http://lib.uliege.be https://matheo.uliege.be

Data-driven Inference of a TF-regulated Metabolic Network for photosynthetic organism modelling: Chlamydomonas Reinhardtii.

Auteur : Waterplas, Dries
Promoteur(s) : Meyer, Patrick
Faculté : Faculté des Sciences
Diplôme : Master en biochimie et biologie moléculaire et cellulaire, à finalité spécialisée en bioinformatique et modélisation
Année académique : 2017-2018
URI/URL : http://hdl.handle.net/2268.2/5438

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

Abstract

Extensive research efforts are currently being spent upon the photosynthetic unicellular alga **Chlamydomonas** ***reinhardtii*** because it presents multiple interesting biological properties as well as potentially highly valuable industrial interests through its capacity to produce substances that can be converted into **biofuel**, namely triacylglycerol (TAG) and biohydrogen (H_2). Such product yields are entirely dependent of the environmental conditions to which the cell is exposed.

Metabolic pathways involved in photosynthesis and biosynthesis of lipids, hydrogen or growth are typically regulated by complex interdependent mechanisms involving multiple fronts of omics (genomics, transcriptomics, proteomics, …), meaning that in order to be able to modelize these pathways we need to study the cell as a **system**, as too many factors are involved to consider reductionist approaches. The increasing success of high-throughput technologies for collecting biological data combined with the increasing understanding of specific complex metabolic pathways now allows research to be conducted on a **genomic scale**. Resulting **systems biology** uses computational and mathematical principles for the modelization of very large complex biological systems, thus providing models able to capture higher-complexity biological events influencing the expression of desired phenotypes, such as TAG production in Chlamydomonas.

Here, we apply **Constraint-Based Modelling** (CBM) techniques on a genome-scale metabolic network reconstruction of C. *reinhardtii,* and assess the predictive power of such constraint-based models by comparing **light-constrained predictions of growth** to experimentally observed cultures under light-limiting growth conditions.

In order to enable the model to capture metabolic shifts in response to important gene-expression changes, as is the case in C. *reinhardtii* when exposed to abiotic stress required for biofuel maximization, we integrate **transcriptomic constraints** upon the existing network. The additional omic layer of regulation allows us to bring in **TF-regulations** upon target genes involved in metabolic pathway expressions. We also proceed to simulate the phenotypic effect of **gene knockouts** by predicting the metabolic reaction of the cell to the perturbation, and show an example where, unlike the state-of-the-art model, the TF-regulated network is capable of detecting metabolic flux perturbations to gene expression changes, thus identifying the scope of genes involved in studied pathways.

Acknowledgements

I would like to thank my promotor Patrick Meyer for propelling me into the world of systems biology, where I have truly found huge interest. I would also like to thank him for the trust he has shown me throughout the year, seldom doubting my work and often letting me try out my own ideas for research despite my tendency to explore various diverging tracks.

For the people in the rest of the building, I also thank my colleagues and friends, and more specifically Manuel Noll, who has truly held my back throughout the whole year and who introduced me into late-night C++ lessons after work, for fun!

There are many people I would like to thank in these acknowledgements, as there are many who have accompanied me in my academic studies up to where I am today. I would expressly like to first thank the University of Liège for allowing me to join the Bioinformatics Master program after my Veterinarian Bachelor studies.

Many thanks go to my long-time friends and especially the ones who have helped me through the redaction of this thesis, mainly Florent, Jérome and Etienne but also Symphorine, Florine, Clément and Peeters.

Finally, my parents, without whom these studies would not have been possible in the first place. I thank them both for their constant trust in my actions and abilities in everyday-life.

Abbreviations list

- **AA** = Amino-Acids
- **BBMC** = Biochemistry and Molecular & Cell Biology
- **CBM** = Constraint-Based Modelling
- **COBRA** = Constraint-Based Reconstruction and Analysis
- **Cre** = Chlamydomonas reinhardtii
- **DREAM** = Dialogue on Reverse Engineering Assessment and Methods
- **FBA** = Flux Balance Analysis
- **FCA** = Flux Coupling Analysis
- **e-Flux** = Expresion Flux
- **ElMo** = Elementary Modes
- **ExPa** = Extreme Pathways
- **GIMME** = Gene Inactivity Moderated by Metabolism and Expression
- **iMAT** = Integrative Metabolic Analysis Tool
- **KNN** = K-Nearest Neighbors
- **LBFBA** = Linear-Bound FBA
- **LP** = Linear Programming
- **MADE** = Metabolic Adjustment by Differential Expression
- **MILP** = Mixed-Integer Linear Programming
- **MOMA:** Minimization of Metabolic Adjustment
- **MRMR** = Maximum Relevance Minimum Redundancy
- **MRS** = Metabolic-Regulatory Solution-space
- **NCBI** = National Center for Biotechnology Information

- **NPQ** = Non-Photochemical Quenching
- **OD** = Optical Density
- **pFBA** = Parsimonious Flux-Balance Analysis
- **plantTFDB** = Plant Transcription Factor Database
- **plnTFDB** = Plant Transcription Factor Database
- **PROM** = Probabilistic Regulation Of Metabolism
- **RNA-seq** = RNA sequencing
- **ROC (curve)** = Receiving Operating Characteristics (for binary classification)
- **ROOM** = Regulatory On/Off Minimization
- **SBML** = Systems Biology Markup Language
- **SQL** = Structured Query Language
- **SRA** = Sequence Read Archive
- **srFBA** = Steady-state Regulatory FBA
- **TAG** = Triacylglycerol
- **TAP** = Tris-Acetate-Phosphate media
- **TF** = Transcription Factor
- **TF-KO** = TF-knockout
- **TGP** = Tris-Glucose-Phosphate media
- **TMP** = Tris-Minimum-Phosphate media
- **TR** = Transcription Regulators | Transcriptional Regulation
- **TRN** = Transcriptional Regulation Network
- **UML** = Unified Modelling Language
- **Y2H** = Yeast Two-Hybrid