Final report

Genomic prediction of essential genes: an *in silico* approach

Hungarian Research Fund (OTKA: 49800)

Csaba Pal, PhD

Perhaps one of the most striking discoveries of modern molecular genetics was the extent by which organisms appear to tolerate mutations or even complete loss of their genes. Systematic single gene deletion studies have revealed that less than 20% of the protein coding genes in yeast (*Saccharomyces cerevisiae*) appear to be indispensable or essential for viability under lab conditions. Our research concentrates largely on yeast (*S. cerevisiae*), and we seek to understand the physiological and evolutionary mechanisms behind this pattern. The following questions sum up our research:

1) Are these seemingly non-essential genes redundant or do they have important contribution under special environmental conditions not yet tested in the laboratory?
2) What's the contribution of gene duplicates and alternative pathways to robustness of metabolic networks against harmful mutations (figure 1)? Is it likely that robustness (e.g. in the form of compensatory genetic interactions) is a directly selected trait?
3) Do non-essential genes evolve at especially high rates across species, as expected if they were under relaxed selection pressure?

To address these issues, we borrowed techniques from bioinformatics, comparative genomics and experimentally tested predictions of genome-scale systems biology models:

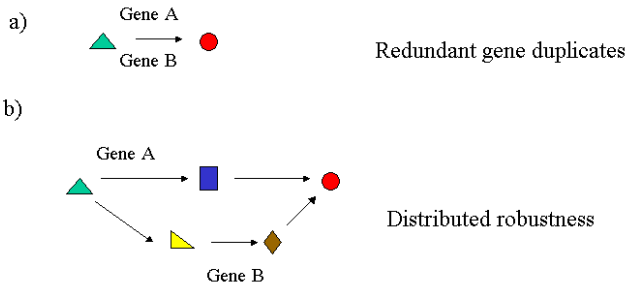
- **Comparative genomics of gene dispensability**. Advances in systems biology and genomics have facilitated a move from studying individual proteins to characterizing global cellular factors. Our systematic surveys indicate that protein evolution is not determined exclusively by selection on protein importance and function, but is also affected by the genomic position of the encoding genes, their expression patterns, their position in biological networks and possibly their robustness to mistranslation. We have identified several cellular and genomic

features that enable reliable characterization of essential[1-4] and dosage sensitive genes[5]. It appears that gene duplications[4], alternative metabolic pathways[4], gene expression level[3,5] and genomic position[1] all have some effect on gene dispensability. After controlling for such confounding factors, the association between gene essentiality and rate of protein evolution is fairly weak[2,3]. These results reinforce the view that the seemingly non-essential genes are under intense purifying selection and have important functional contribution to growth, possibly under special environmental conditions.

- **Metabolic network analyses of gene dispensability**. The high fraction of non-essential genes in yeast and other organisms under laboratory conditions raises the question of what the mechanistic basis for dispensability is, and whether it is the result of selection for buffering or an incidental side product. We analysed these issues using an *in silico* systems biology model of the yeast metabolic network[4]. The model correctly predicts the impact of gene deletions in 88% of the genes studied and *in vivo* metabolic fluxes. Our model predicts that many of the seemingly dispensable genes are important, but under conditions not yet examined in the laboratory. There is ample experimental evidence supporting this view (figure 2). We also found that the presence of gene duplicates can be explained by their role of increasing metabolic fluxes rather than by selection for compensation[4]. The systems biology models combined with phylogenetic analysis allowed us to study several other key issues in evolutionary biology, such as the nature and evolution of minimal genomes[6] and contribution of horizontal gene transfer to network evolution[7].

- **Computational and experimental analyses of genetic interactions**. The systems biology model described above paves the way for gaining novel insights into the nature of genetic interactions[8]. The current emphasis is on 'synthetic lethal' gene pairs, in which case members are mutually capable of compensating null mutations in each other (figure 1). Systematic mapping of these interactions not only provides a profound insight into gene functions[9], but it also contributes to our better understanding of the possible mechanistic bases of robustness against harmful mutations[8]. We performed systems-level flux balance analysis of the yeast (*S. cerevisiae*) metabolic network to identify genetic interactions and then tested the model's predictions with *in vivo* gene-deletion studies. We found that the majority of synthetic genetic interactions are restricted to certain environmental conditions, partly because of the lack of compensation under some (but not all) nutrient conditions. Experimental analyses that used multiple gene deletion strains not only confirmed predictions of the model but also showed that investigation of false predictions may both improve functional annotation within the model and also lead to the discovery of higher-order genetic interactions (e.g. by analysing the effects of deleting three or more genes). More generally, by simultaneously studying the effects of mutations and environmental changes, our work has provided a unified framework for understanding mutational robustness and environmental adaptation[8].

Figures



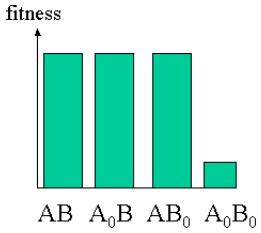


Figure 1. A model for two potential causes of robustness in metabolic networks. (A) The first of the two causes is redundancy of a system's parts: a gene may be dispensable if the genome contains other copies of the same gene with overlapping functions. In our example, both gene A and gene B encode enzymes that can catalyze the same reaction independently of each other, producing a key metabolite (red circle). (B) Alternatively, robustness may be distributed across pathways: the same key metabolite (red circle) can be produced through parallel metabolic routes. (C) Synthetic lethal gene interactions. In both cases, as production of the key metabolite is little affected, single-gene deletant genotypes (A_0B or AB_0) have relatively high fitness, comparable to that of wild type (AB). In contrast, deleting both copies of a duplicated gene or genes sitting on alternative pathways (A_0B_0) is expected to have a drastic effect on fitness.

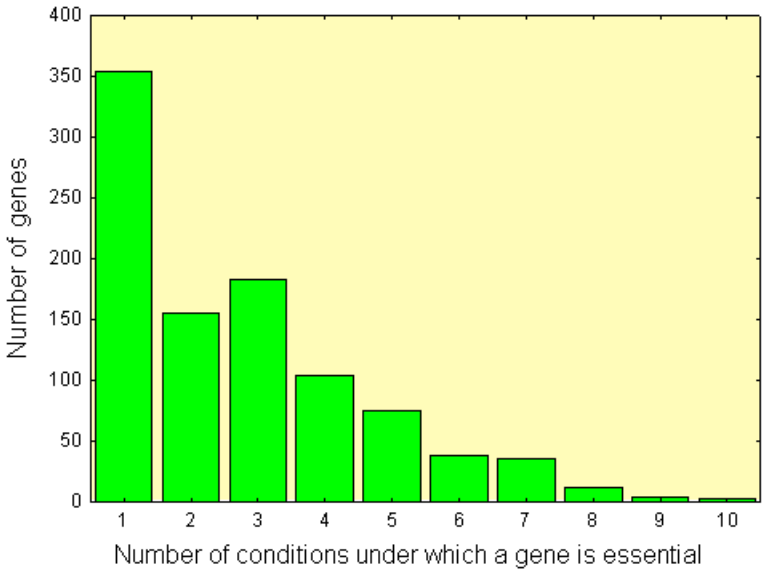


Figure 2 Distribution of conditionally essential genes. Many seemingly nonessential genes in yeast (*S. cerevisiae*) make a contribution to growth under specific conditions. Of 4823 genes not essential for growth on nutrient-rich medium (YPD), 963 exhibited lethality or a strong growth defect under some other conditions. Moreover, most of these conditionally essential genes make a contribution to growth under only one or few of the 31 conditions examined. Gene deletions showing conditional growth phenotypes were compiled from published large-scale screens. Adapted from Harrison et al. 2007[8]

References:

1. Pal, C. & Hurst, L.D. Evidence for co-evolution of gene order and recombination rate. *Nat Genet* **33**, 392-5 (2003).
2. Pal, C., Papp, B. & Hurst, L.D. Genomic function: Rate of evolution and gene dispensability. *Nature* **421**, 496-7; discussion 497-8 (2003).
3. Pal, C., Papp, B. & Lercher, M.J. An integrated view of protein evolution. *Nat Rev Genet* **7**, 337-48 (2006).
4. Papp, B., Pal, C. & Hurst, L.D. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429**, 661-4 (2004).
5. Papp, B., Pal, C. & Hurst, L.D. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**, 194-7 (2003).
6. Pal, C. et al. Chance and necessity in the evolution of minimal metabolic networks. *Nature* **440**, 667-70 (2006).
7. Pal, C., Papp, B. & Lercher, M.J. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet* **37**, 1372-5 (2005).
8. Harrison, R., Papp, B., Pal, C., Oliver, S. & Delneri, D. Plasticity of genetic interactions in metabolic networks of yeast. *Proc. Natl. Acad. Sci. U.S.A.* **104** 2307-2312 (2007).
9. Herrgard, M.J. & Palsson, B.O. Untangling the web of functional and physical interactions in yeast. *J Biol* **4**, 5 (2005).