Metabolomics project
EMBL, Heidelberg 17-21 Jun 19
Project report

Kateřina Břicháčková
Zuzana Halenková
David Novák

Background

We studied computational methods in metabolomics of microbial communities and applied tools developed at EMBL and elsewhere to a diverse array of genomic and metabolic data.

A particular motivation for constructing metabolic models is the importance of metabolism in how organisms and communities function. Moreover, the models allow us to define and determine optima in controlled systems.

Flux Balance Analysis

When characterising a metabolic system, we estimate flux (chemical reaction rate), as determined by enzyme availability, substrate availabilities via uptake from the environment or differential transcription rates, and more. Inhibition and feedback loops inject considerable complexity into the system. Consequently, relying on a naïve kinetic model will likely fail to solve the problem.

Instead, flux balance analysis (FBA) explores a solution space where each dimension corresponds to flux in one reaction. We impose a mass balance constraint (metabolites do not vanish) and other constraints on flux based on supposed biological realities. This narrows down the space of possible solutions considerably. We also add an objective function which is maximised. The result is a linear programming problem.[1] Our analysis relies on the IBM CPLEX Solver.

Additionally, parsimonious flux balance analysis (pFBA) seeks to minimise the number of reactions occurring in the system in order to provide a simpler solution to the problem.

Genome-scale modelling

With genome-scale metabolic models, we try to account for as many metabolic reactions as feasible. A genome-scale stoichiometric matrix is constructed and we solve for fluxes of reactions. In microbial communities, we confront the problem of cell and population growth. In order to simulate it, we create a pseudo-reaction, termed the biomass reaction. This reaction converts key metabolites and cofactors to their chemical 'counterparts' and some amount of biomass (*eg.* 1 gram). This, now, can be taken as the objective function, as long as we are confident in postulating growth as the objective.

Metabolic models of communities

As we transition from single species to whole communities, we make use of species-specific data from manually curated databases to construct a comprehensive model of interactions in terms of metabolite exchanges. The mathematics of community models are similar, but new findings can be inferred. Consider the minimum requirements (*i.e.* number of metabolites which need to be taken in) in a non-interactive model, and a model where metabolite exchanges are permitted (strictly an integer problem). Therefrom we can determine interaction potentials of the given species. On the other hand, overlaps in metabolic requirements among individual species suggest competition.

Darwinian thinking in metabolic models

Furthermore, we find that some metabolites, including amino acids, are exchanged frequently, presumably to provide stability in available metabolites in the face of nutritional fluctuations. This provides a novel interpretation of Darwin's naturalisation hypothesis: a new species that joins cooperating community needs to be able to cooperate in order to succeed. We arrive at an elegant synthesis of the *red queen theory* (actors in evolution have to remain in motion to retain their standing in a community) and the *black queen theory* (the most expensive processes, including metabolic reactions, are prone to being dumped once they can be delegated to somewhere else).

Top-down reconstruction of community-level models

In order to obtain a reasonable community-level model of interactions, one has to check all the required species-level models and synthesise them to build, in a bottom-up fashion, a system which makes sense biologically. This is a time-consuming and potentially tedious. Instead, *CarveMe*, a tool developed at EMBL, automates this step, using a manually curated universal model which is altered according to input data.[2]

The following step is to infer metabolite exchanges between species. This is facilitated by another EMBL tool, *SMETANA*. Running multiple simulations iteratively, each metabolite exchange is specified, among others, by a donor, a receiver and a *SMETANA* score, indicating the fraction of simulations in which the exchange was present (striving to compensate for stochastic initialisation).[3]

Methods

The aforementioned tools feature heavily in our analyses. We mostly used Python 3.6 for our projects. Genome-scale metabolic models for single-species were built using *CarveMe* 1.2.2. Minimal medium was identified and growth phenotypes were simulated by parsimonious flux balance analysis. For that we used *FRAMED*, a Python package for analysis and simulation of metabolic models. A microbial community model was then constructed using *SMETANA* 1.0. We analysed obtained inter-species interactions in Jupyter Notebook and via *McNet*, a web-based tool for the exploration of interaction networks. At last, we searched available literature for evidence and attempted to formulate biological explanations of results.

Project 1 – David Novák

Mucous membrane bacteria with an outgroup

A community of 4 mucous membrane-associated bacterial species (*Neisseria mucosa*[4], *Rodentibacter heylii*[5], *Rothia dentocarosa*[6], *Rothia mucilaginosa*[7]) and 1 species found in phytoplankton (*Marivita cryptomonadis*[8]) was selected, with the latter acting as a form of outgroup in the analysis.

Species-level models were built based on genome using *CarveMe* and minimal media were identified. Subsequently, a community-level model of metabolic interactions was constructed. Sorting metabolic exchanges by *SMETANA* scores showed L-Asparagine, Benzoate and L-Arginine as the most persistently occuring transfers. However, the SMETANA scores for most other exchanges still rarely grew above 0.1.

Exchanges of compounds were further investigated. A network of L-Alanine and D-Alanine exchanges featured in a few of the model reconstructions, with most reliable transfers involving *Neisseria mucosa* as receiver. D-Alanine plays role in peptidoglycan production in formation of biofilms[9]. It has potential significance in protection against dental caries in human mouth bacteria. L-Alanine, furthermore, can be converted to D-Alanine by its corresponding racemase.

Furthermore, L-Asparagine exchanges were persistent, with *SMETANA* scores of 1 for the transfer from *M. cryptomonadis*, *N. mucosa* and *R. dentocariosa* donating to *R. mucilaginosa*. L-Asparagine can be converted via asparagine synthase to glutamate (otherwise synthesised from alpha-ketoglutarate from the TCA cycle). Glutamate then acts as a frequent amino-group donor.

Many other exchanges were described, with *R. mucilaginosa* taking on the frequent role of a receiver in high-scored exchanges. Since *R. dentocariosa* comes from the same genus, I was interested in a model where *R. mucilaginosa* was excluded, hypothesising that *R. dentocariosa* will become a significant consumer in the given community, sucking up metabolites, and that in a *mucilaginosa*-free model, *dentocariosa* would perhaps assume that position. The evidence seemed somewhat in favour of that hypothesis, with a newly persistent L-Glutamate exchange (in 92 % of *SMETANA* iterations) from *R. heylii* to *R. dentocariosa* and other exchanges where *R. dentocariosa* became the chief receiver where *R. mucilaginosa* used to dominate. On the other hand, in the case of benzoate, *N. mucosa* is a donor in both models, but whereas the full model shows *R. mucilaginosa* as the receiver, the new model shows a different receiver, *R. heylii*, with *R. dentocariosa* not stepping in to receive this metabolite.

While it is challenging to draw conclusions from the analysis of this given community, the data generally point toward a competitive relationship between *R. mucilaginosa* and *R. dentocariosa*, as I hypothesised upon observing the initial model.

To obtain plots for all metabolite transfers, the software means to create custom graphs or any of the data used in my analysis, please contact me at *davidnovakcz@hotmail.com*.

Project 2 – Kateřina Břicháčková

The aim of this project is to analyse 3 bacteria species from different environments and the way they interact with each other. The selected species are following:

- *Haemophilus haemoglobinophilus*
- *Veilonella parvula*
- *Lactobacillus iners*

Obtained inter-species interactions were filtered, required SMETANA score was at least 0.5.

The most noticeable thing in the analysis is that *L. iners* is a significant receiver and does not donate anything at all. This result is in agreement with the literature; according to several articles, *L. iners* has unusually small genome (ca. 1 Mbp)[10] similar to that of human symbionts and parasites and lacks 17 protein families present in other *lactobacilli*.[11] 15.6% of its genome is dedicated to various transport mechanisms, suggesting it may acquire many compounds needed for survival from the host or community.[11]

L. iners lacks genes necessary to synthesize any amino acids de novo.[12] That also agrees with the interactions analysis: out of 13 compounds received by *L. iners*, 10 are amino acids. The analysis of 5 other *lactobacilli* species[13] suggests that the amino acids which are missing (i.e. are not received from other 2 species) might be unnecessary for normal growth.

Another compound received by *L. iners* is succinate. According to literature, *L. iners* has an incomplete citric acid cycle containing only fumarate reductase (converts fumarate into succinate).[14] Succinate is not present in minimal medium, which could mean it is not essential and can be made from fumarate, but it can be still helpful to obtain it from environment.

Xanthosine is present in minimal medium for *L. iners* and there is no reaction producing it. In the community model, it is received from *H. haemoglobinophilus* instead.

One surprising thing is the absence of ATP synthase reaction in *L. iners* model. From metabolite balances we see that ATP is produced in pyruvate kinase and phosphoglycerate kinase reactions. This is in contrary to the UniProt database, where different subunits of ATP synthase for *L. iners* are present.[15] These entries were not manually annotated and were inferred from homology, but the complete absence of ATP synthase in *L. iners* is still unlikely. Further investigation would be needed to come to a conclusion.

H. haemoglobinophilus and *V. parvula* exchange 4 compounds in total. Three of them are amino acids and the last one is 2-oxoglutarate. In the single-species model for *V. parvula*, 2-oxoglutarate is contained in minimal medium and there are only transport reactions, no production. The 2-oxoglutarate dehydrogenase reaction is the most highlighted one in the escher map, plus there are other reactions using 2-oxoglutarate, suggesting that it is a very important compound for the growth of *V. parvula*.

To sum up, obtained single-species genome-scale metabolic models and the community model mostly agree with available literature and biological evidence can be found. The absence of ATP synthase in *L. iners* is a possible mistake in the model and should be investigated further.

Project 3 - Zuzana Halenková

In this subproject, a genome-scale metabolic model was built for each of five selected bacteria species and then these were further used for analysis of metabolic interactions in potential community comprising of these species. Five selected species included two *Acinetobacter* species (namely *A. harbinensis* and *A. johnsonii*), two *Corynebacterium* species (namely *C. durum* and *C. fournierii Marseille*) and *Micrococcus luteus*.

Visual inspection of flux-balance analysis maps generated by build_escher_map function from python framed module suggested close relatedness of both *Acinetobacter* species as the distribution of fluxes showed up indistinguishable over all pathways. In comparison with this, maps built for *Corynebacterium* species did not share that kind of similarity with *C. fournierii* emphasizing ATP synthesis over L-lactate import which was accentuated in *C. durum* model. Hypothesis of close relatedness of *A. harbinensis* and *A. johnsonii* within *Acinetobacter* genus is supported by published phylogenetic tree[16], as well as the larger phylogenetic distance of *C. fournierii* and *C. matruchotii* (the closest neighbor of *C. durum*[17]) within *Corynebacterium* genus[18].

In the next step, community models were built using SMETANA. Resulting inter-species metabolite exchanges were filtered by SMETANA score (required SMETANA score greater than 0.1) and also interactions concerning abundantly exchanged inorganic metabolites were removed.

Initially a community consisting of *A. johnsonii*, *C. durum* and *M. luteus* was simulated. In this model, *M. luteus* adopted role of universal donor while *C. durum* maintained the role of general receiver. *A. johnsonii* came out of this model as an independent player, not receiving but also not donating any metabolites. In total five metabolite exchanges were observed between *M. luteus* and *C. durum* with SMETANA score over 0.1 – pantothenate exchange with SMETANA score equal to 1.0 and four amino acid exchanges (alanine, glutamine, asparagine, lysine).

In the next phase, presence of relationships observed in the three-species community was examined in four- and five-species communities, adding *A. harbinensis* and *C. fournierii* to the model (respectively). The role of *C. durum* as a major receiver persisted even in four-species model. Position of *M. luteus* changed from being a sole donor to being also a receiver of glycerol 3-phosphate donated by both *Acinetobacter* species. Nevertheless, by adding *C. fournierii* to the model these donorships of glycerol 3-phosphate disappeared. In this five-species model, both *Corynebacterium* species acted as strong receivers, while *M. luteus* once again adopted position of central donor for both *Corynebacterium* species. Inter-species interactions with greatest SMETANA score once again included donorships of pantothenate and amino acids (alanine, asparagine, glutamine) by *M. luteus* to both *Corynebacterium* species.

The position of *M. luteus* as a donor of pantothenate and amino acids is supported by previously published work where the authors showed, that *M. luteus* strains do not impose an absolute requirement on pantothenate and aforementioned amino acids for their growth.[19]

All three models agreed on *M. luteus* being a major donor in the community and both *Corynebacterium* species being major receivers. Influence of genome size on this relationship can be dismissed as genome size of *M. luteus* is greater than genome size of *C. fournierii* but lesser than genome size of *C. durum* and as genome sizes of all subjected species fall within the range of 1.2 Mb.[20]

Conclusion

We worked with various bacterial species using tools mentioned above, analysing metabolic models for selected species. Furthermore, we combined these to build small (from 3 to 5 species) microbial communities models and tried to uncover the biological explanation of the results. Even though we set out to achieve similar objectives, the focal points of our projects turned out not to be identical. They ranged from searching the literature for species-specific to metabolite-specific phenomena and correlating them with our observations, through statistical analyses and building on top of existing options for visualisation to obtain better tools for inspecting the data.

References

Background

1. Orth, J. D., Thiele, I., & Palsson, B. O. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248. https://doi.org/10.1038/nbt.1614
2. Machado, D., Andrejev, S., Tramontano, M., & Patil, K. R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research*, *46*(15), 7542–7553. https://doi.org/10.1093/nar/gky537
3. Machado, D., Andrejev, S., Tramontano, M., & Patil, K. R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research*, *46*(15), 7542–7553. https://doi.org/10.1093/nar/gky537

PROJECT 1: ***Mucous membrane bacteria with an outgroup***

4. Osses, D. F., Dijkmans, A. C., Van Meurs, A. H., & Froeling, F. M. (2017). Neisseria Mucosa: A New Urinary Tract Pathogen? *Current Urology*, *10*(2), 108–110. https://doi.org/10.1159/000447161
5. Benga, L., Sager, M., & Christensen, H. (2018). From the [Pasteurella] pneumotropica complex to Rodentibacter spp.: an update on [Pasteurella] pneumotropica. *Veterinary Microbiology*, *217*, 121–134. https://doi.org/10.1016/j.vetmic.2018.03.011
6. Ricuarte, J. C., Klein, O., LaBombardi, V., Martinez, V., Serpe, A., & Joy, M. (2001). Rothia dentocariosa Endocarditis Complicated by Multiple Intracranial Hemorrhages. *Southern Medical Journal*, *94*(4), 438–440. https://doi.org/10.1097/00007611-200194040-00018
7. Maraki, S., & Papadakis, I. S. (2015). Rothia mucilaginosa pneumonia: A literature review. *Infectious Diseases*, *47*(3), 125–129. https://doi.org/10.3109/00365548.2014.980843
8. Hwang, C. Y., Bae, G. D., Yih, W., & Cho, B. C. (2009). Marivita cryptomonadis gen. nov., sp. nov. and Marivita litorea sp. nov., of the family Rhodobacteraceae, isolated from marine habitats. *International Journal of Systematic and Evolutionary Microbiology*, *59*(7), 1568–1575. https://doi.org/10.1099/ijs.0.005462-0
9. Qiu, W., Zheng, X., Wei, Y., Zhou, X., Zhang, K., Wang, S., … Li, M. (2016). d-Alanine metabolism is essential for growth and biofilm formation of Streptococcus mutans. *Molecular Oral Microbiology*, *31*(5), 435–444. https://doi.org/10.1111/omi.12146

PROJECT 2 - Kateřina Břicháčková

10. Petrova, M. I., Reid, G., Vaneechoutte, M. & Lebeer, S. (2017). Lactobacillus iners: Friend or Foe? *Trends in Microbiology* 25,182–191
11. Macklaim, J. M., Gloor, G. B., Anukam, K. C., Cribby, S. & Reid, G. (2010) At the crossroads of vaginal health and disease, the genome sequence of Lactobacillus iners AB-1. *Proceedings of the National Academy of Sciences* 108, 4688–4695
12. Mendes-Soares, H., et al. (2014). Comparative Functional Genomics of Lactobacillus Spp. Reveals Possible Mechanisms for Specialization of Vaginal Lactobacilli to Their Environment. *Journal of Bacteriology*, vol. 196, no. 7, 2014, pp. 1458–1470

13. Morishita, T, et al. (1981) Multiple Nutritional Requirements of Lactobacilli: Genetic Lesions Affecting Amino Acid Biosynthetic Pathways. *Journal of Bacteriology, U.S. National Library of Medicine*
14. KEGG PATHWAY: Citrate cycle (TCA cycle) - Veillonella parvula
Available at: https://www.kegg.jp/kegg-bin/show_pathway?vpr00020. (Accessed: 20st June 2019)
15. UniProt: Search ATP synthase in L. iners. Available at: https://www.uniprot.org/uniprot/?query=atp+synthase+organism%3A%22lactobacillus+iners%22&sort=score. (Accessed: 8th July 2019)

PROJECT 3 - Zuzana Halenková

16. Li, W., Zhang, D., Huang, X., & Qin, W. (2014). Acinetobacter harbinensis sp. nov., isolated from river water. *International journal of systematic and evolutionary microbiology, 64 Pt 5*. 1507-13.
17. Riegel, P & Heller, Remy & Prévost, Gilles & Jehl, françois & Monteil, H. (1997). Corynebacterium durum sp. nov., from Human Clinical Specimens. *International journal of systematic bacteriology, 47*. 1107-11.
18. Baek, I., Kim, M., Lee, I., Na, S. I., Goodfellow, M., & Chun, J. (2018). Phylogeny Trumps Chemotaxonomy: A Case Study Involving Turicella otitidis. *Frontiers in microbiology, 9*. 834.
19. Farrior, J.W., & Kloos, W.E. (1975). Amino Acid and Vitamin Requirements of Micrococcus Species Isolated from Human Skin. *International journal of systematic bacteriology, 25*. 80-82.
20. Genome sizes retrieved from www.ncbi.nlm.nih.gov/genome database (Accessed: 9th July 2019)