BMC Evolutionary Biology

Evolutionary rate depends on number of protein-protein interactions independently of gene expression level: Response

Jesse D Bloom (bloom@caltech.edu)
Christoph Adami (adami@caltech.edu)

Evolutionary rate depends on number of protein-protein interactions independently of gene expression level: Response

Jesse D. Bloom[1*] and Christoph Adami[2, 3]

1 Department of Chemistry and Digital Life Laboratory, California Institute of Technology, Pasadena, CA 91125 USA
2 Digital Life Laboratory, California Institute of Technology, Pasadena, CA 91125 USA
3 Keck Graduate Institute, 535 Watson Drive, Claremont, CA 91711 USA
*Corresponding author: bloom@caltech.edu

Evolving proteins are under selection for the ability to perform precise biochemical functions at minimal metabolic cost in a complex cellular environment. One way to investigate the different selective pressures is to examine what factors influence the rate of protein sequence evolution. In a recent study published in 2002, Fraser et al suggested that proteins that participate in more protein-protein interactions are under greater evolutionary constraint [1]. The basis for this claim was a weak but still statistically significant correlation between a protein's rate of sequence evolution and its number of interaction partners as measured by various studies of protein-protein interactions in yeast. However, subsequent studies found this correlation to be highly dependent on the particular choice of protein -protein interactions data set [2, 3].

We resolved this controversy by demonstrating that the correlation between evolutionary rate and the number of interaction partners is linked to a bias towards counting more interactions for abundant proteins[4]. Abundant proteins evolve more slowly [5] and some studies are biased towards finding more protein- protein interactions for abundant proteins [6]. Only those data sets that are biased towards finding more interactions for abundant proteins suggest a correlation between evolutionary rate and the number of interaction partners (Figure 1). Some of our findings have subsequently been echoed by others [7].

Now, Fraser and Hirsh again argue for a meaningful connection between the number of interaction partners and evolutionary rate[8]. We still cannot agree with their analysis. First, we note that the single data set they have re- analyzed is precisely the one which we identified as being the most biased (Figure 1). Their choice to only count interactions for the untagged proteins in mass-spectrometry studies not only fails to account for effects due to the choice of which protein to overexpress (as an interaction is inherently at least pairwise), but in fact increases the net bias in this data set [4]. Fraser and Hirsh also use partial correlation statistics to argue that abundance does not account for all of the correlation. While it is true that some of the data sets still show a statistically significant partial correlation (as we noted in [4]), statistical tests are only as good as the quality of the data to which they are applied and are not a substitute for carefully inspecting the effects of biases in individual data sets. Figure 1 shows a direct linear relationship between the apparent correlation and the bias of the data set, and data sets with no bias show no correlation.

Fraser and Hirsh comment that some of our previous analysis was based on expression levels measured in an aneuploid strain of yeast. This is true but irrelevant, since we observe identical trends if we quantify abundance using codon adaptation index [4] or expression levels from the microarray study preferred by Fraser and Hirsh (data not shown).

We readily acknowledge the possibility that there is a real connection between the number of interaction partners and evolutionary rate hidden in all the noise and biases. However, we feel that the appropriate null hypothesis is that there is no correlation, and we do not believe this null hypothesis has been convincingly disproven.

Hirsh and Fraser's original claim [1] rested on the idea that evolutionary constraints due to protein-protein interactions could be represented by a protein's total number of unique interaction partners. We suggest that if interactions do impose constraints on sequence evolution, they are likely to depend on more subtle factors such as the fraction of a protein's residues directly involved in an intermolecular contact or the total number of monomers present in a macromolecular complex. In fact, one study has investigated the effect of an interaction's type (transient or stable) although this analysis also failed to control for protein abundance [9].

The ultimate lesson of this controversy is that the complexities and interdependencies of protein evolutionary constraints must be properly controlled for. Many factors have now been investigated for their effects on protein evolutionary rate, and one of the interesting conclusions is that protein abundance has a far greater effect [5] than other apparently more intuitively appealing factors such as protein dispensability [10, 11, 12] or the number of interaction partners. The best studies have acknowledged this fact by carefully controlling for protein abundance (see for example [13]), and we suggest that this should become standard procedure in the future.

References

1. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, and Feldman MW. **Evolutionary rate in the protein interaction network**. *Science*, 296:750-752, 2002.
2. Jordan IK, Wolf YI, and Koonin EV. **No simple dependence between protein evolution rate and the number of protein- protein interactions: only the most prolific interactors tend to evolve slowly**. *BMC Evol Biol*, 3:1, 2003.
3. Fraser HB, Wall DP, and Hirsh AE. **A simple dependence between protein evolution rate and the number of protein- protein interactions**. *BMC Evol Biol*, 3:11, 2003.
4. Bloom JD and Adami C. **Apparent dependence of protein evolutionary rate on the number of interactions is linked to biases in protein-protein interactions data sets.** *BMC Evol Biol*, 3:21, 2003.
5. Pal C, Papp B, and Hurst LD. **Highly expressed genes in yeast evolve more slowly**. *Genetics*, 158:927-931, 2001.
6. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, and Bork P. **Comparative assessment of large- scale data sets of protein-protein interactions.** *Nature*, 417:399-403, 2002.
7. Hahn MW, Conant GC, and Wagner A. **Molecular evolution in large genetic networks: does connectivity equal constraint?** J *Mol Evol,* 58:203-211, 2004.
8. Teichmann SA. **The constraints protein- protein interactions place on sequence divergence.** *J Mol Biol*, 324:399-407, 2002. 2
9. Fraser HB and Hirsh AE. **Evolutionary rate depends on number of protein-protein interactions independently of gene expression level** *BMC Evol Biol 4:?,* 2004
10. Jordan IK, Rogozin IB, Wolf YI, and Koonin EV. **Essential genes are more evolutionarily conserved than are nonessential genes in bacteria.** *Genome Res,* 12:962-968, 2002.

11. Pal C, Papp B, and Hurst LD. **Rate of evolution and gene dispensability.** *Nature*, 421:496-498, 2003.
12. Yang J, Gu L, and Li WH . **Rate of protein evolution versus fitness effect of gene deletion.** *Mol Biol Evol*, 20:772- 774, 2003.
13. Pal C, Papp B, and Hurst LD. **Does recombination rate affect the efficiency of purifying selection? the yeast genome provides a partial answer.** *Mol Biol Evol*, 18:2323-2326, 2001. 3
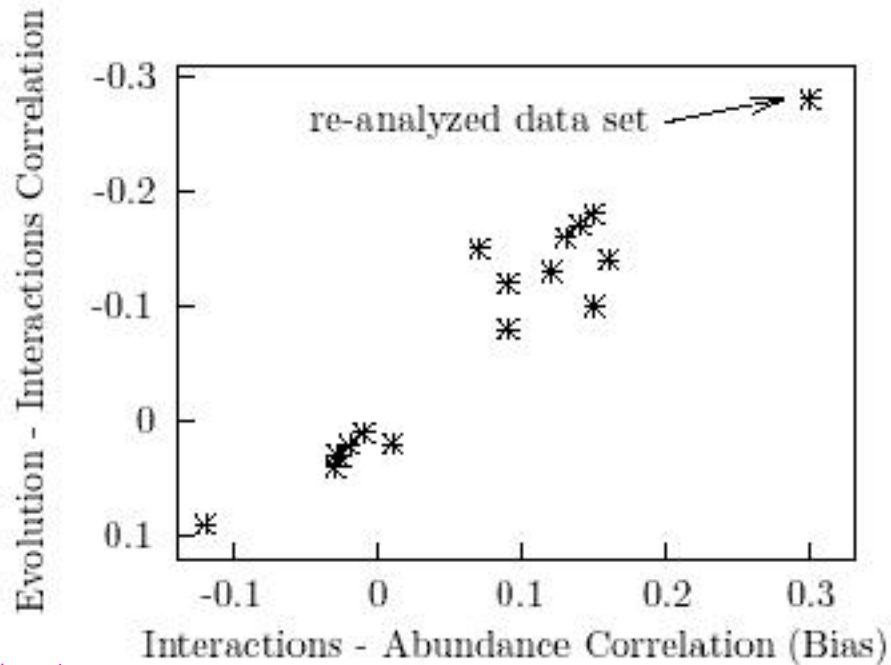
Figure 1