

## ***CiliateGEM: an open-project and a tool for predictions of ciliate metabolic variations and experimental condition design***

Alessio Mancini<sup>1,3\*</sup>, Filmon Eyassu<sup>2\*</sup>, Maxwell Conway<sup>3\*</sup>, Annalisa Occhipinti<sup>3</sup>, Pietro Lio<sup>3</sup>, Claudio Angione<sup>2</sup>, Sandra Pucciarelli<sup>1#</sup>

<sup>1</sup> School of Biosciences and Veterinary Medicine, University of Camerino, Camerino, Italy.

<sup>2</sup> Department of Computer Science and Information Systems, Teesside University, Middlesbrough, UK

<sup>3</sup> Computer Laboratory, University of Cambridge, Cambridge, UK

\* These authors contributed equally to this work

# Corresponding author

e-mail addresses:

Alessio Mancini: [alessio.mancini@unicam.it](mailto:alessio.mancini@unicam.it);

Filmon Eyassu: [filmoneyassu@gmail.com](mailto:filmoneyassu@gmail.com);

Maxwell Conway: [mjc233@cam.ac.uk](mailto:mjc233@cam.ac.uk);

Annalisa Occhipinti: [occhipintiannalisa@gmail.com](mailto:occhipintiannalisa@gmail.com);

Sandra Pucciarelli: [sandra.pucciarelli@unicam.it](mailto:sandra.pucciarelli@unicam.it),

Claudio Angione: [c.angione@tees.ac.uk](mailto:c.angione@tees.ac.uk),

Pietro Lio: [pl219@cam.ac.uk](mailto:pl219@cam.ac.uk).

### **Abstract**

The study of cell metabolism is becoming central in several fields such as biotechnology, evolution/adaptation and human disease investigations. Here we present CiliateGEM, the first metabolic network reconstruction draft of the freshwater ciliate *Tetrahymena thermophila*. We also provide the tools and resources to simulate different growth conditions and to predict metabolic variations. CiliateGEM can be extended to other ciliates in order to set up a meta-model, i.e. a metabolic network reconstruction valid for all ciliates.

Ciliates are complex unicellular eukaryotes of presumably monophyletic origin, with a phylogenetic position that is equal from plants and animals. These cells represent a new concept of unicellular system with a high degree of species, population biodiversity and cell complexity. Ciliates perform in a single cell all the functions of a pluricellular organism, including locomotion, feeding, digestion, and sexual processes.

After generating the model, we performed an in-silico simulation with the presence and absence of glucose. The lack of this nutrient caused a 32.1% reduction rate in biomass synthesis. Despite the glucose starvation, the growth did not stop due to the use of alternative carbon sources such as amino acids.

The future models obtained from CiliateGEM may represent a new approach to describe the metabolism of ciliates. This tool will be a useful resource for the ciliate research community in order to extend these species as model organisms in different research fields. An improved understanding of ciliate metabolism could be relevant to elucidate the basis of biological phenomena like genotype-phenotype relationships, population genetics, and cilia-related disease mechanisms.

### **Background**

A metabolic network represents in an organism the complete set of biochemical reactions suitable to synthesise or break-down metabolites. These reactions drive the production of biomass and energy to support all cellular processes. The reconstruction of a full metabolic network occurring within each cell has advanced from early biochemical studies to algorithmically-generated pathway diagrams starting from genomic sequencing [1]. Genome-scale models (GEMs) offer a comprehensive exploration and a rapid analysis of genomic data. GEMs have been used extensively to study metabolic engineering [2, 3], model-driven drug discoveries [4, 5], prediction of cellular phenotypes after perturbations [6-7], analysis of evolutionary processes [8-11] and models of interspecies interactions [12]. Organism-specific reconstructed metabolic networks may be further implemented to build mathematical models capable of simulating metabolic fluxes [13]. GEM pathway reconstruction has been used in *Caenorhabditis elegans* to predict genes essentiality [14] and to better understand the biology of arthropods [55], including those with a negative impact (vectors of human or animal diseases, agricultural pests). The latter approach is particularly useful to control harmful species and to develop new precautionary strategies [15]. Genome-scale metabolic modelling has also been successfully applied to study metabolic networks in microbes [16], including a Polychlorinated Biphenyl-degrading *Pseudomonas* [17, 18], thermophilic bacteria [19] and members of the human gut microbiota [20].

Among eukaryotic microbes, at least 25 models of *Saccharomyces cerevisiae* have been published since 2003 [21], helping to understand yeast metabolism. Ciliated protozoans may represent an alternative and useful eukaryotic model. Ciliates have been the main subject of projects supported by the EU Framework Programme Horizon 2020 such as the COST Actions [22]. Moreover, the National Centre for Genome Resources and the Gordon and Betty Moore Foundation's supported the Marine Microbiology Initiative (MMI). MMI created a valuable benchmark against the analysis of environmental transcriptomic data [23]. Ciliates, as yeasts, are eukaryotic single cells, but their structural and functional complexity is comparable to human and other metazoan cells. These unicellular organisms are similar to differentiated animal cells with complex functions and membrane-bound structures [24]. The potential number of extant ciliate species has been estimated at nearly 30,000 [25]. They represent an important mediator in the food chain by transforming ultrafine organic matter useful for zooplankton. Being individual cells, they are directly exposed to environmental changes, making them good models for studying cell-stress response and adaptation. Ciliates propagate mainly asexually by transverse fission, even though they perform conjugation, a sexual process that “renews” the genetic material. The complexity of ciliates is further represented by the presence of two different nuclei: the diploid micronucleus involved in conjugation and the polyploid macronucleus. While the first represents the germinal line, the second represents the somatic line. The macronucleus is responsible for gene expression during the vegetative phase.

We describe below a preliminary open software tool (*CiliateGEM*), focused on the *T. thermophila* macronuclear genomic sequences, which allows ciliates researchers to analyse a reconstructed network via Flux Balance Analysis (FBA). By studying this ciliate, we could discover new mechanisms for evolution and adaptation within metabolism, population, species and host-symbiont association. Cellular responses described in *T. thermophila* can be of fundamental importance to understand the biology of all ciliates. *CiliateGEM* is to date the most advanced tool available for ciliates and contains the highest number of curated biochemical reactions.

## Methods

### ***Tetrahymena* model (*CiliateGEM*) reconstruction**

*CiliateGEM* was created using the protocol described by Thiele and Palsson [13]. First, AUGUSTUS (a gene prediction web server [26]), was used to identify all the *Tetrahymena* protein-encoding genes from the macronuclear genome assembly and then to assign the functional roles

through accurate annotation with EC (Enzyme Commission) number. Next, a list of metabolic reactions was generated from KEGG and converted into a stoichiometric matrix. KEGG was used to search for *Tetrahymena* pathways and all the reactions were manually assembled. The specific links are listed in Table S1. A general list of exchange reactions was obtained from the literature [37]. The reaction list was reduced and adapted to fit *Tetrahymena* metabolism (Table S2). Establishing enzyme-reaction relationship is a complex step because an enzyme can be encoded by one or many genes, and each enzyme can catalyse one or more reactions [27]. We defined the growth media and external factors based on *Tetrahymena* experimental studies [28, 29, 30]. This facilitated us in capturing the organism physiological properties. The produced model was then converted into a mathematical one used to simulate the maximal growth with flux-balance analysis (FBA) [31].

FBA requires reactions to be represented as a stoichiometric matrix ( $S$ ), with dimensions of  $m \times n$ , where the metabolites ( $m$ ) are represented as rows and the reactions ( $n$ ) are represented as columns. The stoichiometric matrix is a numerical matrix of stoichiometric coefficients for each metabolite participating in a reaction. FBA assumes the system to be in a pseudo-steady state  $S \cdot v = 0$  holding for internal metabolites (reactants and products of the chemical reactions), where the vector  $v$  represents the flux distribution. An internal metabolite constitutes the model but cannot be imported or exported directly. An “exchange” counterpart of that metabolite, and an exchange reaction are required for modelling import/export reactions. Exchange metabolites can be imported and exported from the system, so they do not satisfy the steady state assumption. The flux distribution  $v$  is a vector of reaction rates and represents a feasible flux of metabolites through the reaction network, where under the principle of mass conservation, the total amount of metabolite consumed and total amount of metabolite produced are equal to zero.

In FBA, directionality and capacity constraints are placed on individual reactions by defining the upper ( $Vmax$ ) and lower ( $Vmin$ ) bounds on the range of values that the flux of each reaction can have ( $Vmin \leq v \leq Vmax$ ). These constraints define the space of allowable flux distributions at which every metabolite is consumed or produced by each reaction in the system. Such flux bounds, coupled with gene-protein-reaction association rules, can also be used to map multiple environmental conditions, using binary threshold-based or continuous valve-based approaches [32, 33]. Despite these constraints, the system is still underdetermined (more unknowns than equations) and, as a result, infinite viable solutions exist. A flux distribution is obtained by defining an objective function that is a scalar product of the vector of flux rates  $v$ , and a vector of weights  $C$ , measuring how each component in the network contributes to the production of a biologically desirable phenotype.

Formally, we adopted the following linear program:

$\max c \cdot v$ , such that

$$S \cdot v = \dot{x}$$

$$Vmin \leq v \leq Vmax$$

$$\dot{x}_i = 0 \text{ if } M_i \in \text{internal metabolites}$$

$$\dot{x}_i \in R \text{ if } M_i \in \text{exchange metabolites}$$

*CiliateGEM*, consists of 545 reactions, 828 metabolites, 64 transport step reactions (which represent the import and export of metabolites between extra- and intra-cellular space), and 84 boundary steps (which represent the input and output of metabolites from extracellular space and a biomass

reaction). The biomass reaction defines a unique objective function to effective growth. This describes the rate at which all of the biomass precursors are made in the correct proportions. Linear programming was used to calculate the optimal flux distribution that maximizes growth. The model was encoded in SBML format [34] and fulfils MIRIAM requirements [35]. It was then imported in MATLAB (version R2016b). The simulation was carried out using COBRA [36] toolbox with the linear programming solver GLPK. All biomass flux values were given in millimoles/hour/grams of dry weight (mmol/gDW/hr). CiliateGEM is provided as Supplementary Material (CiliateModel.mat).

## Results

### *CiliateGEM: an open project and a methodological pipeline*

The complexity of ciliates makes it necessary to adopt a new approach to study their metabolism. To understand these organisms is fundamental to consider that they can feed, move, and reproduce in a single cell. Ciliate complexity includes also cell compartmentalisation, which ensures the optimal environment for each specific metabolic reaction (i.e. optimal lysosomes pH for macromolecule hydrolysis).

To date, *CiliateGEM* allows the analysis of their reconstructed networks using the ciliate *Tetrahymena thermophila* as a case study. We choose this organism because *Tetrahymena* micro [38] and macro-nuclear [39] genomes have been sequenced and its biology has been extensively studied. The goal of this open project is to create a tool to simulate the growth of *Tetrahymena* cells in different conditions and then to expand it to all ciliates. The development and the refinement of this tool, coupled with the COBRA toolbox [56], can lead researchers to predict gene essentiality and genotype-phenotype relationship. *CiliateGEM* has been obtained using a bottom-up approach, from genome annotation to a mathematical model. The steps we followed are represented in Fig. 1.

### *Tetrahymena growth simulation*

One of the first experimental in vitro studies of metabolic pathways in ciliates were performed using *Tetrahymena pyriformis* as a model organism [28, 29, 30]. *Tetrahymena* cells were grown in a standard mixture of  $^{14}\text{C}$  radiolabelled substrates including glucose, glycerol, pyruvate and glutamate. The incorporation of  $^{14}\text{C}$  in  $\text{CO}_2$ , glycogen and nucleic acids was measured with ranges of values depending on the carbon labelled position in the substrate structure. Borowitz et al. [29] reported a higher incorporation of  $^{14}\text{C}$  into  $\text{CO}_2$  after one hour of incubation with [ $^{14}\text{C}$ ]glucose than in the same experiment with [ $^{14}\text{C}$ ]glycerol, i.e. 43.9-64.2 nmol/ $10^6$  cells versus 19.8-19.9 nmol/ $10^6$  cells. Stein et al. [30] further reported a higher radiolabelled C incorporation in  $\text{CO}_2$  from [ $^{14}\text{C}$ ]glucose than from pyruvate, glutamate and glycerol. The measured values of the labelled carbon incorporation into the  $\text{CO}_2$  product were as follows: [ $^{14}\text{C}$ ]glucose (106-255 nmol/ $10^6$  cells); [ $^{14}\text{C}$ ]pyruvate (55.4-121 nmol/ $10^6$  cells); [ $^{14}\text{C}$ ]glutamate (1.90-3.96 nmol/ $10^6$  cells); [ $^{14}\text{C}$ ]glycerol (11.0-13.6 nmol/ $10^6$  cells). Then, they tested glycogen as product, and the [ $^{14}\text{C}$ ]glucose resulted the preferred substrate with the highest incorporation rates (313-496 nmol/ $10^6$  cells). A comparable incorporation rate between glycerol and pyruvate was reported (~6.5 nmol/ $10^6$  cells), as well as an absent incorporation from glutamate (<0.2 nmol/ $10^6$  cells). The results differed from those of  $\text{CO}_2$  among the secondary substrates (glutamate, pyruvate and glycerol), but were in agreement on the highest used substrate (glucose).

Our initial simulation showed the maximum growth when glucose, glutamate, glycerol and pyruvate were used as growth media. In agreement with literature [29, 30], simulations of CiliateGEM showed that glucose is utilised as the preferred carbon source (Fig. 2). The uptake for glucose in our

model (-10 mmol/gDW/hr) is proportionally much higher than those of the secondary substrates (glutamate, pyruvate and glycerol) described by the authors in the previous in-vitro experiments [29, 30]. We allowed such uptake to saturate the system so the full amount is not necessarily utilised because of the thermodynamics constraints. Glucose is metabolised through glycolysis in the cytosol. A simulation in the presence of the above metabolites but without glucose showed a 32.1% reduction rate on biomass synthesis (Fig. 3). Despite the reduced biomass production, the growth without glucose is largely maintained by utilising alternative carbon sources such as amino acids (Fig. 3), as shown by the higher increase of biomass production for reactions R00891 (i.e. L-Serine + Hydrogen sulfide  $\rightleftharpoons$  L-Cysteine + H<sub>2</sub>O) and R00258 (L-Alanine + 2-Oxoglutarate  $\rightleftharpoons$  Pyruvate + L-Glutamate).

After glucose starvation, reactions R10822 [ADP + DNA(n) + 5'-Phospho-DNA(m)  $\rightleftharpoons$  AMP + Orthophosphate + DNA(n+m)] and R00381 [ATP + DNA(n) + 5'-Phospho-DNA(m)  $\rightleftharpoons$  AMP + Diphosphate + DNA(n+m)] showed the highest positive and negative values. Both reactions are catalysed by DNA ligases during DNA repair or recombination. These two reactions differ only on ATP or ADP cofactors. ATP is the phosphorylated form of ADP and is the preferred cofactor because phosphoanhydride bonds store a high amount of energy. Since ATP is synthesised from ADP during glycolysis by phosphorylation, we can speculate that in absence of glucose DNA ligation by ligase is performed by using ADP molecules as an alternative. DNA ligase using ADP has been reported from the aerobic hyper-thermophilic archaeon *Aeropyrumpernix* K1 [40]. Other opposite values were reported for reactions R02549 (4-Aminobutyraldehyde + NAD<sup>+</sup> + H<sub>2</sub>O  $\rightleftharpoons$  4-Aminobutanoate + NADH + H<sup>+</sup>) and R01986 (4-Aminobutyraldehyde + NADP<sup>+</sup> + H<sub>2</sub>O  $\rightleftharpoons$  4-Aminobutanoate + NADPH + H<sup>+</sup>), both involving in the arginine-catabolism pathway. Like R10822 and R00381 they only differ on cofactors, in this case NAD and NADP. While the structural difference between these two molecules is only the phosphate group, NADH participates in catabolic reactions, i.e. reactions that break down molecules to release energy, while NADPH participates in anabolic reactions, namely those that consume energy in order to build up or synthesise larger molecules. Finally, reactions R04241 (ATP + THF-polyglutamate(n) + L-Glutamate  $\rightleftharpoons$  ADP + Orthophosphate + THF-polyglutamate(n+1)) and R04242 (THF-polyglutamate + n H<sub>2</sub>O  $\rightleftharpoons$  Tetrahydrofolate + n L-Glutamate) showed an increase of biomass production in the simulation without glucose. Both reactions are involved in folate biosynthesis, an essential cofactor for DNA and amino acid synthesis.

## Discussion

Studies on *Tetrahymena* have led ciliate researchers to landmark discoveries on cellular mechanisms, including the first cytoskeletal motor [41] and programmed translational frameshifting [42]. Some other studies have been awarded two Nobel prizes: to T. Cech and S. Altman for the discovery of catalytic RNA in 1989; and to E. H. Blackburn, C. W. Greider and J. W. Szostak for their work on telomeres in 2009. In general, mechanisms first described in ciliates have proven to be of widespread occurrence and of fundamental importance for the biology of all eukaryotes. Furthermore, ciliates are the only unicellular organisms in which programmed DNA elimination during somatic differentiation, widespread in animals [43], has been deeply characterised [44].

In this study, we set the basis of the open project *CiliateGEM* and we constructed a metabolic model draft of *T. thermophila*, made freely available to researchers. We checked the model robustness by simulating the organism cell growth in the presence or absence of glucose (the preferred carbon source utilised by this organism). Then we compared glucose to glutamate, pyruvate and glycerol consumption. Despite the organisms shows a decreased rate of biomass production, the growth without glucose is maintained by using alternative carbon sources, as amino acids and the tetrahydrofolate synthesis (essential cofactor in several metabolic reactions). This

metabolic variation appears to increase the activity of ADP dependent DNA ligase I, possibly involved in DNA repair or recombination (R10822). We also speculated that in absence of glucose, ADP may be the preferred energy molecule instead of ATP, and the metabolism is switched towards catabolic reactions in order to obtain energy for growth. Our results showed consistency with previous studies on *Tetrahymena pyriformis* [28, 29, 30], therefore suggesting that CiliateGEM could provide a framework for the entire community.

*CiliateGEM* is freely available for academics and can be fully customised and updated. Our long-term objective is to achieve a “meta-model” that can be applied for metabolic simulations using other ciliates genomic resources (Table S4). The impact of *CiliateGEM* is to allow researchers to predict the response to changes in experimental conditions (e.g. starvation), and consequently to optimise the design of experimental protocols. Although this study focuses on *T. thermophila*, the findings may be shared with other ciliates as *Paramecium* and *Euplotes* species, including Antarctic strains [45]. In the future, we aim to reconstruct a multi-compartment model in symbiosis with algae by incorporating our datasets with an algal genome-scale model [46] (53% of reactions in the *CiliateGEM* are common in algae). Moreover, the complex systems of ciliates membranous compartments associated to phagocytosis and trafficking is strictly correlated with nutrients uptake and metabolism [47, 48]. We believe that understanding the role of compartments is even more important in multi compartmentalised unicellular organisms than in multicellular organisms (a list of ciliate organelles is reported in Table S3). Reactions from different ciliates can be integrated in order to build up a hypothesis-driven scalable metabolic model based on common reactions.

In conclusion, our work confirms that *Tetrahymena* can be used as model organism also for metabolic network reconstruction. Even though *Tetrahymena* is much less known from the metabolic point of view than others model organisms such as *Saccharomyces* and *C. elegans* (including the metabolic network operating during grown at laboratory-controlled conditions), the availability of “omic” resources allows reliable modelling of metabolic fluxes. Tools and resources obtained for *Tetrahymena* can then be applied to other ciliates and can elucidate genotype-phenotype mapping. A key point of this project is to create a collaborative network useful to expand and validate CiliateGEM with in vitro experiments on metabolic reactions already collected inside the model, and ultimately add new reactions. Collaboration between research teams in this type of projects is essential to speed up the process and to join the expertise of different research groups. A consensus modelling approach, already successfully adapted for many metabolic reconstructions including human and yeast, might help researchers to understand the ciliate differences at genera, species, subspecies and populations level. Furthermore, with a combination of shared resources, it will be possible to study full ecosystems (food chain from endosymbionts, associated bacteria, predators-preys). These models, coupled with multi-omic approaches [49], could also highlight aspects of evolutionary biology and biogeography, sympatric and allopatric speciation.

## Acknowledgements

This research work was supported by the European Commission Marie Skłodowska-Curie Actions H2020 RISE Metable - 645693

## References

(1) Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol.* 2010; 28: 977–982.

- (2) McAnulty MJ, Yen JY, Freedman BG, Senger RS. Genome-scale modeling using flux ratio constraints to enable metabolic engineering of clostridial metabolism in silico. *BMC Syst Biol.* 2012. 6:42.
- (3) Kim B, Kim WJ, Kim DI, Lee SY. Applications of genome-scale metabolic network model in metabolic engineering. *J Ind Microbiol Biotechnol.* 2015; 42: 339–348.
- (4) Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol.* 2011. 7:501.
- (5) Shaked I, Oberhardt MA, Atias N, Sharan R, Ruppin E. Metabolic Network Prediction of Drug Side Effects. *Cell Syst.* 2016. 2(3):209-13.
- (6) Eyassu F, Angione C. Modelling pyruvate dehydrogenase under hypoxia and its role in cancer metabolism. *Royal Society open science.* 2017 Oct 1;4(10):170360.
- (7) Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM, Palsson BØ. Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci U S A.* 2013; 110(50):20338-43.
- (8) Angione C, Lió P, Predictive analytics of environmental adaptability in multi-omic network models. *Scientific reports,* 2015; 5.
- (9) Babaei P, Ghasemi-Kahrizsangi T, Marashi SA. Modeling the differences in biochemical capabilities of pseudomonas species by flux balance analysis: How good are genome-scale metabolic networks at predicting the differences? *The Scientific World Journal.* 2014; vol 2014
- (10) van Heck RG, Ganter M, Martins Dos Santos VA, Stelling J. Efficient Reconstruction of Predictive Consensus Metabolic Network Models. *PLoS Comput Biol.* 2016;12(8):e1005085.
- (11) Koehorst JJ, van Dam JC, van Heck RG, Saccenti E, Dos Santos VA, Suarez-Diez M, Schaap PJ. Comparison of 432 Pseudomonas strains through integration of genomic, functional, metabolic and expression data. *Sci Rep.* 2016;6:38699.
- (12) Levy R, Borenstein E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proceedings of the National Academy of Sciences of the United States of America.* 2013; 110(31): 12804–12809.
- (13) Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols.* 2010; 5(1): 93–121.
- (14) Yilmaz LS, Walhout AJ. A *Caenorhabditis elegans* Genome-Scale Metabolic Network Model. *Cell Syst.* 2016; 2(5):297-311.
- (15) Baa-Puyoulet P, Parisot N, Febvay G, et al. ArthropodaCyc: a CycADS powered collection of BioCyc databases to analyse and compare metabolism of arthropods. *Database: The Journal of Biological Databases and Curation.* 2016;2016:baw081.
- (16) Vijayakumar S, Conway M, Lió P, Angione C. Seeing the wood for the trees: a forest of methods for optimization and omic-network integration in metabolic modelling. *Briefings in Bioinformatics.* 2017. May 30.

- (17) Taffi, M, Paoletti, N, Angione, C, Pucciarelli, S, Marini, M, Liò, P: Bioremediation in marine ecosystems: a computational study combining ecological modeling and flux balance analysis. *Front Genet.* 2014; 5: 319-325
- (18) Taffi, M, Paoletti, N, Liò, P, Pucciarelli, S, Marini, M: Bioaccumulation modelling and sensitivity analysis for discovering key players in contaminated food webs: the case study of PCBs in the adriatic sea. *Ecol. Model.* 2015; 306: 205–215
- (19) Dahal S, Poudel S, Thompson RA. Genome-Scale Modeling of Thermophilic Microorganisms. *Adv Biochem Eng Biotechnol.* 2016; 160:103-119
- (20) Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, Greenhalgh K, Jäger C, Baginska J, Wilmes P, Fleming RM, Thiele I. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat Biotechnol.* 2017; 35(1):81-89.
- (21) Heavner BD, Price ND. Comparative Analysis of Yeast Metabolic Network Models Highlights Progress, Opportunities for Metabolic Reconstruction. *PLoS Comput Biol.* 2015; 13;11(11):e1004530.
- (22) Cost action: (<http://www.cost.eu/COSTActions/bmbs/BM1102>)
- (23) Marine Microbiology Initiative (MMI): <http://marinemicroeukaryotes.org/>
- (24) Bright LJ, Kambesis N, Nelson SB, Jeong B, Turkewitz AP. Comprehensive Analysis Reveals Dynamic and Evolutionary Plasticity of Rab GTPases and Membrane Traffic in *Tetrahymena thermophila*. *PLoS Genetics.* 2010;6(10):e1001155.
- (25) Warren A, Patterson DJ, Dunthorn M, et al. Beyond the “Code”: A Guide to the Description and Documentation of Biodiversity in Ciliated Protists (Alveolata, Ciliophora). *J. Eukaryot. Microbiol.* 2017; 64: 539–554.
- (26) Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, 2005; 33:465-467
- (27) Poolman MG, Bonde BK, Gevorgyan A, Patel HH, Fell DA. Challenges to be faced in the reconstruction of metabolic networks from public databases. *Systems biology* 2006; 153(5):379–384.
- (28) Raugi GJ, Liang T, Blum JJ. A quantitative analysis of metabolite fluxes along some of the pathways of intermediary metabolism in *Tetrahymena pyriformis*. *Journal of Biological Chemistry.* 1975; 250(15):5866–5876.
- (29) Borowitz MJ, Stein RB, Blum JJ. Quantitative analysis of the change of metabolite fluxes along the pentose phosphate and glycolytic pathways in *Tetrahymena* in response to carbohydrates. *J Biol Chem.* 1977 Mar 10;252(5):1589-605.
- (30) Steint B, Blum JJ. Quantitative Analysis of Intermediary Metabolism in *Tetrahymena*. *Journal of Biological Chemistry* 1980. 255(9):4198–4205.
- (31) Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nature biotechnology*, 2010. 28(3): 245–248.

- (32) Angione C. Integrating splice-isoform expression into genome-scale models characterizes breast cancer metabolism. *Bioinformatics*. 2017 Sep 8.
- (33) Opdam S, Richelle A, Kellman B, Li S, Zielinski DC, Lewis NE. A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. *Cell Syst*. 2017; 22;4(3):318-329.
- (34) Hucka M, Bergmann FT, Hoops S, Keating SM, Sahle S, Schaff JC, Smith LP, Wilkinson DJ. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core. *J Integr Bioinform*. 2015; 12(2):266.
- (35) Laibe C, Le Novère N. MIRIAM Resources: tools to generate and resolve robust cross-references in Systems Biology. *BMC systems biology*, 2007; 1:58.
- (36) Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BØ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc*. 2011; 6(9):1290-307.
- (37) Nogales J, Palsson BO, Thiele Ines. A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. *BMC Systems Biology*. 2008; 2:1-79.
- (38) Hamilton EP, Kapusta A, Huvos PE, et al. Structure of the germline genome of *Tetrahymena thermophila* and relationship to the massively rearranged somatic genome. *eLife*. 2016;5:e19090.
- (39) Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Delcher AL. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol*. 2006; 4(9):e286.
- (40) Jeon SJ, Ishikawa K. A novel ADP-dependent DNA ligase from *Aeropyrum pernix* K1. *FEBS Lett*. 2003; 550(1-3):69-73.
- (41) Gibbons IR, Rowe AJ 1965. Dynein: A protein with adenosine triphosphatase activity from cilia. *Science* 149:424–426
- (42) Lobanov AV, Heaphy SM, Turanov AA, Gerashchenko MV, Pucciarelli S, Devaraj RR, Xie F, Petyuk VA, Smith RD, Klobutcher LA, Atkins JF, Miceli C, Hatfield DL, Baranov PV, Gladyshev VN. Position-dependent termination and widespread obligatory frameshifting in *Euplotes* translation. *Nat Struct Mol Biol*. 2017;24(1):61-68.
- (43) Wang J, Davis RE. Programmed DNA elimination in multicellular organisms. *Curr Opin Genet Dev*. 2014;27:26-34.
- (44) Nowacki M, Landweber LF. 2009 Epigenetic inheritance in ciliates. *Current opinion in microbiology*.;12(6):638-643. doi:10.1016/j.mib.2009.09.012.
- (45) Pucciarelli S, La Terza A, Ballarini P, Barchetta S, Yu T, Marziale F, Passini V, Methe B, Detrich HW III, Miceli C (2009) Molecular cold-adaptation of protein function and gene regulation: the case for comparative genomic analyses in marine ciliated protozoa. *Mar Genomics* 2, 57–66.

- (46) Gomes de Oliveira Dal'Molin C, Quek LE, Palfreyman RW, Nielsen LK. AlgaGEM – a genome-scale metabolic reconstruction of algae based on the *Chlamydomonas reinhardtii* genome. *BMC Genomics*, 2011; 12(Suppl 4), p.55.
- (47) Nusblat AD, Bright LJ, Turkewitz AP. Conservation and innovation in *Tetrahymena* membrane traffic: proteins, lipids, and compartments. *Methods in cell biology*. 2012; 109: 141.
- (48) Guerrier S, Plattner H, Richardson E, Dacks JB, Turkewitz AP. An evolutionary balance: conservation vs innovation in ciliate membrane trafficking. *Traffic*. 2017; 18(1):18-28.
- (49) Angione C, Liò P, Pucciarelli S, Can B, Conway M, et al. 2016. *Bioinformatics Challenges and Potentialities in Studying Extreme Environments*, Vol. 9874 LNCS
- (50) Stover NA, Krieger CJ, Binkley G, Dong Q, Fisk DG, Nash R, Sethuraman A, Weng S, Cherry JM. Tetrahymena Genome Database (TGD): a new genomic resource for *Tetrahymena thermophila* research. *Nucleic Acids Res*. 2006
- (51) Arnaiz O, Cain S, Cohen J, Sperling L. ParameciumDB: a community resource that integrates the *Paramecium tetraurelia* genome sequence with genetic data. *Nucleic Acids Res*. 2007 Jan;35(Database issue):D439-44.
- (52) Arnaiz O, Sperling L. ParameciumDB in 2011: new tools and new data for functional and comparative genomics of the model ciliate *Paramecium tetraurelia*. *Nucleic Acids Res*. 2011 Jan;39(Database issue):D632-6.
- (53) Swart EC, Bracht JR, Magrini V, Minx P, Chen X, Zhou Y, Khurana JS, Goldman AD, Nowacki M, Schotanus K, Jung S, Fulton RS, Ly A, McGrath S, Haub K, Wiggins JL, Storton D, Matese JC, Parsons L, Chang WJ, Bowen MS, Stover NA, Jones TA, Eddy SR, Herrick GA, Doak TG, Wilson RK, Mardis ER, Landweber LF. The *Oxytricha trifallax* macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. *PLoS Biol*. 2013;11(1):e1001473.
- (54) Burns J, Kukushkin D, Lindblad K, Chen X, Jonoska N, Landweber LF.: a database of ciliate genome rearrangements. *Nucleic Acids Res*. 2016 Jan ;44(D1):D703-9. doi: 10.1093/nar/gkv1190
- (55) Baa-Puyoulet P, Parisot N, Febvay G, Huerta-Cepas J, Vellozo AF, Gabaldón T, Calevro F, Charles H, Colella S. ArthropodaCyc: a CycADS powered collection of BioCyc databases to analyse and compare metabolism of arthropods. *Database*. 2016 Jan 1;2016:baw081.
- (56) Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdottir HS, Keating SM, Vlasov V, Wachowiak J, Magnúsdóttir S. Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3. 0. arXiv preprint arXiv:1710.04038. 2017 Oct 11.

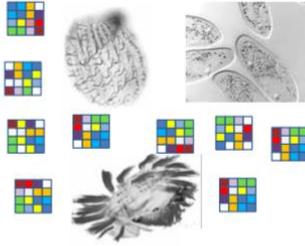
## Figure legends

**Figure 1. CiliateGEM pipeline.** *CiliateGEM* was constructed using the protocol by Thiele and Palsson (2010), as well as manual curation. To characterise the metabolic networks, we have gathered all core pathways from different organisms, including bacteria. *CiliateGEM* is provided in Matlab format as Supplementary Material (file name: *CiliateGEM.mat*).

**Figure 2. Rate of biomass synthesis by CiliateGEM from different substrates.** The *CiliateGEM* model was allowed to utilise different carbon sources for growth. Glucose, glutamate and pyruvate consumption (illustrated by negative flux) directly affects the growth rate of *CiliateGEM* (depicted by positive flux values).

**Figure 3. Differential biomass production (%) after glucose starvation.** Values for growth with (Gg) and without glucose (Gwg) were used in this formula  $(Gwg-Gg)/|Gg|*100$ .

**1** All protein encoding genes of the model ciliate (*T. thermophila* in this study) have been identified and annotated using AUGUSTUS [26]



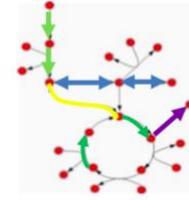
**2** Functional roles were assigned to the identified genes with Enzyme Commission (EC) number

EC 1.1.1.1
Gene - Reaction map
Pathway
EC 1.1.1.1
Gene - Reaction map
Pathway
EC 1.1.1.1
Gene - Reaction map
Pathway

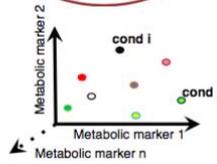
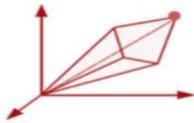
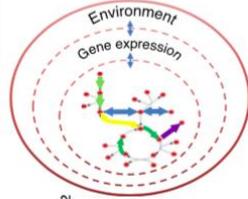
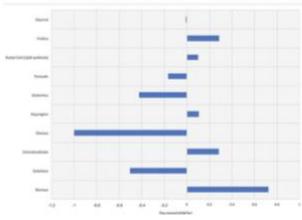
Metabolites

	Reactions			
	1	2	...	n
A	-1			-1
B	-1	-1		1
C	1	-2		
D		1		-1
...				
m				

Stoichiometric matrix **S**



**5** The resulting model was used to simulate the maximal growth using FBA and to investigate the *T. thermophila* reaction network



**4** Experimental studies were used to define the growth media and external factors

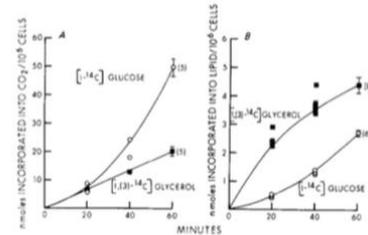


Fig. 1

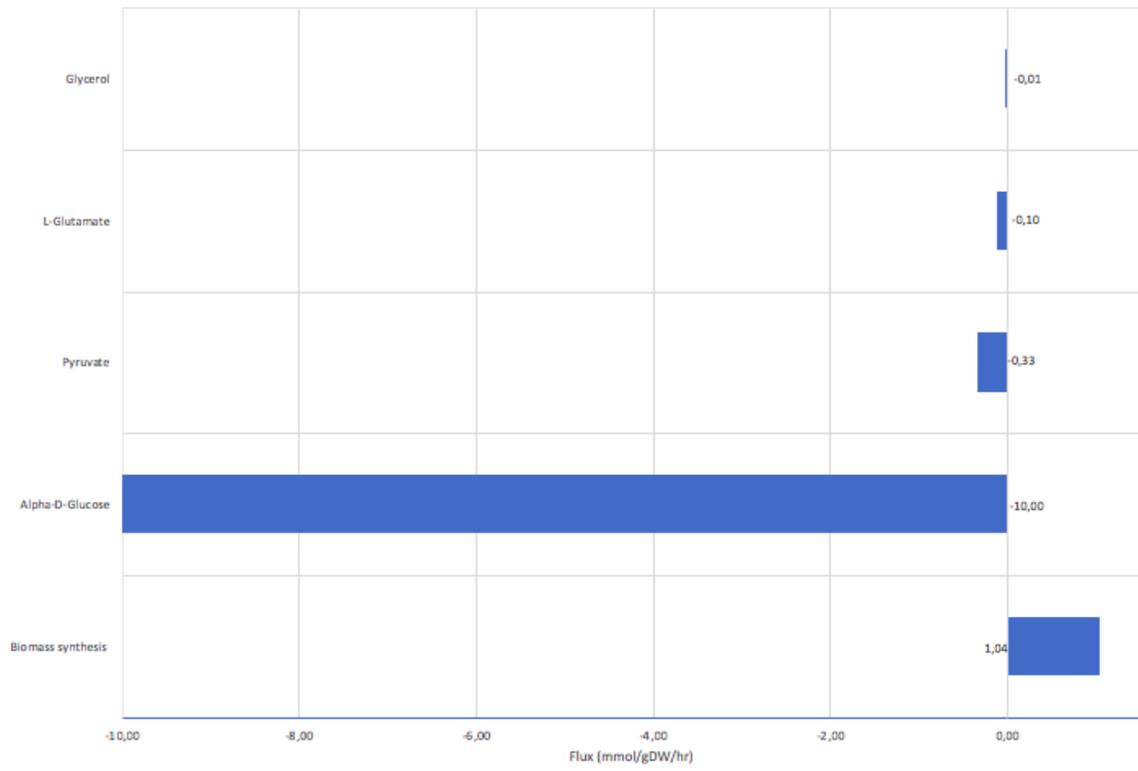


Fig. 2

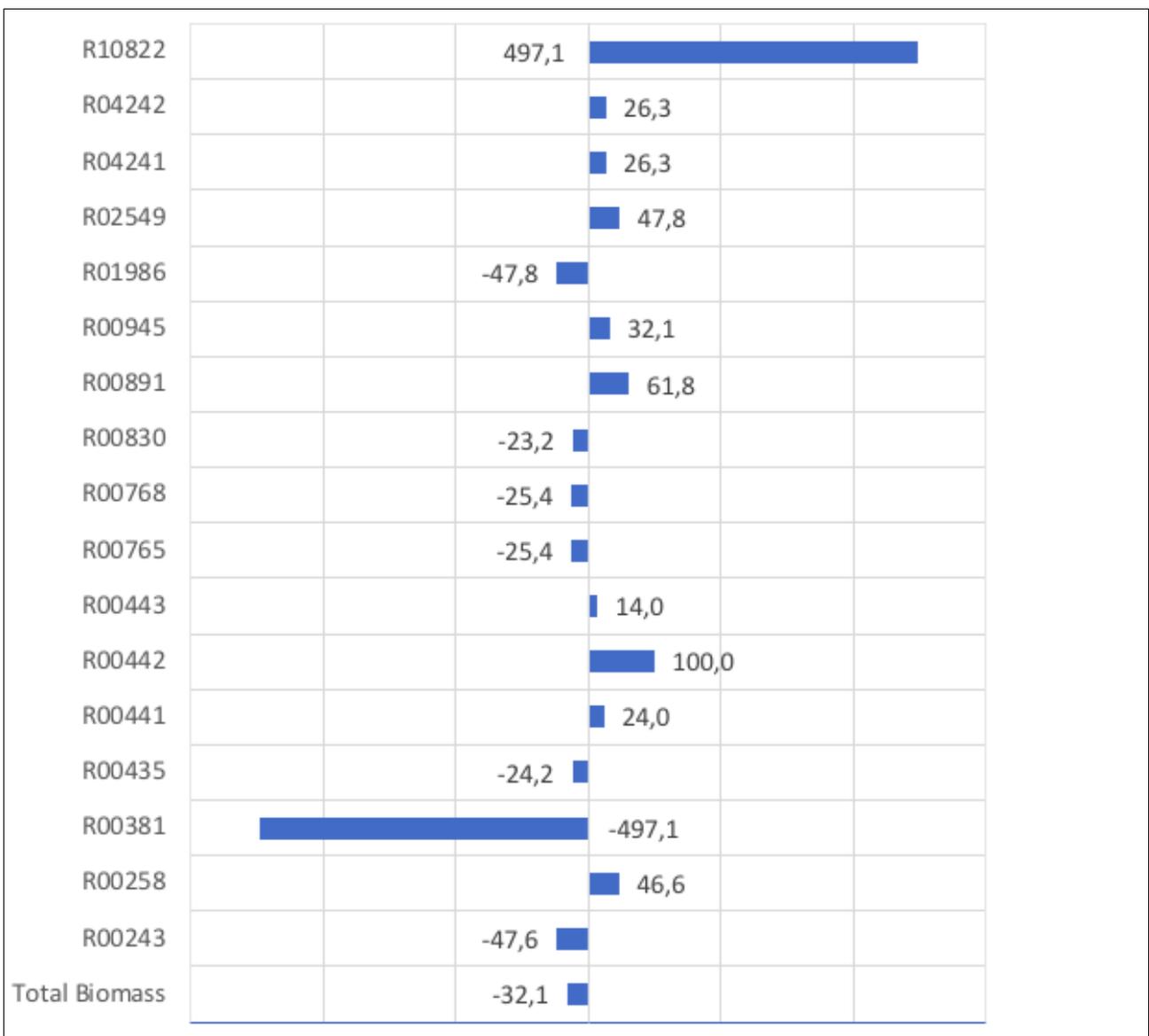


Fig. 3