.1003580.

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 Rd, 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. and FAASG Consortium (2017). Functional Annotation of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. *BMC Genomics* 18 (1): 484. DOI: 10.1186/s12864-017-3862-8.

Maier, T., Schmidt, A., Güell, M., Kühner, S., Gavin, A.-C., Aebersold, R. and Serrano, L. (2011). Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Molecular Systems Biology* 7 (1): 511. DOI: 10.1038/msb.2011.38.

Mendoza, S. N., Olivier, B. G., Molenaar, D. and Teusink, B. (2019). A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biology* 20 (1): 1–20. DOI: 10.1186/s13059-019-1769-1.

Montañez, R., Medina, M. A., Solé, R. V. and Rodríguez-Caso, C. (2010). When metabolism meets topology: reconciling metabolite and reaction networks. *Bioessays* 32 (3): 246–56. DOI: 10.1002/bies.200900145.

Moxley, J. F., Jewett, M. C., Antoniewicz, M. R., Villas-Boas, S. G., Alper, H., Wheeler, R. T., Tong, L., Hinnebusch, A. G., Ideker, T., Nielsen, J. and Stephanopoulos, G. (2009). Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. *Proceedings of the National Academy of Sciences of the United States of America* 106 (16): 6477–82. DOI: 10.1073/pnas.0811091106.

Mundheim, H., Aksnes, A. and Hope, B. (2004). Growth, feed efficiency and digestibility in salmon (S*almo salar* L.) fed different dietary proportions of vegetable protein sources in combination with two fish meal qualities. *Aquaculture* 237 (1): 315–331. DOI: 10.1016/j.aquaculture.2004.03.011.

Nam, H., Conrad, T. M. and Lewis, N. E. (2011). The role of cellular objectives and selective pressures in metabolic pathway evolution. *Current Opinion in Biotechnology* 22 (4): 595–600. DOI: 10.1016/j.copbio.2011.03.006.

Nichols, P. D., Glencross, B., Petrie, J. R. and Singh, S. P. (2014). Readily available sources of long-chain omega-3 oils: is farmed Australian seafood a better source of the good oil than wild-caught seafood? *Nutrients* 6 (3): 1063–79. DOI: 10.3390/nu6031063.

Opdam, S., Richelle, A., Kellman, B., Li, S., Zielinski, D. C. and Lewis, N. E. (2017). A systematic evaluation of methods for tailoring genome-scale metabolic models. *Cell Systems* 4 (3): 318–329.e6. DOI: 10.1016/j.cels.2017.01.010.

Orth, J. D., Palsson, B. Ø. and Fleming, R. M. T. (2010a). Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide. *EcoSal Plus* 4 (1). DOI: 10.1128/ecosalplus.10.2.1.

Orth, J. D., Thiele, I. and Palsson, B. O. (2010b). What is flux balance analysis? *Nature Biotechnology* 28 (3): 245–248. DOI: 10.1038/nbt.1614.

Pacheco, M. P., Pfau, T. and Sauter, T. (2015). Benchmarking procedures for high-throughput context specific reconstruction algorithms. *Frontiers in Physiololgy* 6: 410. DOI: 10.3389/fphys.2015.00410.

Palsson, B. Ø. (2015). *Systems biology: constraint-based reconstruction and analysis*. Cambridge University Press. DOI: 10.1017/CBO9781139854610.

Ponce-de-León, M., Apaolaza, I., Valencia, A. and Planes, F. J. (2020). On the inconsistent treatment of gene-protein-reaction rules in context-specific metabolic models. *Bioinformatics* 36 (6): 1986–1988. DOI: 10.1093/bioinformatics/btz832.

Price, N. D., Reed, J. L. and Palsson, B. Ø. (2004). Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology* 2 (11): 886–97. DOI: 10.1038/nrmicro1023.

Ramsden, C. E., Hibbeln, J. R., Majchrzak, S. F. and Davis, J. M. (2010). *n*-6 Fatty acid-specific and mixed polyunsaturate dietary interventions have different effects on CHD risk: a meta-analysis of randomised controlled trials. *British Journal of Nutrition* 104 (11): 1586–600. DOI: 10.1017/S0007114510004010.

Reed, J. L. and Palsson, B. Ø. (2004). Genome-scale *in silico* models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. *Genome Research* 14 (9): 1797–805. DOI: 10.1101/gr.2546004.

Richelle, A., Joshi, C. and Lewis, N. E. (2019a). Assessing key decisions for transcriptomic data integration in biochemical networks. *PLOS Computational Biology* 15 (7): e1007185. DOI: 10.1371/journal.pcbi.1007185.

Richelle, A., Chiang, A. W., Kuo, C.-C. and Lewis, N. E. (2019b). Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. *PLOS Computational Biology* 15 (4): e1006867. DOI: 10.1371/journal.pcbi.1006867.

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. and Lewis, N. E. (2020). What does your cell really do? Model-based assessment of mammalian cells metabolic functionalities using omics data. *bioRxiv*: 2020.04.26.057943. DOI: 10.1101/2020.04.26.057943.

Robaina Estévez, S. and Nikoloski, Z. (2014). Generalized framework for context-specific metabolic model extraction methods. *Frontiers in Plant Science* 5: 491. DOI: 10.3389/fpls.2014.00491.

Robaina Estévez, S. and Nikoloski, Z. (2017). On the effects of alternative optima in context-specific metabolic model predictions. *PLOS Computational Biology* 13 (5): e1005568. DOI: 10.1371/journal.pcbi.1005568.

Rossell, S., van der Weijden, C. C., Lindenbergh, A., van Tuijl, A., Francke, C., Bakker, B. M. and Westerhoff, H. V. (2006). Unraveling the complexity of flux regulation: a new method demonstrated for nutrient starvation in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* 103 (7): 2166–71. DOI: 10.1073/pnas.0509831103.

Salzberg, S. L. (2019). Next-generation genome annotation: we still struggle to get it right. *Genome Biology* 20 (1): 92. DOI: 10.1186/s13059-019-1715-2.

Sánchez, B. J., Li, F., Kerkhoven, E. J. and Nielsen, J. (2019). SLIMEr: probing flexibility of lipid metabolism in yeast with an improved constraint-based modeling framework. *BMC Systems Biology* 13 (1): 4. DOI: 10.1186/s12918-018-0673-8.

Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M. and Sauer, U. (2012). Multidimensional optimality of microbial metabolism. *Science* 336 (6081): 601–4. DOI: 10.1126/science.1216882.

Shamir, M., Bar-On, Y., Phillips, R. and Milo, R. (2016). Snapshot: timescales in cell biology. *Cell* 164 (6): 1302–1302.e1. DOI: 10.1016/j.cell.2016.02.058.

Shen, F., Li, J., Zhu, Y. and Wang, Z. (2016). Systematic investigation of metabolic reprogramming in different cancers based on tissue-specific metabolic models. *Journal of Bioinformatics and Computational Biology* 14 (5): 1644001. DOI: 10.1142/S0219720016440017.

Shendure, J. and Aiden, E. L. (2012). The expanding scope of DNA sequencing. *Nature Biotechnology* 30 (11): 1084–94. DOI: 10.1038/nbt.2421.

Shlomi, T., Cabili, M. N., Herrgård, M. J., Palsson, B. Ø. and Ruppin, E. (2008). Network-based prediction of human tissue-specific metabolism. *Nature Biotechnology* 26 (9): 1003–10. DOI: 10.1038/nbt.1487.

Siriwach, R., Matsuda, F., Yano, K. and Hirai, M. Y. (2020). Drought stress responses in context-specific genome-scale metabolic models of *Arabidopsis thaliana*. *Metabolites* 10 (4). DOI: 10.3390/metabo10040159.

Sulheim, S. (2020). Assembly and application of genome-scale metabolic models to study *Streptomyces coelicolor* and *Prochlorococcus*. PhD thesis. Norwegian University of Science and Technology (NTNU).

Thiele, I. and Palsson, B. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols* 5 (1): 93–121. DOI: 10.1038/nprot.2009.203.

Tibbetts, S. M. (2018). The potential for 'next-generation', microalgae-based feed ingredients for salmonid aquaculture in context of the blue revolution. In: *Fish Nutrition (Third Edition)*. Ed. by E. Jacob-Lopes, L. Q. Zepka and M. I. Queiroz. InTechOpen. Chap. 8: 151–175. DOI: 10.5772/intechopen.73551.

Tocher, D. R. (2003). Metabolism and functions of lipids and fatty acids in teleost fish. *Reviews in Fisheries Science* 11 (2): 107–184. DOI: 10.1080/713610925.

Toroghi, M. K., Cluett, W. R. and Mahadevan, R. (2016). A multi-scale model of the whole human body based on dynamic parsimonious flux balance analysis. *IFAC-PapersOnLine* 49 (7): 937–942. DOI: https://doi.org/10.1016/j.ifacol.2016.07.319.

Troell, M., Naylor, R. L., Metian, M., Beveridge, M., Tyedmers, P. H., Folke, C., Arrow, K. J., Barrett, S., Crépin, A.-S., Ehrlich, P. R., Gren, A., Kautsky, N., Levin, S. A., Nyborg, K., Österblom, H., Polasky, S., Scheffer, M., Walker, B. H., Xepapadeas, T. and de Zeeuw, A. (2014). Does aquaculture add resilience to the global food system? *Proceedings of the National Academy of Sciences of the United States of America* 111 (37): 13257–63. DOI: 10.1073/pnas.1404067111.

Turchini, G. M., Torstensen, B. E. and Ng, W.-K. (2009). Fish oil replacement in finfish nutrition. *Reviews in Aquaculture* 1 (1): 10–57. DOI: 10.1111/j.1753-5131.2008.01001.x.

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A.-K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., Alm, T., Edqvist, P.-H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J. and Pontén, F. (2015). Tissue-based map of the human proteome. *Science* 347 (6220): 1260419. DOI: 10.1126/science.1260419.

Varma, A. and Palsson, B. O. (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and Environmental Microbiology* 60 (10): 3724–31. DOI: 10.1128/AEM.60.10.3724-3731.1994.

Vlassis, N., Pacheco, M. P. and Sauter, T. (2014). Fast reconstruction of compact context-specific metabolic network models. *PLOS Computational Biology* 10 (1): e1003424–. DOI: 10.1371/journal.pcbi.1003424.

Wang, Y., Eddy, J. A. and Price, N. D. (2012). Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Systems Biology* 6 (1): 153. DOI: 10.1186/1752-0509-6-153.

Washburn, M. P., Koller, A., Oshiro, G., Ulaszek, R. R., Plouffe, D., Deciu, C., Winzeler, E. and Yates 3rd, J. R. (2003). Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* 100 (6): 3107–12. DOI: 10.1073/pnas.0634629100.

Wessely, F., Bartl, M., Guthke, R., Li, P., Schuster, S. and Kaleta, C. (2011). Optimal regulatory strategies for metabolic pathways in *Escherichia coli* depending on protein costs. *Molecular Systems Biology* 7: 515. DOI: 10.1038/msb.2011.46.

Xu, Y., Holic, R. and Hua, Q. (2020). Comparison and analysis of published genome-scale metabolic models of *Yarrowia lipolytica*. *Biotechnology and Bioprocess Engineering* 25 (1): 53–61. DOI: 10.1007/s12257-019-0208-1.

Yizhak, K., Benyamini, T., Liebermeister, W., Ruppin, E. and Shlomi, T. (2010). Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics* 26 (12): i255–60. DOI: 10.1093/bioinformatics/btq183.

Zakhartsev, M., Rotnes, F., Gulla, M., Øyås, O., van Dam, J. C. J., Suarez-Diez, M., Grammes, F., HafÞórsson, R. A., van Helvoirt, W., Koehorst, J. J., Schaap, P. J., Jin, Y., Mydland, L. T., Gjuvsland, A. B., Sandve, S. R., dos Santos, V. A. P. M. and Vik, J. O. (2021). SALARECON connects the Atlantic salmon genome to growth and feed efficiency. *bioRxiv*: 2021.06.03.446971. DOI: 10.1101/2021.06.03.446971.

Zhang, W., Li, F. and Nie, L. (2010). Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology* 156: 287–301. DOI: 10.1099/mic.0.034793-0.

Zhang, Z., Wang, J., Wang, J., Wang, J. and Li, Y. (2020). Estimate of the sequenced proportion of the global prokaryotic genome. *Microbiome* 8 (1): 134. DOI: 10.1186/s40168-020-00903-z.

Zomorrodi, A. R., Suthers, P. F., Ranganathan, S. and Maranas, C. D. (2012). Mathematical optimization applications in metabolic networks. *Metabolic Engineering* 14 (6): 672–86. DOI: 10.1016/j.ymben.2012.09.005.

Zur, H., Ruppin, E. and Shlomi, T. (2010). iMAT: an integrative metabolic analysis tool. *Bioinformatics* 26 (24): 3140–3142. DOI: 10.1093/bioinformatics/btq602.

Appendix A. Clustermap of model contents

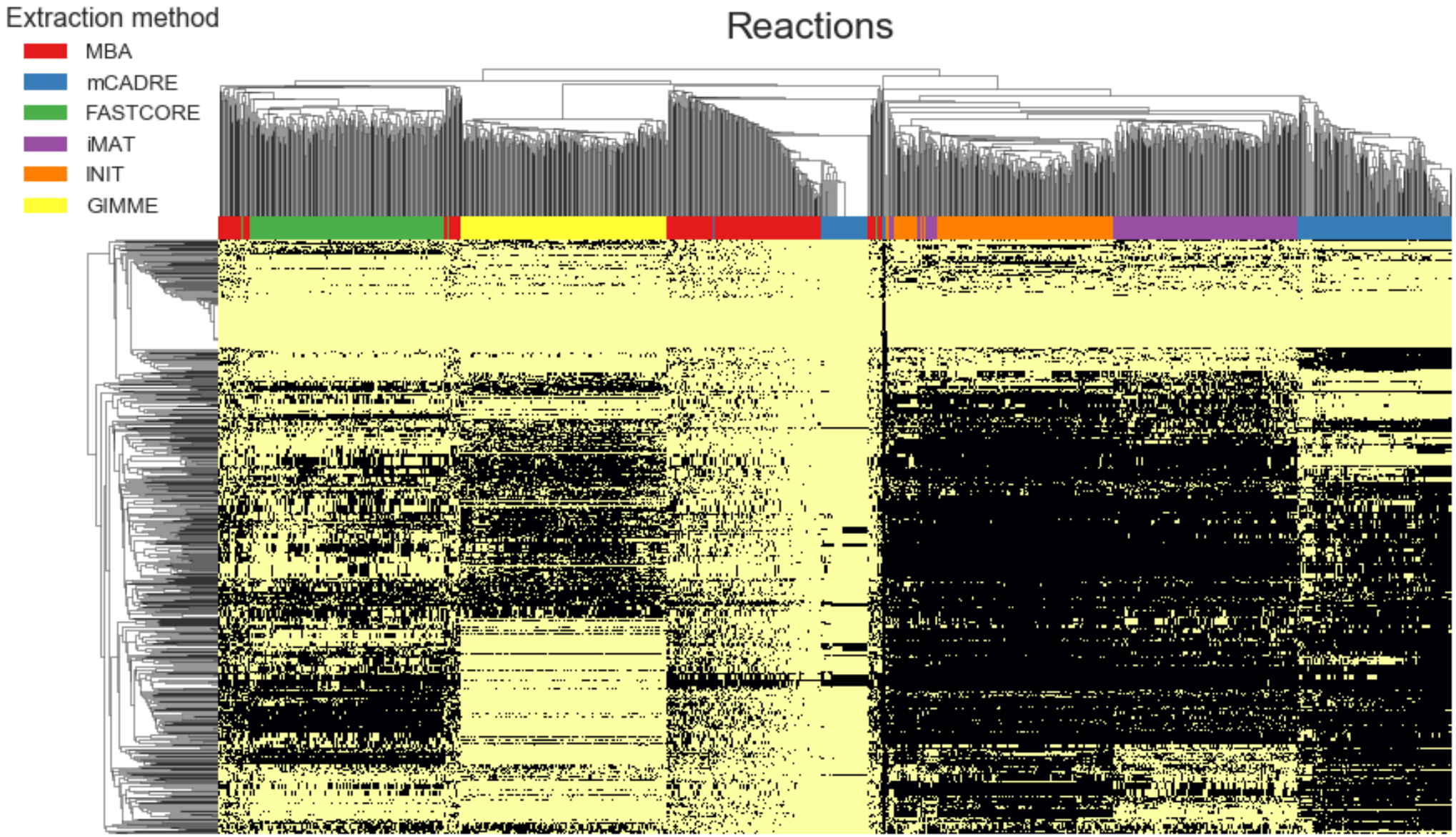


Figure A.1: Reaction inclusion in extracted models. The presence of reactions is similar for models extracted with the same model extraction method, and are generally well clustered together.

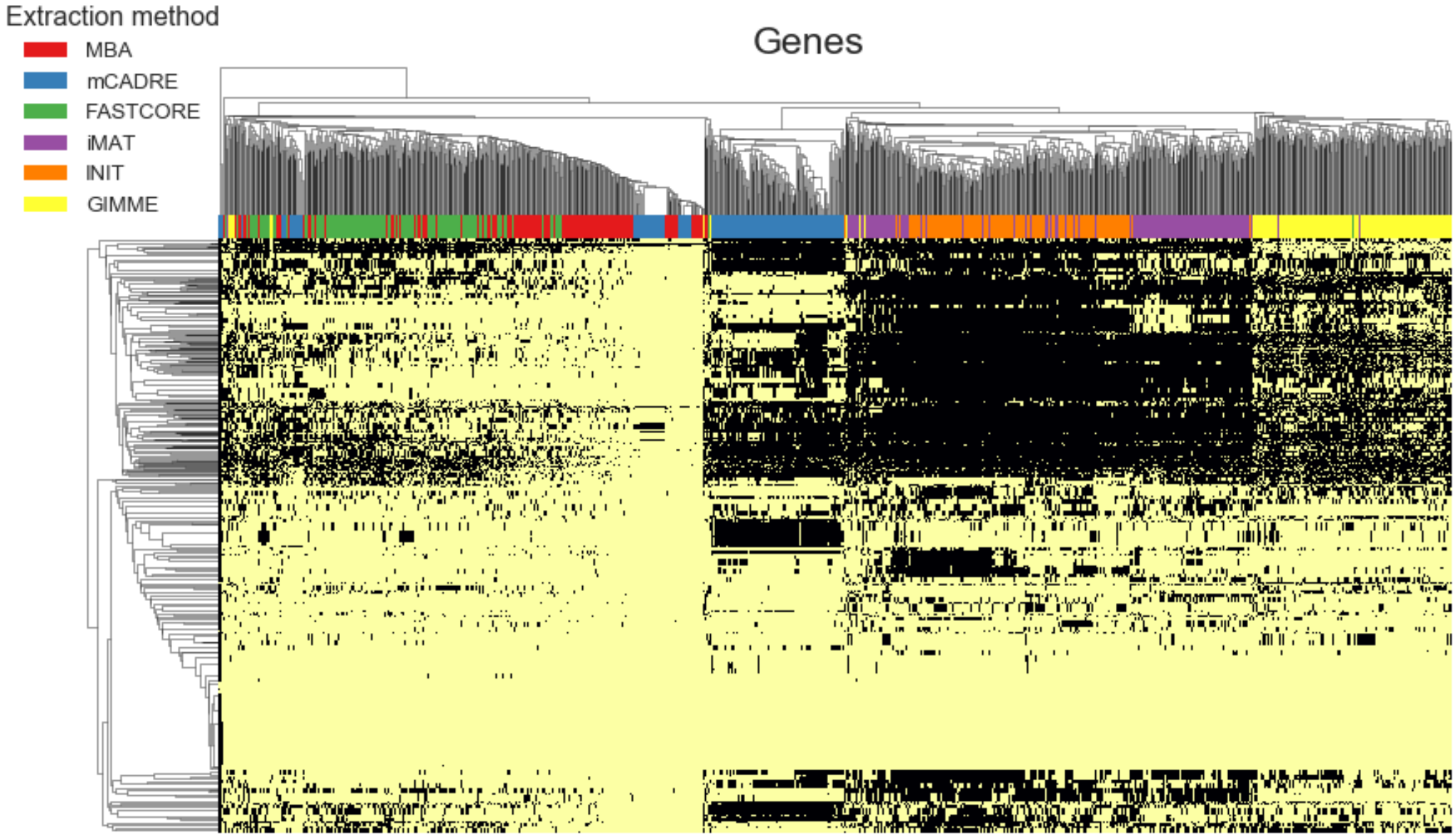


Figure A.2: Gene inclusion in extracted models. The presence of genes is similar for models extracted with the same model extraction method, and are generally well clustered together.

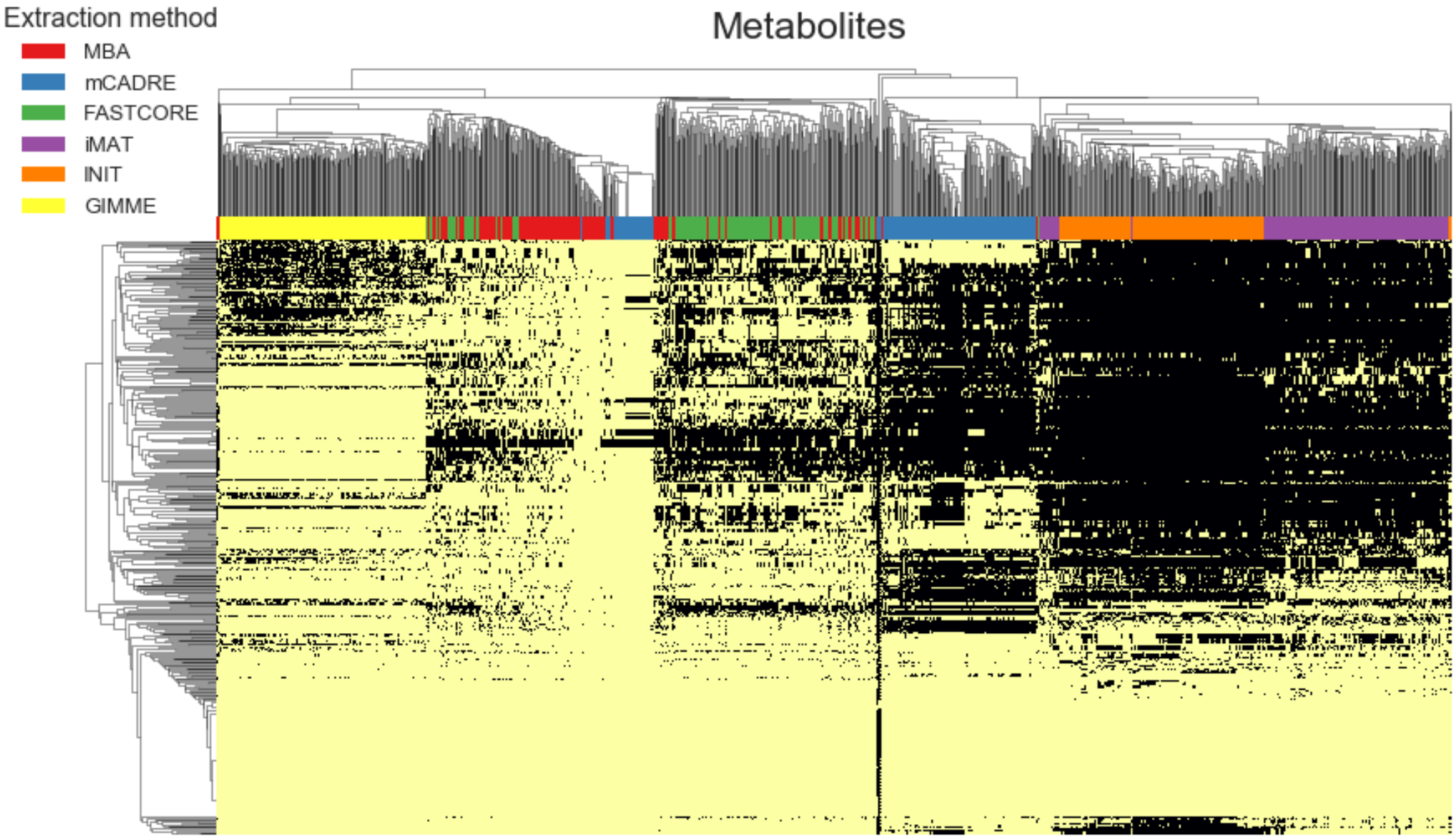
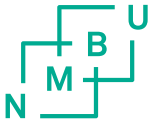


Figure A.3: Metabolite inclusion in extracted models. The presence of metabolites is similar for models extracted with the same model extraction method, and are generally well clustered together.



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway