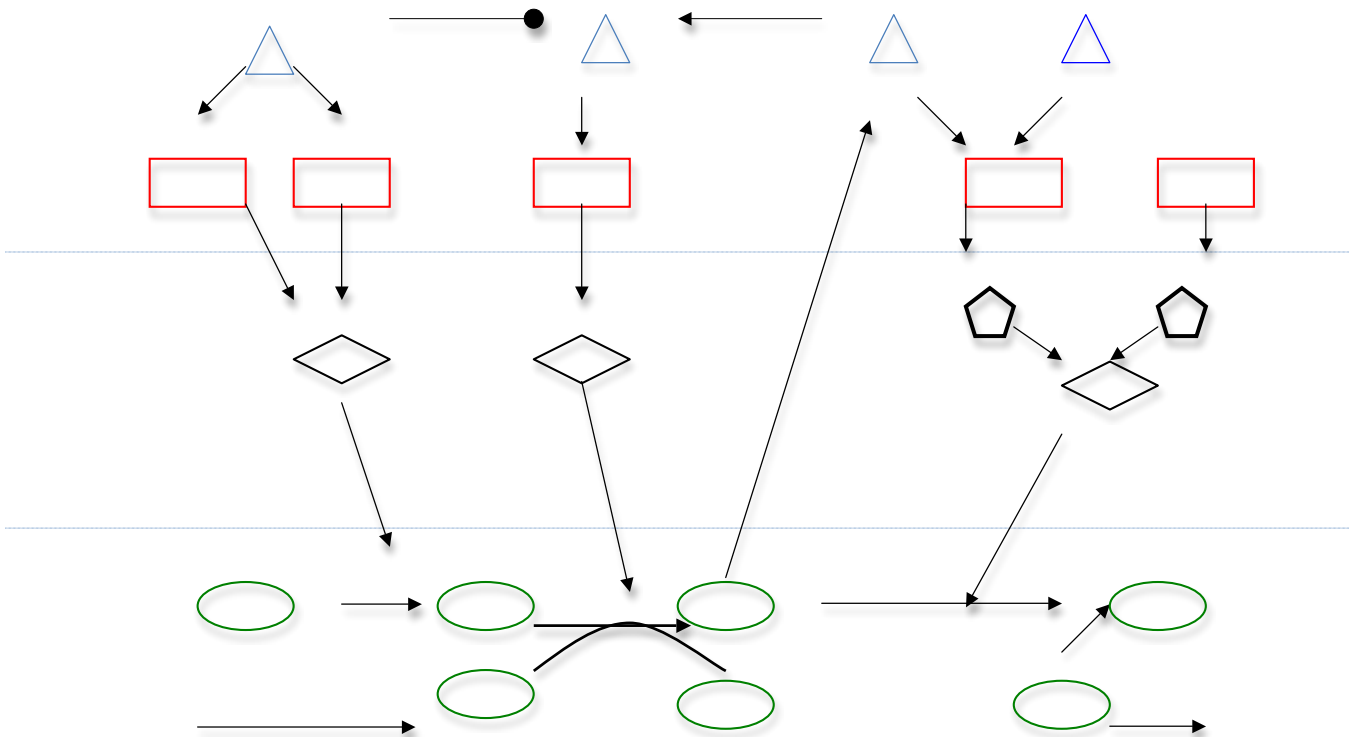


Evolutionary Models for Combined Regulation-Metabolism Graphs

Jotun Hein and Gail Preston 7.12.09

Motivation. Evolutionary models have been made for protein interaction networks and for metabolic pathways. Such models have the advantage that a homologous set of networks can then be analyzed, which is a typical situation as information are often gathered in multiple species. Since metabolism and genes interact, it is fruitful to model these in together and such models have recently been published (Shlomi et al., 2007; Yeang and Vingron, 2006). In a single organism this allows the analysis of data including both metabolic flux and expression levels, such as knockout experiments. Making evolutionary models for integrated networks would have the advantages that analysis from multiple species can now be performed and if information was obtained in one organism, such models would provide a natural framework for how to transfer this knowledge to a target specie.

Technical background - We have two classes of networks and for each we need to integrate two components: Dynamics and evolution. Below is illustrated the models by Shlomi et al. (2007) – all we need is to add evolution to the integrated model.



The integrated model as described in Shlomi et al. (2007). Different representations of the two classes of networks can be chosen, but for simplicity we can assume that the regulatory is boolean and the metabolic system is described by mass action equations or just by a stoichiometric matrix/reaction graph. Blue triangles are TF (transcription factors) that can interact with genes (red rectangles). The genes make proteins (black diamonds), that might go together, that works as enzymes activating reactions in the the metabolism. Metabolites (green ovals) might activate or repress TFs.

i.a. dynamic models for gene regulation – a long series of models exists. The main decision here is whether we want to have a discrete (most often boolean – on/off) states for the individual genes.

i.b dynamic models for metabolism – again a series of models exists and levels of description. The simplest would be a pure stoichiometry graph that describes the possible set of reactions and fluxes. Reactions could then be turned off and on, dependent the state of the gene for the corresponding enzyme. More complicated than this would be to give each reaction a mass action kinetics or do have a specific flux, whose components are turned on/off dependent on the status of the genes between the corresponding enzymes.

ii.a. evolutionary models for gene regulatory networks – a variety of models have been considered from very simple to very realistic (Lynch, 2007; Quale and Bullock, 2006; Jenkins and Stekel, 2009).

ii.b. evolutionary models for metabolisms – adding/deleting reactions according to some death-birth process with some level of dependency seems natural and has been analyzed by for instance Mithani et al. (2009).

iii. Making a general integrated model including both dynamics and evolution is in principle simple: Again we want to lose/gain edges, but in an integrated model nodes will be additionally labeled into genes, TFs, metabolites and enzymes.

Project. The full analysis of real data relating to metabolism and regulation for two organisms would imply many components and in this project will focus on a single component, namely the formulation of an evolutionary model for integrated regulation-metabolism networks and simulations to investigate properties of the model. Such a model will

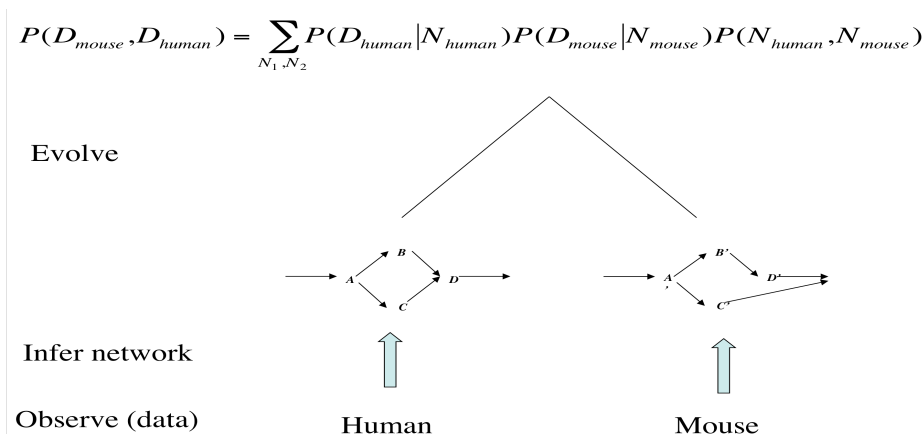
Work Plan.

- Week 1-2: Read key papers from the literature list and make preliminary contents of the report.
- Week 3-5: Implement the basic representation of a RM network and the evolutionary operations that can act on it. Simulate the long term behaviour of this model under different parameters. It is of interest to know which parameters gives networks similar to real ones.
- Week 6-8: Using MCMC investigate evolutionary paths from N1 to N2, ie condition on the end point as well.
- Week 9-10: Finish report and discuss natural extensions of the work.

References:

- Babu et al. (2004) Structure and evolution of transcriptional regulatory networks. *Curr. Op. Struc. Biol.*, **14**, 283-291753.
- Covert et al. (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 429.92-97
- Covert et al. (2008) Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli *Bioinformatics* 24.28. 2044-50
- Jenkins and Stekel (2009) A New Model for Investigating the Evolution of Transcription Control Networks *Artificial Life* 15: 259–291
- Lee et al. (2008) Dynamic Analysis of Integrated Signaling, Metabolic, and Regulatory Networks. *PLOS Comput Biol.* 4.5.1-20
- Lynch, M (2007) The evolution of genetic networks by non-adaptive processes *Nature Review Genetics* 8.803-813
- Mithani et al. (2009) A stochastic model for the evolution of metabolic networks with neighbor dependence. *Bioinformatics (advance access)*
- Quayle, A. and Bullock, S. (2006) Modelling the evolution of genetic regulatory networks. *Journal of Theoretical Biology*, 238 (4), pp. 737-750
- Reil (1999) Dynamics of Gene Expression in an Artificial Genome - Implications for Biological and Artificial Ontogeny. *ECAL* 457-466
- Shlomi et al. (2007) A genome scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol* 3:101
- Yeang and Vingron (2006) "A joint model of regulatory and metabolic networks" *BMC Bioinformatics*. 7.332-33.

Scaling to a Dphil. This project is ideal as a pre-study for a Dphil. The full problems for a non-composite networks can be exemplified for human and mouse as data (expression data/metabolic dynamics) has been observed from the same (homologous) system in human and mouse. Key to the problem is to calculate the probability of the two networks, P. Inference of the unobservable networks can be done if some networks contribute more to the probability of the data. P can be calculated by summing over all possible pairs of networks and then also over all possible histories creating the pair of networks. Thus relative to the pilot project (that focussed on the novel component of evolving a network), there are two new components: real data network inference [P(D|N)]. However,



The main additional components in a full projects are:

- Real data. Biological networks suitable for analysis using this approach will be identified in consultation with Gail Preston, whose research addresses the function, evolution and regulation of metabolic networks in pathogenic microorganisms.
- Simultaneous Network inference. There is a network inference problem for both gene regulatory networks and metabolic pathways, while the former is by far the most studied. For a review of the former see (Markowitz and Spang, 2007) and for a few papers on the latter, please see (Arita 200a,b; Boyer et al.(2003))

References:

- Arita (2000) Metabolic reconstruction using shortest path Simulation Practice and Theory 8.1-2.109-12
- Arita (2000) Graph Modelling of Metabolism JS Art Intel. 15.4.??
- Boyer (2003) Ab Initia reconstruction of metabolic pathways *Bioinformatics* 19.2.26-34
- Markowitz and Spang (2007) Inferring cellular networks – a review *BMC Bioinf* 2007, 8(Suppl 6):S5
- Tucker et al. (2007) Estimating parameters for generalized mass action models using constraint propagation *Math Biosci* 208.2.607 620