



ReCodLiver0.9: Overcoming Challenges in Genome-Scale Metabolic Reconstruction of a Non-model Species

Eileen Marie Hanna^{1,2†}, Xiaokang Zhang^{2†}, Marta Eide³, Shirin Fallahi⁴, Tomasz Furmanek⁵, Fekadu Yadetie³, Daniel Craig Zielinski⁶, Anders Goksøyr³ and Inge Jonassen^{2*}

¹ Department of Computer Science and Mathematics, Lebanese American University, Byblos, Lebanon, ² Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway, ³ Department of Biological Sciences, University of Bergen, Bergen, Norway, ⁴ Department of Mathematics, University of Bergen, Bergen, Norway, ⁵ Institute of Marine Research, Bergen, Norway, ⁶ Department of Bioengineering, University of California, San Diego, La Jolla, CA, United States

OPEN ACCESS

Edited by:

Tim Young,
Auckland University of Technology,
New Zealand

Reviewed by:

Gregorio Peron,
Ca' Foscari University of Venice, Italy
Hunter N. B. Moseley,
University of Kentucky, United States

*Correspondence:

Inge Jonassen
inge.jonassen@uib.no

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Metabolomics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 04 August 2020

Accepted: 22 October 2020

Published: 26 November 2020

Citation:

Hanna EM, Zhang X, Eide M, Fallahi S, Furmanek T, Yadetie F, Zielinski DC, Goksøyr A and Jonassen I (2020) ReCodLiver0.9: Overcoming Challenges in Genome-Scale Metabolic Reconstruction of a Non-model Species. *Front. Mol. Biosci.* 7:591406. doi: 10.3389/fmolb.2020.591406

The availability of genome sequences, annotations, and knowledge of the biochemistry underlying metabolic transformations has led to the generation of metabolic network reconstructions for a wide range of organisms in bacteria, archaea, and eukaryotes. When modeled using mathematical representations, a reconstruction can simulate underlying genotype-phenotype relationships. Accordingly, genome-scale metabolic models (GEMs) can be used to predict the response of organisms to genetic and environmental variations. A bottom-up reconstruction procedure typically starts by generating a draft model from existing annotation data on a target organism. For model species, this part of the process can be straightforward, due to the abundant organism-specific biochemical data. However, the process becomes complicated for non-model less-annotated species. In this paper, we present a draft liver reconstruction, ReCodLiver0.9, of Atlantic cod (*Gadus morhua*), a non-model teleost fish, as a practicable guide for cases with comparably few resources. Although the reconstruction is considered a draft version, we show that it already has utility in elucidating metabolic response mechanisms to environmental toxicants by mapping gene expression data of exposure experiments to the resulting model.

Keywords: genome-scale metabolic reconstruction, Atlantic cod, less-annotated species, model curation, environmental toxicology

1. INTRODUCTION

Induced by the availability of next-generation DNA sequencing as well as high-throughput genomic, proteomic, and metabolomic data, genome-scale metabolic models (GEMs) have become fundamental tools in the systems biology of metabolism. To develop a GEM, metabolic genes and their product enzymes are assembled into a network of metabolic reactions carried out by an organism (Palsson, 2006). Such networks can be used as a platform to answer relevant biological questions, by transforming biochemical knowledge into a mathematical format and computing appropriate physiological states (Orth et al., 2010; Schellenberger et al., 2011). Using Boolean

logic representation, gene-protein-reaction (GPR) associations depict direct connections between genotype and metabolic capability (Nielsen, 2017). GEMs have a wide scope of applications, among which are the contextualization of omics data, hypothesis-driven discovery, support for metabolic engineering, modeling interactions among cells and organisms, drug targeting, and prediction of enzyme functions (Oberhardt et al., 2009; Gu et al., 2019).

A comprehensive bottom-up protocol to generate high-quality reconstructions was presented a decade ago (Thiele and Palsson, 2010). It consists of four main phases presented as an exhaustive list of steps. First, a draft reconstruction is generated based on the genome annotation of a target organism and available information in biochemical databases, usually in an automated manner. Noticeably, the quality of the draft model is highly impacted by the quality of the genome annotation, in addition to the abundance of data on the organism in public repositories. The draft is then subject to manual refinement that revises and curates all gene and reaction entries. Next, the reconstruction can be converted into mathematical format which is thereafter analyzed for gap filling and tested for stoichiometric balance. Several computational tools were developed to cover this process with automated functionalities (Gu et al., 2019). They primarily vary in programming languages, linked databases, input template models, and supported target organisms.

Marine organisms, including fish, represent important resources for human societies, and will be essential components of a growing blue bioeconomy. As we are entering the UN Decade of Ocean Science for Sustainable Development (Visbeck, 2018; Ryabinin et al., 2019), understanding and predicting the effects of environmental stressors, including xenobiotic exposures, on marine species will be one of the main challenges. In this context, metabolic reconstructions can prove very useful. Notwithstanding, when it comes to teleost GEMs, only two models are presented in the literature, MetaFishNet (Li et al., 2010) and ZebraGEM (Bekaert, 2012; van Steijn et al., 2019). MetaFishNet covers five fish species which makes it challenging to analyze organism-specific data using this model. In addition, subsystems in this GEM are not interconnected or compartmentalized. The first version of the ZebraGEM, a zebrafish GEM, did not contain GPR associations. A very recent version, ZebraGEM 2.0 (van Steijn et al., 2019), now includes those associations, with standardized component names, and expanded reactions list.

The Atlantic cod (*Gadus morhua*) is a key species in the North Atlantic marine ecosystem. Its genome was first published in 2011 (Star et al., 2011) and has been followed up by more extensive genome sequencing, enabling novel genomic and transcriptomic assemblies (Torresen et al., 2017) as well as a Trinity assembly (Zhang et al., 2020) and a very recent assembly gadMor3.0 (GenBank assembly accession: GCA_902167405). These resources allow extensive gene mapping and bioinformatic analysis, such as the identification of the gene complement of the Atlantic cod cytochrome P450 superfamily (Karlsen et al., 2012) or the total chemical defensome (Zhang et al., 2019). Due to its habitat near offshore installations as well as coastal industries and communities, the Atlantic cod serves as a relevant model

species for studying the effects of and responses to anthropogenic activities (Dale et al., 2019). In addition, the lipid-rich liver of cod serves as a magnet for lipophilic environmental pollutants, increasing the bioaccumulation within each fish with age and the biomagnification of compounds through the food chain.

The Atlantic cod has lost the MHC II and related genes of the adaptive immune system (Star et al., 2011). Furthermore, we have recently shown that cod, together with several other teleost fish, have lost the pregnane x receptor (*pxr*, *nr1i2*) gene from their genomes (Eide et al., 2018). This xenosensor is activated by a wide range of drugs and pollutants, and regulates the transcription of genes involved in all phases of the xenobiotic and steroid biotransformation (Kliwer et al., 2002). PXR also mediates regulation of hepatic energy metabolism in humans, by inducing lipogenesis and inhibiting fatty acid beta-oxidation (Wada et al., 2009). The absence of a *pxr* gene thus leads to questions regarding how the Atlantic cod responds to environmental contaminants that bind this receptor in other species. Results from an *in vivo* exposure study, where cod was injected the legacy polychlorinated biphenyl congener PCB153, a compound known to bind PXR in other species, showed indications that lipid metabolism was affected (Yadatie et al., 2017). In view of these findings, an ensuing effort was initiated to generate a metabolic reconstruction of Atlantic cod liver, with a focus on lipid and xenobiotic metabolism pathways. In this paper, we present and discuss our metabolic reconstruction progress in terms of steps, methods, choice of tools, and challenges.

2. METHODS

Selecting the Appropriate Tool

Many computational tools have been developed to help automate steps of the GEM reconstruction process which is usually laborious and time-consuming (Thiele and Palsson, 2010). The choice of tool for our reconstruction work on Atlantic cod was predominantly narrowed down by the fact that this is a non-model and less-annotated organism. *De novo* reconstruction based on metabolic databases do not apply in such cases, but rather GEMs of related scope needs to serve as input. Another important feature for us was the tool's support for eukaryote modeling. In view of these constraints, the Reconstruction, Analysis, and Visualization of Metabolic Networks (RAVEN 2.0) MATLAB toolbox (Wang et al., 2018) was selected. It is used for semi-automated draft model reconstruction and provides curation, simulation, and constraint-based capabilities that are compatible with the COBRA toolbox. RAVEN has been used to generate GEMs for several organisms including bacteria, archaea, parasites, fungi, and various human tissues. We also note that the toolbox allows *de novo* draft reconstruction based on metabolic pathway databases, namely KEGG and MetaCyc, although that it is not applicable to our case study. Most importantly, RAVEN can generate draft models for a target organism based on protein homology, using existing high-quality GEMs of organisms at an appropriate evolutionary distance. Two main functions are applied in this process. `getBlast` or `getBlastFromExcel` construct a structure including homology measurements between the target organism and the template

organisms. Then, `getModelFromHomology` uses this structure along with existing template GEMs to create a draft model containing reactions associated with orthologous genes. The subsequent draft is then subject to curation and refinement. The RAVEN toolbox includes several functions that are implemented to support subsequent steps.

We also highlight a few approaches. `CarveME` (Machado et al., 2018) uses a top-down approach, based on Python command-line, to convert curated reactions from the BiGG database into organism-specific GEMs. `AutoKEGGRec` (Karlsen et al., 2018) is a novel tool which is based on the KEGG pathways database (Aoki and Kanehisa, 2005) and is fully integrated with the COBRA toolbox (Schellenberger et al., 2011). `AuReMe` (Aite et al., 2018) presents a customized platform and pipelines to generate GEMs while preserving metadata and ensuring model traceability. It is also linked to the BiGG knowledgebase (Duarte et al., 2007) and the `MetaCyc` database (Caspi et al., 2013). For a thorough review of GEM reconstruction tools, we refer to the recent paper by Mendoza et al. (2019). The tools available for non-model species reconstructions (Henry et al., 2010; Boele et al., 2012; Hanemaaijer et al., 2017; Machado et al., 2018; Wang et al., 2018) are listed in **Supplementary Table S1**.

Choosing the Reconstruction Template

With the Atlantic cod liver as our reconstruction scope, we surveyed the literature for a relevant GEM that could be used as an input template to the RAVEN 2.0 toolbox. In the absence of tissue-specific models for fish species and the availability of a human liver GEM that greatly focuses on lipid metabolism, a primary trade-off was between using a generic template from a phylogenetically closer species, i.e., zebrafish, or using a liver-specific template from human that falls within our reconstruction scope. The `iHepatocytes2322` (Mardinoglu et al., 2014) consensus GEM was our template of choice. The reactions and associated proteins included in `iHepatocytes2322` were assigned into eight compartments. More details can be found in the original publication (Mardinoglu et al., 2014).

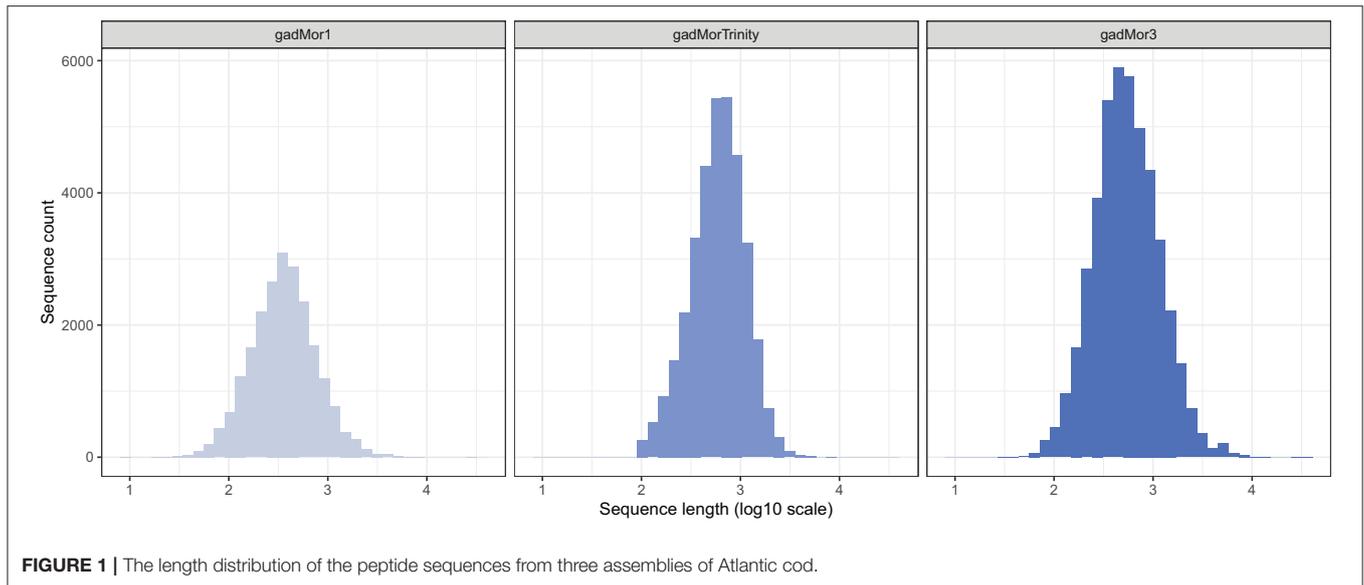
Of the two available fish GEMs, the `MetaFishNet` (Li et al., 2010) involves five fish species: zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), Takifugu (*Takifugu rubripes*), Tetraodon (*Tetraodon nigroviridis*), and stickleback (*Gasterosteus aculeatus*). Here, cDNA sequences from those species were analyzed and the corresponding list of metabolic genes was formed based on gene ontology. This list was then used to identify enzymes based on their orthology to human genes. Metabolic reactions were integrated from EHMN, the human metabolic network at UCSD (BiGG) (Duarte et al., 2007), and the *D. rerio* metabolic network from KEGG (Aoki and Kanehisa, 2005). The other fish model is `ZebraGEM` (Bekaert, 2012). It is a comprehensive GEM based on the literature and *D. rerio* biochemical data collected from various resources. This model is manually curated with gaps extensively checked and filled. An updated version, `ZebraGEM 2.0`, now includes the GPR associations and essential reactions, which were not previously covered in the first version. Both versions simplify the description of lipid metabolism processes by mainly discarding reactions specific to fatty acids. Therefore, essential pathways in our

reconstruction scope were excluded from `ZebraGEM`, which makes it unfit to our case study.

Initially, the global reconstruction of the human metabolic network, `Recon 1`, was published in 2007 (Duarte et al., 2007). Several network and supporting information updates have followed. A recent `Recon 3D` version includes a thorough catalogue of GPR associations in addition to structural information on metabolites and enzymes (Brunk et al., 2018). This version also comprises a three-dimensional metabolite and protein structure data which allow integrated analyses of metabolic functions in humans. The Edinburgh Human Metabolic Network (EHMN) was presented concurrently to `Recon 1` (Ma et al., 2007). The Human Metabolic Reaction (HMR) database (Agren et al., 2012) was later generated by integrating `Recon 1` and EHMN with biochemical reactions from KEGG (Aoki and Kanehisa, 2005) and `MetaCyc` (Caspi et al., 2013) databases. An updated version (`HMR 2.0`) with emphasis on lipid metabolism was later introduced (Mardinoglu et al., 2014). Several subsequent tissue- and context-specific GEMs were released. They include `iAdipocyte1809`, `iMyocytes2419`, and `CardioNet` (Mardinoglu et al., 2013; Geng and Nielsen, 2017). We emphasize the consensus GEM for hepatocytes, `iHepatocytes2322`, which was assembled from proteomics data and previously published liver models. It is a consensus hepatocyte GEM that is mainly generated from proteomics data of several existing liver models, namely `HepatoNet1` (Gille et al., 2010), `iLJ1046` (Jerby et al., 2010), `iAB676` (Bordbar et al., 2011), and `iHepatocyte1154` (Agren et al., 2012). It is well-annotated as it includes all the protein-coding genes and reactions in those liver models. Notably, and in alignment with our GEM context, `iHepatocytes2322` extensively covers lipid metabolism functions, such as the uptake of remnants of lipoproteins, the formation and degradation of lipid droplets, and secretion of synthesized lipoproteins, as well as xenobiotic biotransformation. It consists of 2,322 genes, 5,686 metabolites, and 7,930 reactions, separated into eight different compartments based on the HMR 2.0 database. Although the template is not fish-specific, the important subcellular functions and mechanisms within the compartments are conserved through evolution, particularly among fish and humans (Schlegel and Stainier, 2007; Li et al., 2010). Indeed, Zebrafish is widely used model to study human metabolic diseases (Schlegel and Stainier, 2007; Seth et al., 2013; Gut et al., 2017). However, there are differences in fat storage sites between Atlantic cod and humans, the molecular bases of which are not known. In Atlantic cod and other lean fish fat is stored mainly in the liver, whereas in humans it is stored in adipose tissues. In fact, cod liver normally contains 50–60% of fat (Li et al., 1986) that is stored in peroxisomes and cytosolic lipid droplets, and these subcellular compartments are also included in `iHepatocytes2322`.

Generating an Atlantic Cod Liver Draft Model

As described, we used the RAVEN 2.0 toolbox to generate a draft metabolic reconstruction of the Atlantic cod liver, based on protein homology, and using `iHepatocytes2322` as our input template.



In addition to the template model as input, RAVEN 2.0 requires two FASTA files including the peptide sequences of the target and the template organisms. The literature and public repositories contain four main versions of Atlantic cod genome annotations. The first assembly, gadMor1, was published in 2011 (Star et al., 2011). An updated version, gadMor2, with extensive genome sequencing enabling novel genomic and transcriptomics assemblies followed in 2017 (Torresen et al., 2017). Those two versions along with multiple RNA-seq data from multiple tissues and developmental stages were used to create gadMorTrinity (Zhang et al., 2020) which includes less-fragmented and more comprehensive sequences. gadMor3.0 (GenBank assembly accession: GCA_902167405), here gadMor3, is a recently published chromosome-level reference assembly generated using long-read sequencing technology. The length distribution of the peptide sequences from these three assemblies are shown in **Figure 1**. We used gadMor1, gadMorTrinity, and gadMor3 to examine the potential improvement that these annotations could bring in the context of metabolic reconstructions. A separate set of draft models was generated from each assembly.

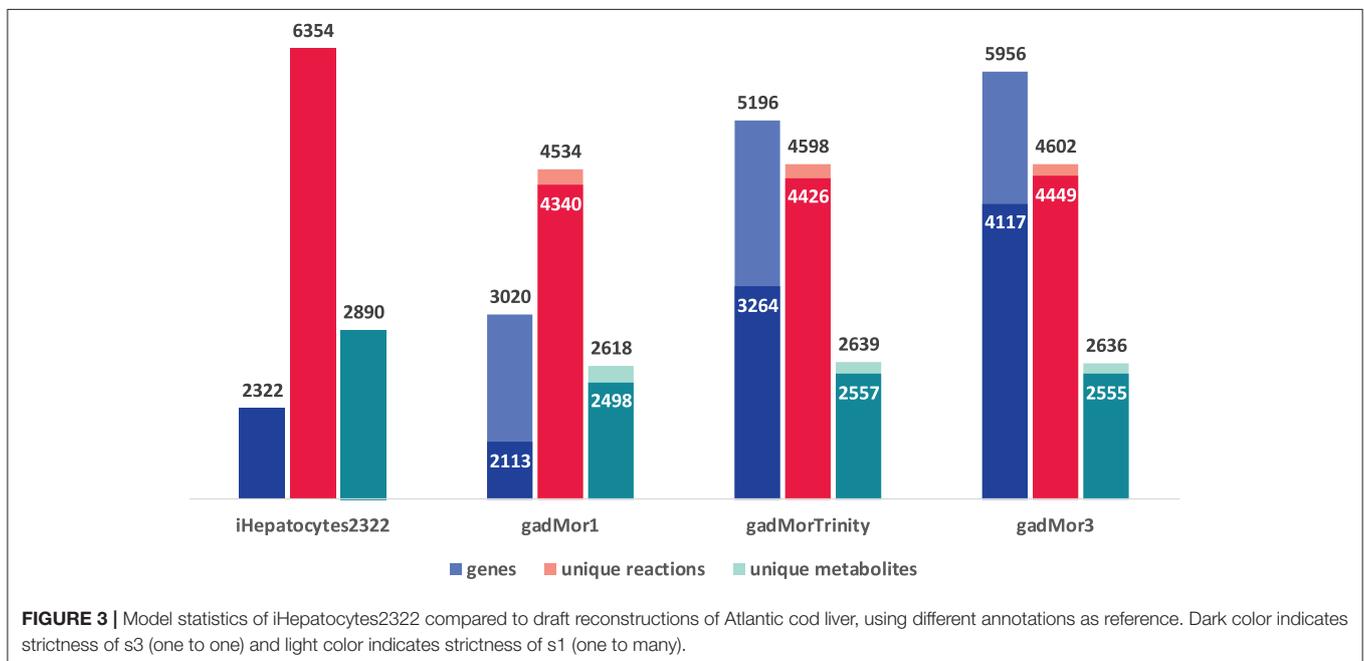
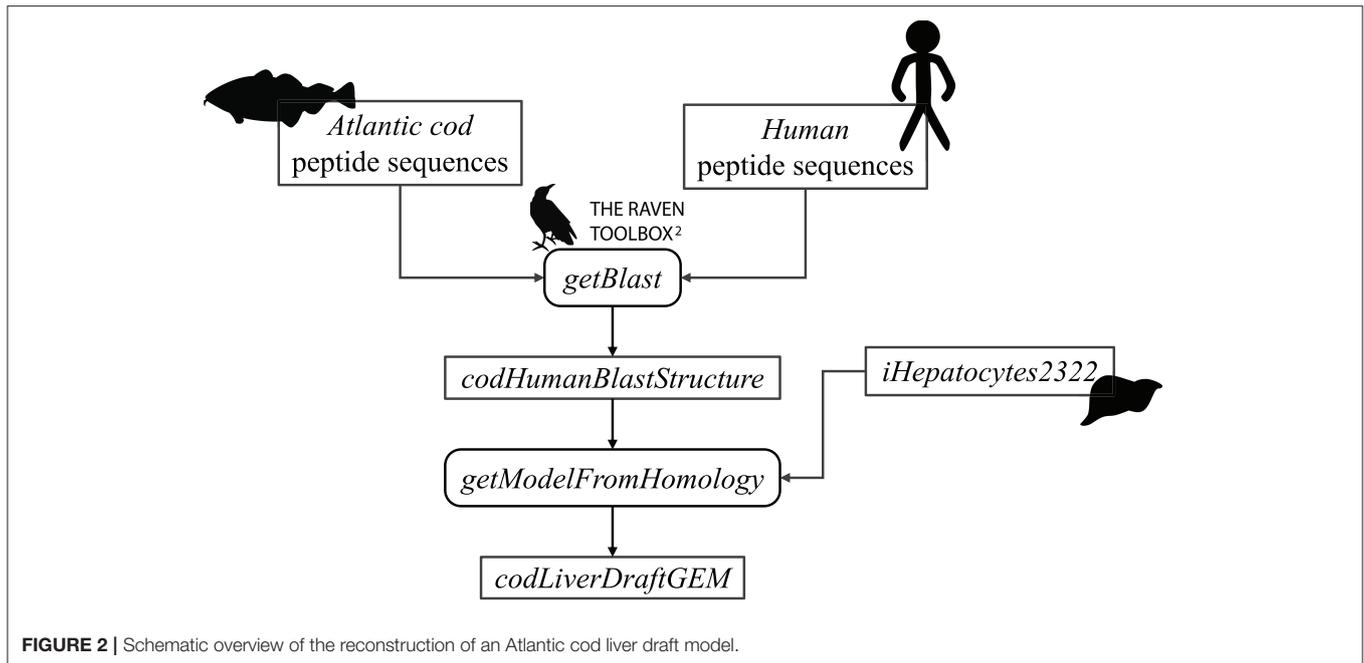
First, the FASTA files were used as input to the getBlast function. A BLAST structure, was created from the reciprocal BLAST analyses between the target and the template organism, using the BLASTp algorithm (Altschul et al., 1990). Next, this structure along with the template model were input to the getModelFromHomology function to generate a draft cod GEM by extracting reactions involving cod genes that are orthologous to human genes in iHepatocytes2322. **Figure 2** shows an overview of the draft model reconstruction process. The process can be tuned based on different input parameters.

We used the RAVEN 2.0 toolbox to generate several draft models based on different strictness parameters and Atlantic cod genome annotations. The strictness parameter specifies which reactions are included in the draft model, mainly based on how

genes from the template GEM organism and from the target organism are mapped against each other. When strictness is equal to 1 or 2, gene mapping is done for all pairs based on BLASTp score threshold, in both or in one direction, respectively. Alternatively, when strictness is set to 3, all BLASTp results are checked and only the ones with the lowest *E*-value are kept for all gene pairs in each direction separately. Then, target genes are mapped to template genes for all pairs which have acceptable BLASTp results, in both directions. In addition, maxE, minLen, and minIde are adjustable thresholds for the maximum *E*-value, the minimum alignment length, and the minimum identity acceptable in the gene mapping process. Their respective default values are 10^{-30} , 200, and 40%.

3. RESULTS

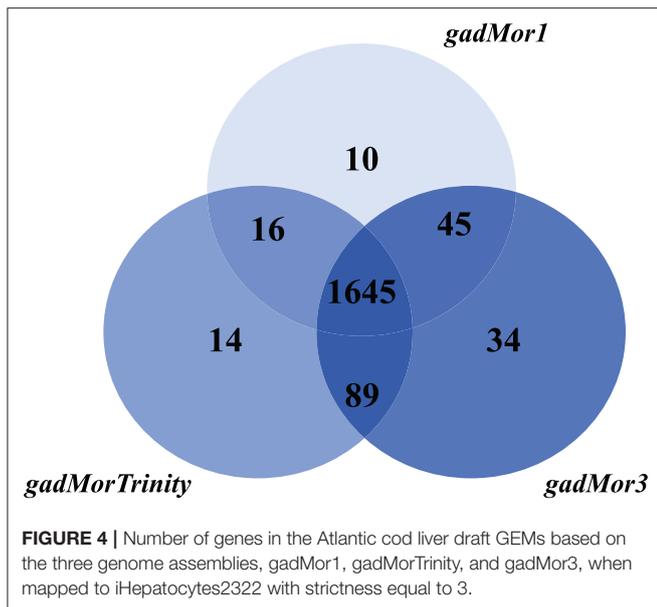
The graph in **Figure 3** shows the model statistics corresponding to iHepatocytes2322 and different cod annotations, based on strictness values of 1 (s1) and 3 (s3) for one-to-many and best one-to-one mapping between human and cod genes, respectively. First, we note that s1 leads to a greater number of matched genes in the draft models. Such high number of matches can be partly attributed to the fish specific whole-genome duplication during teleost fish evolution, where fish genomes may have more than one orthologs for some human genes (Li et al., 2010; Glasauer and Neuhaus, 2014; Grimholt, 2018; Ravi and Venkatesh, 2018). However, a higher proportion of these matches are likely to be a result of splice variants and transcript assembly artifacts. Accordingly, the more annotated the cod genome, the more duplicate matches with human genes in the template are expected. In this context, the trend in the number of genes in the draft models can be attributed to the fact that gadMor1 has the lowest while gadMor3 has the highest number of annotated



sequences. The number of reactions and metabolites across all resultant drafts is close for both strictness values.

The current version of the RAVEN toolbox adds a reaction to the draft if at least one of its catalyzing enzymes is mapped between species. Hence, the number of reactions and metabolites as produced by the printModelStats function in the toolbox does not reflect the actual gene mapping between iHepatocytes2322 and each of the draft models. One way to properly examine different cod annotation outcomes, given the known fish genome duplication, is to compare

the number of mapped genes from iHepatocytes2322 with each draft, with strictness s3. But the current RAVEN only generates the mapped genes included in the draft model, we thus added this feature to the getModelFromHomology function and pulled the request to RAVEN's GitHub page (<https://github.com/SysBioChalmers/RAVEN/pull/301>). The Venn diagram in **Figure 4** shows the output of this function. As the draft model from gadMor3 is the most complete, we used it as the baseline and included the other two as complementary sources. All three draft models were deposited in



BioModels (Malik-Sheriff et al., 2020) and assigned the identifiers MODEL2010090001, MODEL2010090002, MODEL2010090003. They are also linked to the FAIRDOMHub (Wolstencroft et al., 2017) assay “Draft metabolic reconstruction model of Atlantic cod liver” (<https://doi.org/10.15490/fairdomhub.1.assay.1365.2>).

A draft reconstruction typically contains missing genes and reactions which are gaps to be filled later. In cases where a template model is used as input, such gaps result from unmapped genes, and subsequent reactions, between the template and the target organisms. In order to examine the degree of gaps in our draft model, we considered the “Metabolism of xenobiotics by cytochrome P450” subsystem which is highly relevant to our case study. A map showing the gap filling of this subsystem was visualized by Escher (King et al., 2015) and can be found on FAIRDOMHub at <https://doi.org/10.15490/fairdomhub.1.datafile.3892.1>. The pathways of benzo[a]pyrene (BaP) and 1,2-Dibromoethane from this subsystem are shown in **Figure 5**.

By manual curation, we set several criteria for gap filling: 1. by searching relevant literature, several reactions were identified as spontaneous reactions, i.e., not catalyzed by any enzyme, and included; 2. when the chemistry of the reaction made it clear that it belonged to a process catalyzed by a certain enzyme family (i.e., cytochrome P450 monooxygenases), they were assigned to this family even though the specific gene homolog was not identified.

4. CASE STUDY OF ATLANTIC COD EXPOSED TO BENZO(A)PYRENE

Mapping gene expression data to the metabolic pathway can give a better understanding of the mechanism by which a toxicant affects the organism. Here, we show a case study of *ex vivo* exposure of Atlantic cod liver slices to benzo(a)pyrene (BaP) (Yadatie et al., 2018). As seen in **Figure 5A**, all enzymes

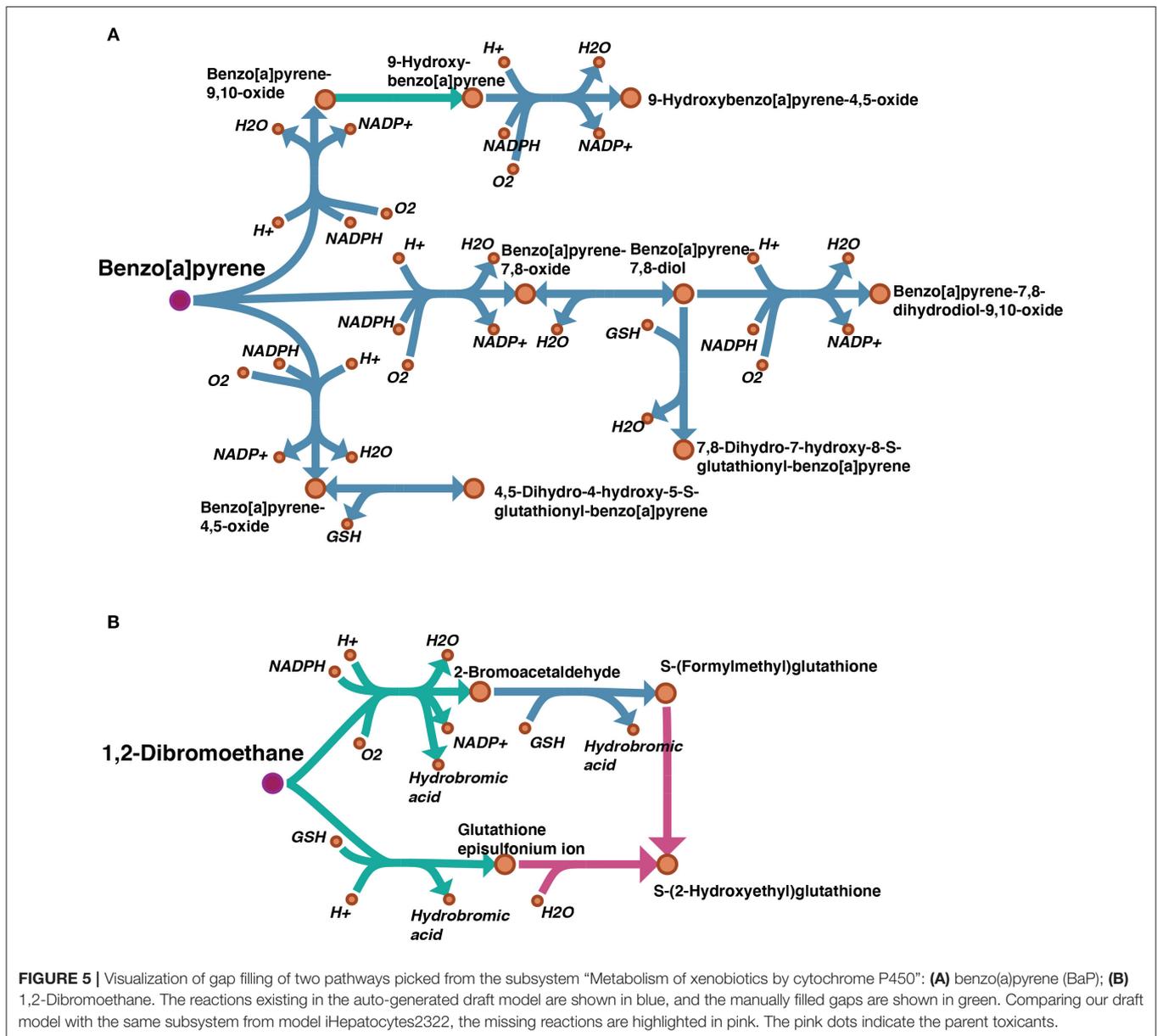
related to this metabolic pathway were mapped in our model following manual curation. In more details, the metabolic pathway of BaP (adapted from Schlenk et al., 2008) shows that there are two main biotransformation routes of BaP (**Figure 6**).

Combining the metabolic pathway and gene expression data, we show that the main biotransformation route of BaP in Atlantic cod liver is likely toward the most reactive metabolite, benzo(a)pyrene-7,8-dihydrodiol-9,10-epoxide (**Figure 6**). Simultaneously, an alternate clearance pathway dependent upon the SULT and UGT enzymes is actually down-regulated, demonstrating clear preference for the primary activation pathway. This shift can be explained by the fact that BaP itself is a well-known ligand for the aryl hydrocarbon receptor (AHR), an important ligand-activated transcription factor regulating the expression of several enzymes in xenobiotic biotransformation pathways, including CYP1A, and recently characterized in cod (Aranguren-Abadia et al., 2019). Although this mechanism is not specific for cod and well described previously (e.g., Kleinow et al., 1998; Schlenk et al., 2008), this demonstrates how a useful insight into evaluation of harmful compounds can be gained by metabolic reconstructions. The expression profiles of the genes can be found in the **Supplementary Table S2**.

5. DISCUSSION AND FUTURE PERSPECTIVES

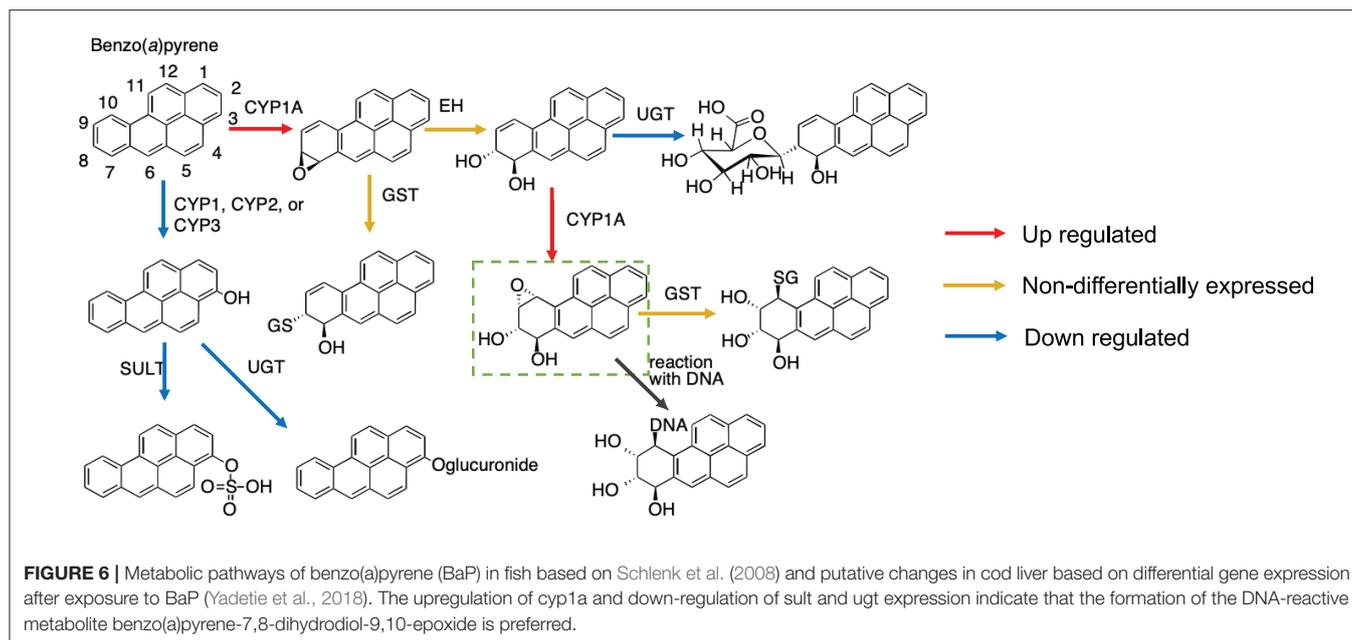
In this paper, we presented our experience in reconstructing a draft metabolic network for Atlantic cod liver. As our genome-scale reconstruction work is still on-going, the main intention in this paper is to highlight possible challenges of working on non-model less-annotated organisms. Draft reconstruction for model species is typically the fastest phase in the process, due to the abundance of biochemical knowledge and annotations. However, it becomes less straightforward with a lack of such data. Thanks to enabling computational tools, an alternative is to use existing generic or tissue-specific GEMs that fall within the context of the target reconstruction. In our case, a human liver model was used as a template to generate a draft cod liver model. The fact that it is a well-annotated consensus liver GEM compensated for known physiological and metabolic differences between cod and human, which could be resolved in the next model curation phases.

As demonstrated for the “Metabolism of xenobiotics by cytochrome P450” subsystem, gap filling is an essential part of model curation that need to be applied to all the included subsystems. We survey main gap filling approaches and then describe subsequent reconstruction steps. Several tools were presented to help automate this step. Among those approaches is Meneco (Prigent et al., 2017), a topological gap filling approach, written in Python, that identifies missing reactions through reformulating the gap filling problem as a combinatorial optimization problem. fastGapFill (Thiele et al., 2014) is available in the COBRA toolbox. It selects the set of reactions that optimize a linear score representing the quantitative metabolic production of the metabolic network. Other existing tools can identify missing reactions by integrating taxonomic (Benedict et al., 2014), compartment modularity (Mintz-Oron et al., 2012), or



phenotypic knowledge such as experimental flux data (Herrgard and Palsson, 2006; Christian et al., 2009; Kumar and Maranas, 2009; Vitkin and Shlomi, 2012). However, approaches relying on complementary knowledge to perform the gap filling process are not applicable to less-annotated organisms. Accordingly, fastGapFill from the COBRA toolbox could be a suitable choice to automatically fill the gaps in the draft model. Following this step, the refined set of reactions needs to be manually curated since some of them might be added arbitrarily to fulfill selected model criteria, such as model connectivity, by the gap filling tool. The model is then ready for evaluation using flux balance analysis (FBA) (Orth et al., 2010) and phenotype datasets. FBA aims to simulate the cell behavior through finding a set of optimal metabolic fluxes which optimize a metabolic objective,

which is often biomass production in bacteria or cancer cells. The metabolic objective of particular cell types in multi-cellular eukaryotes is more difficult to determine however, but methods to infer such objectives have been developed (Bordbar et al., 2020). The simulation results from FBA are compared with experimental values of biomass growth and metabolic fluxes and the model is subjected to refinement until it behaves as desired. The latter allows the reconstructed metabolic network to be analyzed using constraint-based modeling approach. Once those steps are completed, the model can be used to simulate different biological conditions and study the metabolic potential of the organism. For example, it is possible to perform gene knockout studies *in silico* to identify essential genes for various biological conditions using the constraint-based model of the



reconstructed network. Thus, insights can be gained into how important physiological processes in the organism, including energy metabolism, growth, and reproductive processes, can be affected by environmental toxicants.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://figshare.com/>, doi: 10.6084/m9.figshare.c.5168303.v2.

AUTHOR CONTRIBUTIONS

EMH designed the study and wrote the manuscript draft. EMH and XZ performed the *in-silico* work. TF generated the updated Atlantic cod genome assembly. DCZ proposed the idea of the case study and helped XZ with the subsequent analysis and visualization of the results. ME, FY, and AG provided input on all aspects related to biology and environmental toxicology and performed the experimental work of the case study. SF identified and summarized potential tools for gap analysis and helped formatting the manuscript. AG and IJ supervised the work. EMH

and XZ worked on the final version of the manuscript which was reviewed and proofread by all authors.

FUNDING

This work was supported by the Research Council of Norway to the Center for Digital Life Norway (DLN) project dCod 1.0: decoding the systems toxicology of Atlantic cod [grant number 248840] and the Norwegian research school in bioinformatics, biostatistics, and systems biology (NORBIS).

ACKNOWLEDGMENTS

This manuscript has been released as a pre-print (Hanna et al., 2020). We would like to thank Prof. Bernhard Palsson's group and the RAVEN group for joining the workshops and their helpful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2020.591406/full#supplementary-material>

REFERENCES

- Agren, R., Bordel, S., Mardinoglu, A., Pornputtpong, 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 Comput. Biol.* 8:e1002518. doi: 10.1371/journal.pcbi.1002518
- Aite, M., Chevallier, M., Frioux, C., Trottier, C., Got, J., Cortés, M. P., et al. (2018). Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale metabolic models. *PLoS Comput. Biol.* 14:e1006146. doi: 10.1371/journal.pcbi.1006146
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Aoki, K. F., and Kanehisa, M. (2005). Using the KEGG database resource. *Curr. Protoc. Bioinformatics* 11, 1–12. doi: 10.1002/0471250953.bi112s11

- Aranguren-Abadía, L., Lille-Langøy, R., Madsen, A. K., Karchner, S. I., Franks, D. G., Yadetie, F., et al. (2019). Molecular and functional properties of the Atlantic cod (*Gadus morhua*) aryl hydrocarbon receptors AhR1a and AHR2A. *Environ. Sci. Technol.* 2019, 1033–1044. doi: 10.1021/acs.est.9b05312
- Bekaert, M. (2012). Reconstruction of Danio rerio metabolic model accounting for subcellular compartmentalisation. *PLoS ONE* 7:e49903. doi: 10.1371/journal.pone.0049903
- Benedict, M. N., Mundy, M. B., Henry, C. S., Chia, N., and Price, N. D. (2014). Likelihood-based gene annotations for gap filling and quality assessment in genome-scale metabolic models. *PLoS Comput. Biol.* 10:e1003882. doi: 10.1371/journal.pcbi.1003882
- Boele, J., Olivier, B. G., and Teusink, B. (2012). FAME, the flux analysis and modeling environment. *BMC Syst. Biol.* 6:8. doi: 10.1186/1752-0509-6-8
- Bordbar, A., Feist, A. M., Usaite-Black, R., Woodcock, J., Palsson, B. O., and Famili, I. (2011). A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. *BMC Syst. Biol.* 5:180. doi: 10.1186/1752-0509-5-180
- Bordbar, A., Lewis, N. E., Schellenberger, J., Palsson, B. Ø., and Jamshidi, N. (2020). Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. *Mol. Syst. Biol.* 6:422. doi: 10.1038/msb.2010.68
- Brunk, E., Sahoo, S., Zielinski, D. C., Altunkaya, A., Dräger, A., Mih, N., et al. (2018). Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* 36:272. doi: 10.1038/nbt.4072
- Caspi, R., Foerster, H., Fulcher, C. A., Kaipa, P., Krummenacker, M., Latendresse, M., et al. (2013). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* 42, D459–D471. doi: 10.1093/nar/gkt1103
- Christian, N., May, P., Kempa, S., Handorf, T., and Ebenhö, O. (2009). An integrative approach towards completing genome-scale metabolic networks. *Mol. Biosyst.* 5, 1889–1903. doi: 10.1039/b915913b
- Dale, K., Müller, M. B., Tairova, Z., Khan, E. A., Hatlen, K., Grung, M., et al. (2019). Contaminant accumulation and biological responses in Atlantic cod (*Gadus morhua*) caged at a capped waste disposal site in Kollevåg, Western Norway. *Mar. Environ. Res.* 145, 39–51. doi: 10.1016/j.marenvres.2019.02.003
- Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., et al. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1777–1782. doi: 10.1073/pnas.0610772104
- Eide, M., Rydbeck, H., Tørresen, O. K., Lille-Langøy, R., Puntervoll, P., Goldstone, J. V., et al. (2018). Independent losses of a xenobiotic receptor across teleost evolution. *Sci. Rep.* 8:10404. doi: 10.1038/s41598-018-28498-4
- Geng, J., and Nielsen, J. (2017). *In silico* analysis of human metabolism: reconstruction, contextualization and application of genome-scale models. *Curr. Opin. Syst. Biol.* 2, 29–38. doi: 10.1016/j.coisb.2017.01.001
- Gille, C., Bölling, C., Hoppe, A., Bulik, S., Hoffmann, S., Hübner, K., et al. (2010). HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol. Syst. Biol.* 6:411. doi: 10.1038/msb.2010.62
- Glasauer, S. M., and Neuhauss, S. C. (2014). Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genomics* 289, 1045–1060. doi: 10.1007/s00438-014-0889-2
- Grimholt, U. (2018). Whole genome duplications have provided teleosts with many roads to peptide loaded MHC class I molecules. *BMC Evol. Biol.* 18:25. doi: 10.1186/s12862-018-1138-9
- Gu, C., Kim, G. B., Kim, W. J., Kim, H. U., Lee, S. Y. (2019). Current status and applications of genome-scale metabolic models. *Genome Biol.* 20:121. doi: 10.1186/s13059-019-1730-3
- Gut, P., Reischauer, S., Stainier, D. Y., and Arnaout, R. (2017). Little fish, big data: zebrafish as a model for cardiovascular and metabolic disease. *Physiol. Rev.* 97, 889–938. doi: 10.1152/physrev.00038.2016
- Hanemaaijer, M., Olivier, B. G., Röling, W. F., Bruggeman, F. J., and Teusink, B. (2017). Model-based quantification of metabolic interactions from dynamic microbial-community data. *PLoS ONE* 12:e173183. doi: 10.1371/journal.pone.0173183
- Hanna, E. M., Zhang, X., Eide, M., Fallahi, S., Furmanek, T., Yadetie, F., et al. (2020). ReCodLiver0. 9: Overcoming challenges in genome-scale metabolic reconstruction of a non-model species. *bioRxiv* 2020:162792. doi: 10.1101/2020.06.23.162792
- Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Linsay, B., and Stevens, R. L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* 28:977. doi: 10.1038/nbt.1672
- Herrgård, M. J., and Palsson, B. Ø. (2006). Genome-scale models of metabolic and regulatory networks. *Netw. Models Appl.* 232.
- Jerby, L., Shlomi, T., and Ruppín, E. (2010). Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol. Syst. Biol.* 6:401. doi: 10.1038/msb.2010.56
- Karlsen, E., Schulz, C., and Almaas, E. (2018). Automated generation of genome-scale metabolic draft reconstructions based on KEGG. *BMC Bioinformatics* (2018) 19:467. doi: 10.1186/s12859-018-2472-z
- Karlsen, O. A., Puntervoll, P., and Goksøyr, A. (2012). Mass spectrometric analyses of microsomal cytochrome P450 isozymes isolated from β -naphthoflavone-treated Atlantic cod (*Gadus morhua*) liver reveal insights into the cod CYPome. *Aquat. Toxicol.* 108, 2–10. doi: 10.1016/j.aquatox.2011.08.018
- King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., and Palsson, B. O. (2015). Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Comput. Biol.* 11:4321. doi: 10.1371/journal.pcbi.1004321
- Kleinow, K. M., James, M. O., Tong, Z., and Venugopalan, C. S. (1998). Bioavailability and biotransformation of benzo(a)pyrene in an isolated perfused *in situ* catfish intestinal preparation. *Environ. Health Perspect.* 106, 155–166. doi: 10.1289/ehp.98106155
- Kliwer, S. A., Goodwin, B., and Willson, T. M. (2002). The nuclear pregnane X receptor: a key regulator of xenobiotic metabolism. *Endocr. Rev.* 23, 687–702. doi: 10.1210/er.2001-0038
- Kumar, V. S., and Maranas, C. D. (2009). GrowMatch: an automated method for reconciling *in silico/in vivo* growth predictions. *PLoS Comput. Biol.* 5:e1000308. doi: 10.1371/journal.pcbi.1000308
- Li, S., Pozhitkov, A., Ryan, R. A., Manning, C. S., Brown-Peterson, N., et al. (2010). Constructing a fish metabolic network model. *Genome Biol.* 11:R115. doi: 10.1186/gb-2010-11-11-r115
- Lie, Ø., Lied, E., and Lambertsen, G. (1986). Liver retention of fat and of fatty acids in cod (*Gadus morhua*) fed different oils. *Aquaculture* 59, 187–196. doi: 10.1016/0044-8486(86)90003-7
- Ma, H., Sorokin, A., Mazein, A., Selkov, A., Selkov, E., Demin, O., et al. (2007). The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol. Syst. Biol.* 3:135. doi: 10.1038/msb4100177
- Machado, D., Andrejev, S., Tramontano, M., and Patil, K. R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* 46, 7542–7553. doi: 10.1093/nar/gky537
- Malik-Sheriff, R. S., Glont, M., Nguyen, T. V., Tiwari, K., Roberts, M. G., Xavier, A., et al. (2020). BioModels-15 years of sharing computational models in life science. *Nucl. Acids Res.* 48, D407–D415. doi: 10.1093/nar/gkz1055
- Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Nookaew, I., Jacobson, P., et al. (2013). Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Mol. Syst. Biol.* 9:649. doi: 10.1038/msb.2013.5
- Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Uhlen, M., and Nielsen, J. (2014). Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat. Commun.* 5:3083. doi: 10.1038/ncomms4083
- Mendoza, S. N., Olivier, B. G., Molenaar, D., and Teusink, B. (2019). A systematic assessment of current genome-scale metabolic reconstruction tools. *BioRxiv* 2019:558411. doi: 10.1101/558411
- Mintz-Oron, S., Meir, S., Malitsky, S., Ruppín, E., Aharoni, A., and Shlomi, T. (2012). Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proc. Natl. Acad. Sci. U.S.A.* 109, 339–344. doi: 10.1073/pnas.1100358109
- Nielsen, J. (2017). Systems biology of metabolism. *Annu. Rev. Biochem.* 86, 245–275. doi: 10.1146/annurev-biochem-061516-044757
- Oberhardt, M. A., Palsson, B. Ø., and Papin, J. A. (2009). Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* 5:320. doi: 10.1038/msb.2009.77
- Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010). What is flux balance analysis? *Nat. Biotechnol.* 28:245. doi: 10.1038/nbt.1614
- Palsson, B. Ø. (2006). *Systems Biology: Properties of Reconstructed Networks*. Cambridge: Cambridge University Press.

- Prigent, S., Frioux, C., Dittami, S. M., Thiele, S., Larhlimi, A., Collet, G., et al. (2017). Meneco, a topology-based gap-filling tool applicable to degraded genome-wide metabolic networks. *PLoS Comput. Biol.* 13:e1005276. doi: 10.1371/journal.pcbi.1005276
- Ravi, V., and Venkatesh, B. (2018). The divergent genomes of teleosts. *Annu. Rev. Anim. Biosci.* 6, 47–68. doi: 10.1146/annurev-animal-030117-014821
- Ryabinin, V., Barbière, J., Haugan, P., Kullenberg, G., Smith, N., McLean, C., et al. (2019). The UN decade of ocean science for sustainable development. *Front. Mar. Sci.* 6:470. doi: 10.3389/fmars.2019.00470
- Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., Orth, J. D., Feist, A. M., et al. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* 6:1290. doi: 10.1038/nprot.2011.308
- Schlegel, A., and Stainier, D. Y. (2007). Lessons from “lower” organisms: what worms, flies, and zebrafish can teach us about human energy metabolism. *PLoS Genet.* 3:e199. doi: 10.1371/journal.pgen.0030199
- Schlenk, D., Celander, M., Gallagher, E. P., George, S., James, M., Kullman, S. W., et al. (2008). Biotransformation in fishes. *Toxicol. Fish.* 153–234. doi: 10.1201/9780203647295.ch4
- Seth, A., Stemple, D. L., and Barroso, I. (2013). The emerging use of zebrafish to model metabolic disease. *Dis. Models Mech.* 6, 1080–1088. doi: 10.1242/dmm.011346
- Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T. F., et al. (2011). The genome sequence of Atlantic cod reveals a unique immune system. *Nature* 477:207. doi: 10.1038/nature10342
- Thiele, I., and Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5:93. doi: 10.1038/nprot.2009.203
- Thiele, I., Vlassis, N., and Fleming, R. M. (2014). fastGapFill: efficient gap filling in metabolic networks. *Bioinformatics* 30, 2529–2531. doi: 10.1093/bioinformatics/btu321
- Tørresen, O. K., Star, B., Jentoft, S., Reinart, W. B., Grove, H., Miller, J. R., et al. (2017). An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. *BMC Genomics* 18:95. doi: 10.1186/s12864-016-3448-x
- van Steijn, L., Verbeek, F. J., Spaink, H. P., and Merks, R. M. (2019). Predicting metabolism from gene expression in an improved whole-genome metabolic network model of *Danio rerio*. *Zebrafish* 16, 348–362. doi: 10.1089/zeb.2018.1712
- Visbeck, M. (2018). Ocean science research is key for a sustainable future. *Nat. Commun.* 9:690. doi: 10.1038/s41467-018-03158-3
- Vitkin, E., and Shlomi, T. (2012). MIRAGE: a functional genomics-based approach for metabolic network model reconstruction and its application to cyanobacteria networks. *Genome Biol.* 13:R111. doi: 10.1186/gb-2012-13-11-r111
- Wada, T., Gao, J., and Xie, W. (2009). PXR and CAR in energy metabolism. *Trends Endocrinol. Metab.* 6, 273–279. doi: 10.1016/j.tem.2009.03.003
- Wang, H., Marcišauskas, S., Sánchez, B. J., Domenzain, I., Hermansson, D., Agren, R., et al. (2018). RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLoS Comput. Biol.* 14:e1006541. doi: 10.1371/journal.pcbi.1006541
- Wolstencroft, K., Krebs, O., Snoep, J. L., Stanford, N. J., Bacall, F., Golebiewski, M., et al. (2017). FAIRDOMHub: a repository and collaboration environment for sharing systems biology research. *Nucl. Acids Res.* 45, D404–D407. doi: 10.1093/nar/gkw1032
- Yadatie, F., Oveland, E., Døskeland, A., Berven, F., Goksøyr, A., and Karlsen, O. A. (2017). Quantitative proteomics analysis reveals perturbation of lipid metabolic pathways in the liver of Atlantic cod (*Gadus morhua*) treated with PCB 153. *Aquat. Toxicol.* 185, 19–28. doi: 10.1016/j.aquatox.2017.01.014
- Yadatie, F., Zhang, X., Hanna, E. M., Aranguren-Abadía, L., Eide, M., Blaser, N., et al. (2018). RNA-Seq analysis of transcriptome responses in Atlantic cod (*Gadus morhua*) precision-cut liver slices exposed to benzo [a] pyrene and 17 α -ethynylestradiol. *Aquat. Toxicol.* 201, 174–186. doi: 10.1016/j.aquatox.2018.06.003
- Zhang, X., Eide, M., Karlsen, O. A., Hanna, E. M., Brun, M., Blaser, N., et al. (2019). “The chemical defensome of Atlantic cod (*Gadus morhua*): how does it differ from defensome networks in other teleost species?” in *20th International Symposium on Pollutant Responses In Marine Organisms, 2019* (Charleston, SC).
- Zhang, X., Furmanek, T., Jonassen, I., Goksøyr, A. (2020). Transcript assembly and peptide sequences of Atlantic cod. *Collection.* doi: 10.6084/m9.figshare.c.5168303.v2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Hanna, Zhang, Eide, Fallahi, Furmanek, Yadatie, Zielinski, Goksøyr and Jonassen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.