Ph. D. thesis summary

Exploring the network of genetic interactions and evolutionary innovations with metabolic modeling

Balázs Szappanos

Supervisor: Balázs Papp, Ph.D.

Ph.D. School of Biology

Biological Research Centre, Hungarian Academy of Sciences

Szeged, Hungary

2016

Introduction

Recently the development of experimental techniques led to the exponential increase in the available biological data. This allowed the investigation of biological systems via a holistic approach and led to the emergence of systems biology. Metabolism is one of the most well-known cellular subsystems and therefore a suitable model for exploring and answering the questions of systems biology. It is also possible to build mathematical models of metabolism which are capable of predicting the growth effect of different perturbations like gene deletion. Evolutionary biology could also profit from the systems biology approach. Understanding the mapping between genotype and phenotype is one of the main goals of evolutionary biology. Genotype-to-phenotype mapping on a systematic scale is possible with metabolic models (Papp et al., 2011).

Our work aims to answer some of the basic questions of evolutionary genetics and evolutionary innovations using metabolic models. First, we investigated the interaction between the growth effects of the deletion of metabolic genes in a pairwise manner. We also described the systematic patterns among these genetic interactions (Szappanos et al., 2011). Second, we examined the importance of underground activities of enzymes in the evolution of metabolic pathways (Notebaart et al., 2014).

Investigating genetic interactions by using metabolic modelling

Genetic interaction occurs when the effect of the simultaneous deletion of two genes is different than what we expect based on the effects of the single gene deletions. Genetic interaction can be negative when the effect of the double mutant is stronger than expected and positive when the effect is weaker. Genetic interactions have an important role in the development of diseases with genetic background (Lehner, 2007) and describing unknown genes (Boone et al., 2007). Genetic interactions can also shape the adaptive fitness landscape and therefore affect the trajectory of adaptation (Phillips, 2008).

During our work we simulated the genetic interactions by applying the flux balance analysis (FBA) method on the yeast metabolic network (Segrè et al., 2005). Predictions of the model were compared with a systematic experimental dataset of genetic interactions (data produced in the lab of our collaborator, Charles Boone). This dataset contains genetic interaction data for more than 176 000 metabolic gene pairs.

Aims

First, we intended to test the capability of the metabolic model to predict individual genetic interactions and the general patterns of the network of genetic interactions. Furthermore we planned to develop an automated method to refine the metabolic model by correcting its

false predictions. Finally we wanted to test the hypotheses of the model refinement algorithm.

Methods

- Using FBA to predict the growth effect of single and double gene deletions.
- Calculating functional pleiotropy of the metabolic genes using FBA.
- Development of a genetic algorithm-based method to refine the predictions of the model.

Results

1. Genetic interaction degree correlates with gene dispensability

Experimental genetic interaction datasets show an interesting pattern: the more interactions a gene has the higher the growth effect of the deletion of the gene will be (Costanzo et al., 2010). We confirmed this pattern with the metabolic model as well. After that we used the model to find a mechanistic explanation to this phenomenon. One potential explanation is that a gene with multiple genetic interactions has multiple biological functions (show high pleiotropy) as well. Our results showed that there is a strong negative correlation between the number of biosynthetic processes affected by a gene (that is, pleiotropy) and the dispensability of the gene. Besides, pleiotropy and the genetic interaction degree show positive correlation. Based on this we assume that the effect of deletion of

genes with multiple biological functions (highly pleiotropic) is influenced by several other genes, each which compensate a different biological function of the highly pleiotropic gene.

2. Metabolic model captures only a minority of empirical interactions

Genetic interactions among metabolic genes were calculated using FBA and compared with the experimental dataset coming from the Boone lab. We found that experimentally identified interactions are highly overrepresented among predicted interactions confirming that *in silico* genetic interactions are biologically relevant. However, the model captures only a small fraction of the interactions in the experimental dataset. We intended to improve this shortcoming of the model.

3. Automated refinement of the metabolic model

In order to improve the accuracy of the model to predict genetic interactions we developed a machine-learning method that automatically generates hypotheses to explain negative interaction between genes. Our method minimizes model mispredictions globally by creating and evolving a population of modified models. Several modifications were suggested that, together, considerably improved the fit of the model to our genetic interaction map (100–267% increase in recall and 44–59% increase in precision).

4. Experimental validation of the refinement of the *in silico* NAD biosynthesis pathways

Our automated algorithm revealed that the removal of only one or two reactions from the model corrects the prediction of four negative interactions between alternative NAD biosynthesis pathways. That is, the algorithm deleted one of the three NAD biosynthesis pathways available in the yeast model. We failed to find yeast homologs for the enzymes of this pathway suggesting that the presence of this pathway in the yeast model is the result of misannotation. Next, we interrogated the metabolic model to deduce specific predictions for experimental testing. We tested these predictions experimentally and confirmed that the investigated third NAD biosynthesis pathway is indeed not available in yeast. Laboratory experiments were carried out by my colleague Béla Szamecz. Our results show that machine learning methods coupled with high-throughput experimentation, have the potential to close the iterative cycles of generating and testing new hypotheses, leading to at least partial automation of biological discoveries.

The evolutionary importance of underground metabolic reactions

It is well known that many enzymes have limited substrate specificities and can catalyze, albeit at low rates, reactions other than those for which they have evolved. These so called underground

reactions are prevalent, their activities are low but mutations can increase it (Khersonsky and Tawfik, 2010).

Understanding the way how new metabolic pathways evolve is one of the main goals of evolutionary biology and systems biology. The prevailing theory is that new pathways are patched together from already existing enzymes coming from different parts of the network. Underground activities have a key role in this evolutionary process. Our aim was to understand the importance of underground reactions in the adaptation to new environmental conditions.

Aims

Our knowledge about how underground reactions can act as an evolutionary raw material in the evolution of new metabolic functions is limited. Do underground reactions remain isolated, or can they potentially be wired into the native network and allow the organism to survive in novel environments? How often can underground reactions cause a new beneficial phenotype? To answer these questions we systematically collected a list of well-known underground activities of *E. coli* enzymes and integrated them in a genome-wide *E. coli* metabolic model.

Methods

- Collecting *E. coli* underground reactions and integrating them in the metabolic network using the BRENDA database and literature data.

- Applying FBA method to simulate the fitness effect of underground reactions.

- Analyzing underground activities with protein overexpression experiments.

Results

1. Underground reactions can be wired into the native *E. coli* metabolic network

We tested whether the collected underground reactions share their substrate and product with the native reactions or they are isolated. We found that only 22% of the underground reactions are isolated from the rest of the metabolic network and 45% can be fully integrated therefore capable of building new biochemical pathways in association with the native reactions.

2. Underground reactions can provide fitness advantage under certain conditions

In order to estimate the evolutionary potential of underground reactions we calculated their fitness advantage in 2754 *in silico* conditions. We found 19 conditions where the presence of an underground reaction is essential extending the number of conditions

available for the *E. coli* model with 2.9%. This result shows that underground reactions have a great evolutionary potential considering that the underground reactions extend the metabolic model with merely 10.8%. Besides, in further 31 environments underground reactions provided a clear quantitative growth advantage.

3. Investigating the growth effect of single underground reactions

In the next step we calculated the fitness effect of integrating each underground reaction separately to the native network. We found that in every condition where underground reactions are essential the presence of a single underground reaction is enough. Besides, it is rare that the same reaction is beneficial in multiple conditions, that is, the growth effect of the underground reactions is conditional.

4. *In silico* predictions of the effect of underground reactions overlap with experimental data

In order to validate our simulations we compared our data with the results of a systematic experimental screen. Our experiments are based on the fact that the effect of underground activities can be enhanced by overexpressing the enzyme. We overexpressed 4269 *E. coli* gene in 194 conditions, each containing a different carbon source. Laboratory experiments were carried out by my colleague Bálint Kintses. We found that our *in silico* model could predict 44% of the conditions emerged from the experimental screen. Moreover the overlap between the experimental and predicted condition – gene

pairs was 2 orders of magnitude higher than randomly expected. These results suggest that our network of *in silico* underground reactions already covers a significant part of the evolutionary raw material available for short-term adaptation to novel environments.

References

Boone, C., Bussey, H., and Andrews, B.J. (2007). Exploring genetic interactions and networks with yeast. Nat. Rev. Genet. *8*, 437–449.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L.Y., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. Science *327*, 425–431.

Khersonsky, O., and Tawfik, D.S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. Annu. Rev. Biochem. *79*, 471–505.

Lehner, B. (2007). Modelling genotype–phenotype relationships and human disease with genetic interaction networks. J. Exp. Biol. *210*, 1559–1566.

Notebaart, R.A., Szappanos, B., Kintses, B., Pál, F., Györkei, Á., Bogos, B., Lázár, V., Spohn, R., Csörgő, B., Wagner, A., et al. (2014). Network-level architecture and the evolutionary potential of underground metabolism. Proc. Natl. Acad. Sci. *111*, 11762–11767.

Papp, B., Notebaart, R.A., and Pál, C. (2011). Systems-biology approaches for predicting genomic evolution. Nat. Rev. Genet. *12*, 591–602.

Phillips, P.C. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. Nat. Rev. Genet. *9*, 855–867.

Segrè, D., DeLuna, A., Church, G.M., and Kishony, R. (2005). Modular epistasis in yeast metabolism. Nat. Genet. *37*, 77–83.

Szappanos, B., Kovács, K., Szamecz, B., Honti, F., Costanzo, M., Baryshnikova, A., Gelius-Dietrich, G., Lercher, M.J., Jelasity, M., Myers, C.L., et al. (2011). An integrated approach to characterize genetic interaction networks in yeast metabolism. Nat. Genet. *43*, 656–662.

List of publications

Szappanos, B.*, Fritzemeier, J.*, Csörgő, B.*, Lázár, V., Lu, X., Fekete, G., Bálint, B., Herczeg, R., Nagy, I., Notebaart, R.A., et al. (2016). Adaptive evolution of complex innovations through stepwise metabolic niche expansion. *Nat Commun* 7, 11607.
IF: 11.329

Karcagi, I.*, Draskovits, G.*, Umenhoffer, K., Fekete, G., Kovács, K., Méhi, O., Balikó, G., **Szappanos, B.**, Györfy, Z., Fehér, T., et al. (2016). Indispensability of Horizontally Transferred Genes and Its Impact on Bacterial Genome Streamlining. *Mol Biol Evol* 33, 1257–1269.
IF: 13.649

Notebaart, R.A.*, **Szappanos, B.***, Kintses, B.*, Pál, F., Györkei, Á., Bogos, B., Lázár, V., Spohn, R., Csörgő, B., Wagner, A., et al. (2014). Network-level architecture and the evolutionary potential of underground metabolism. *PNAS* 111, 11762–11767.
IF: 9.674

Lázár, V.*, Nagy, I.*, Spohn, R.*, Csörgő, B., Györkei, Á., Nyerges, Á., Horváth, B., Vörös, A., Busa-Fekete, R., Hrtyan, M., Bogos, B., Méhi, O., Fekete, G., **Szappanos, B.**, Balázs, K., Papp, B., Pál, Cs. (2014). Genome-wide analysis captures the determinants of the antibiotic cross-resistance interaction network. *Nat Commun* 5.
IF: 11.470

VanderSluis, B.*, Hess, D.C.*, Pesyna, C., Krumholz, E.W., Syed, T., **Szappanos, B.**, Nislow, C., Papp, B., Troyanskaya, O.G., Myers, C.L., et al. (2014). Broad metabolic sensitivity profiling of a prototrophic yeast deletion collection. *Genome Biol.* 15, R64.
IF: 10.810

Lázár, V.*, Pal Singh, G., Spohn, R., Nagy, I., Horváth, B., Hrtyan, M., Busa-Fekete, R., Bogos, B., Méhi, O., Csörgő, B., Pósfai, G., Fekete, G., **Szappanos, B.**, Kégl, B., Papp, B., Pál, C. (2013). Bacterial evolution of antibiotic hypersensitivity. *Molecular Systems Biology* 9.
IF: 14.099

Szalay-Bekő, M.*, Palotai, R.*, **Szappanos, B.**, Kovács, I.A., Papp, B., and Csermely, P. (2012). ModuLand plug-in for Cytoscape: determination of hierarchical layers of overlapping network modules and community centrality. *Bioinformatics* 28, 2202–2204.
IF: 5.323

Lázár, V.*, Ecsedi, S., Vízkeleti, L., Rákosy, Z., Boross, G., **Szappanos, B.**, Bégány, A., Emri, G., Adány, R., and Balázs, M. (2012). Marked genetic differences between BRAF and NRAS mutated primary melanomas as revealed by array comparative genomic hybridization. *Melanoma Res.* 22, 202–214.
IF: 2.518

Nagy, L.G.*, Házi, J., **Szappanos, B.**, Kocsubé, S., Bálint, B., Rákhely, G., Vágvölgyi, C., and Papp, T. (2012). The Evolution of Defense Mechanisms Correlate with the Explosive Diversification of Autodigesting Coprinellus Mushrooms (Agaricales, Fungi). *Syst Biol* 61, 595–607.
IF: 12.169

Papp, B.*, **Szappanos, B.**, and Notebaart, R.A. (2011). Use of genome-scale metabolic models in evolutionary systems biology. *Methods Mol. Biol* 759, 483–497.
IF: NA

Szappanos, B.*, Kovács, K.*, Szamecz, B., Honti, F., Costanzo, M., Baryshnikova, A., Gelius-Dietrich, G., Lercher, M.J., Jelasity, M.,

11

Myers, C.L., et al. (2011). An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat. Genet* 43, 656–662.
IF: 35.532

Angyán, A.F.*, **Szappanos, B.***, Perczel, A., and Gáspári, Z. (2010). CoNSEnsX: an ensemble view of protein structures and NMR-derived experimental data. *BMC Struct Biol* 10, 39.
IF: 2.258

Szappanos, B.*, Süveges, D., Nyitray, L., Perczel, A., and Gáspári, Z. (2010). Folded-unfolded cross-predictions and protein evolution: the case study of coiled-coils. *FEBS Lett* 584, 1623–1627.
IF: 3.601

Gáspári, Z.*, Várnai, P., **Szappanos, B.**, and Perczel, A. (2010). Reconciling the lock-and-key and dynamic views of canonical serine protease inhibitor action. *FEBS Lett* 584, 203–206.
IF: 3.601

* first authors