16 Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U. S. A.* 85, 2444–2448
17 Mewes, H.W. *et al.* (2002) MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* 30, 31–34
18 Ho, Y. *et al.* (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415, 180–183
19 Gavin, A.C. *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141–147

Sarah A. Teichmann*
M. Madan Babu
MRC Laboratory of Molecular Biology, Hills Road, Cambridge, UK CB2 2QH.
*e-mail: sat@mrc-lmb.cam.ac.uk

Conservation of gene co-regulation in prokaryotes and eukaryotes

Response from Berend Snel, Peer Bork and Martijn A. Huynen

Currently we have a wealth of data on various kinds of functional relations between proteins obtained from comparative genome analysis and high-throughput experiments (reviewed in [1]). Teichmann and Babu discuss, on account of our recent work, protein interaction networks obtained from comparative genome analysis. They take the opportunity to discuss the conservation of co-regulation and its correlation with functional relations among pairs of proteins.

We would like to raise some issues in detecting the conservation (or absence thereof) of co-regulation using gene order; how we think the variations in the cellular network in various species can be studied; and how to determine and interpret the higher order structure in networks of functional relations.

At first glance, the amount of conserved gene order (on average 15–20% among divergent species; data obtained using [2]) suggests that co-regulation is poorly conserved in prokaryotes. However, the fact two genes are not neighbours does not mean that they are not co-regulated. They can still be part of the same regulon. McGuire and co-workers [3] showed that the conservation of two genes in one operon can be used as a powerful indicator of shared regulatory elements in front of those genes in species where they are not in the same operon. In general, it is hard to prove the absence of co-regulation and thus to study its evolution. Even when expression data do not indicate co-regulation, it remains unclear how functional relations evolve because for proteins that have been shown to physically interact, coexpression cannot always be detected [1].

Nevertheless, the cellular network obviously differs between species. How can we detect the taxon-specific relations that underlie these differences when the conservation of co-regulation is such an important signal in detecting functional relations? In our work on gene order, we identify genes (so called linkers) that repeatedly occur in completely non-intersecting operon settings [4]. Such mutually exclusive associations can indicate that member proteins of the orthologous group from different species have distinct functional relationships in their respective species. To use Teichmann's example of the variation of the citric acid cycle among published genomes, we observed that the operon organization in *Methanobacterium thermoautotrophicum*, in which pyruvate ferredoxin oxidoreductase and fumarase are linked, supports the suggestion that the incomplete citric acid cycle in this species runs in the reductive direction [5]. With the recent sequencing of *Methanopyrus kandleri* [6], this operon organization can be observed to be conserved: the genomic indication of the alternative mode of interaction between these proteins is in itself conserved and therefore significant. We believe that, as more genomes become available, such 'conservation of variation' will be an important tool to study the evolution of functional interactions by comparative genome analysis.

To determine and interpret the emerging networks of functional relations, our work presented a study of a protein interaction network predicted from conserved gene neighborhood [4]. This network is a small-world scale-free network. Interestingly, its high degree of local clustering suggests that functional modules exist. We succeed in delineating such sets of interacting proteins from the network. In parallel, similar interacting sets have been identified by Rogozin and co-workers [7] and van Nimwegen and co-workers [8]. Comparative genome analysis thus allows the identification of a level of functional interaction intermediate between that of pairwise interactions and that of the complete genome, which, we think, constitutes what a functional module is, from the perspective of a cell or organism.

References

1 von Mering, C. *et al.* (2002). Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417, 399–403
2 Korbel, J.O. *et al.* (2002). SHOT: a web server for the construction of genome phylogenies. *Trends Genet.* 18, 158–162
3 McGuire, A.M. and Church, G.M. (2000) Predicting regulons and their cis-regulatory motifs by comparative genomics. *Nucleic Acids Res.* 28, 4523–30
4 Snel, B. *et al.* (2002). The identification of functional modules from the genomic association of genes. *Proc. Natl. Acad. Sci. U. S. A.* 99, 5890–5895
5 Huynen, M.A. *et al.* (1999) Variation and evolution of the citric acid cycle: a genomic approach. *Trends Microbiol.* 7, 281–291
6 Slesarev, A.I. *et al.* (2002) The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4644–4649
7 Rogozin, I.B. *et al.* (2002) Connected gene neighborhoods in prokaryotic genomes. *Nucleic Acids Res* 30, 2212–2223
8 van Nimwegen, E. *et al.* (2002) Probabilistic clustering of sequences: Inferring new bacterial regulons by comparative genomics. *Proc. Natl. Acad. Sci. U. S. A.* 99, 7323–7328

Berend Snel
Martijn A. Huynen
Nijmegen Center for Molecular Life Sciences, p/a Center for Molecular and Biomolecular Informatics, Toernooiveld 1, 6525 ED Nijmegen, Netherlands.
e-mail: b.snel@cmbi.kun.nl

Peer Bork
European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

Published online: 08 August 2002

  PII: S0167-7799(02)02040-1