The likelihood that two proteins interact might depend on the proteins' age

Pedro Beltrao

(pedrobeltrao@gmail.com)

Abstract

It has been previously shown [1] that *S. cerevisiae* proteins preferentially interact with proteins of the same estimated likely time of origin. To study this observation further, the protein interaction networks of *S. cerevisiae* and *H. sapiens* were analyzed taking into account an estimate for the age of the proteins in these species. These estimates were obtained by studying the presence and absence of putative orthologs in other eukaryotic species. In this work preliminary results are described that point to a dependence of the likelihood of protein interaction on the proteins' age. The probability of two proteins interactions was found to be linearly dependent on the time the proteins have co-existed in the species.

Methods and Results

Protein-protein interactions for *S. cerevisiae* were obtained from BIND, excluding any interactions derived from protein complexes. Protein-protein interactions for *H. sapiens* were obtained from the Human Protein Reference database[2] and from two high-throughput studies[3-4] excluding any interactions derived from protein complexes. I considered only proteins that were represented in the interactomes (i.e. with one or more interactions).

In order to create groups of *S. cerevisiae* proteins with different average age I used the reciprocal best blast hit method to determine the most likely ortholog in eleven other yeast species (see figure 1 for species names).The same was done for *H. sapiens* proteins using eleven other eukaryotic species (see figure 2 for species names). *S. cerevisiea* and *H. sapiens* proteins with putative orthologs, in all species tested, were considered to be ancestral proteins and were grouped into group A. To obtain groups of proteins with decreasing average age of origin, proteins were selected according to the absence of identifiable orthologs in other species (see figure 1 and 2). It is important to note that these groups of decreasing average protein age are overlapping. Group F is contained in E, both are contained in D and so forth. Non overlapping groups of proteins with decreasing time of origin could have been selected but the lower numbers obtained might in a later stage make statistical analysis more difficult.

The phylogenetic trees in figure 1 and 2 (obtained with MEGA 3.1) are neighborhood joining trees obtained by concatenating 10 proteins from the ancestral group A (from both species). They are mostly used to have a graphical representation of the species divergence. It is important to note that, in figure 2, the placement of *C. familiaris* does not correspond with other published phylogenetic trees. It might be due to the proteins selected for the tree construction. I should consider using different combinations of ancestral proteins to check the robustness of the tree.

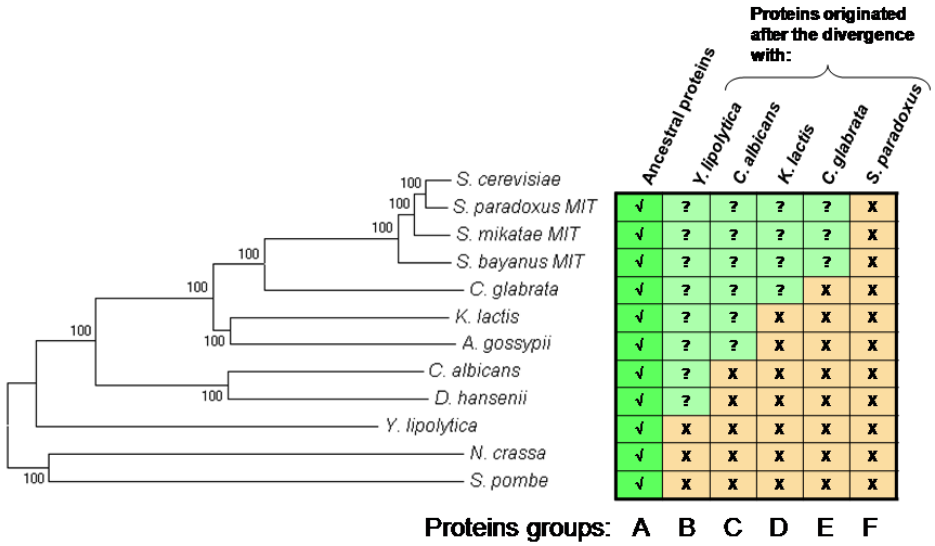

Figure 1 – Selection of *S. cerevisiae* protein groups with decreasing average age of origin estimated by patters of presence and absence of identifiable orthologs in other yeast species

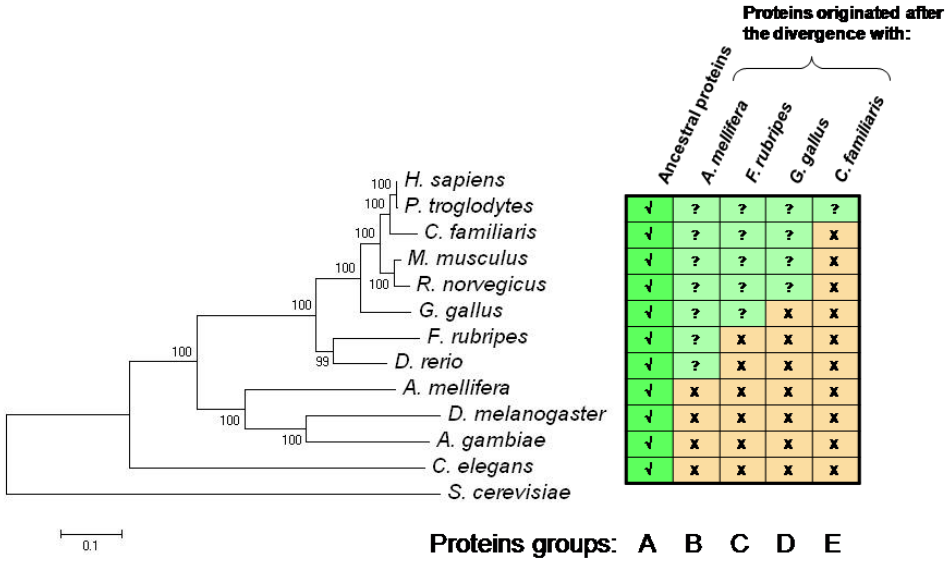

Figure 2 – Selection of *H. sapiens* protein groups with decreasing average age of origin estimated by patters of presence and absence of identifiable orthologs in other species

To determine the effect of protein age on the likelihood of interaction with ancestral proteins I counted the number of interactions between proteins in group A and the other groups of proteins for *S. cerevisiae* (see table 1) and for *H. sapiens* (see table 2)

Table 1 – Likelihood for protein-protein interactions correlates with the age of the interacting partner for *S. cerevisiae* proteins. Group A contains the predicted most ancient proteins. Groups B to F contain proteins with predicted decreasing average age.

Reference group	Proteins in group A	Proteins in reference group	Prot. interactions between groups	Likelihood for prot. interaction
B	848	1783	1784	0.00118
C	848	1387	1233	0.001048
D	848	858	656	0.000902
E	848	741	536	0.000853
F	848	434	242	0.000658

Likelihood for protein interactions within group A = 0.003168

Table 2 – Likelihood for protein-protein interactions correlates with the age of the interacting partner for *H. sapiens* proteins. Group A contains the predicted most ancient proteins. Groups B to E contain proteins with predicted decreasing average age.

Reference group	Proteins in group A	Proteins in reference group	Prot. interactions between groups	Likelihood for prot. interactions
B	604	3057	1112	0.000602
C	604	1104	310	0.000465
D	604	746	182	0.000404
E	604	152	32	0.000349

Likelihood for protein interactions within group A = 0.002844

From the data it is possible to observe that protein-interactions within groups (within group A) are more likely than protein-interactions between groups. This is in agreement with the results from Qin *et al.*[1]. Also the likelihood for a protein to interact with an ancestral protein depends on the age of this protein. This simple analysis suggests that the younger the protein is the less likely it is to interact with an ancestral protein.

I redid the analysis for the human interactome, excluding yeast-two-hybrid interactions from the dataset. As it can be seen in table 3, the results are qualitatively the same. There is a small increase in the likelihood of interaction with the ancestral proteins for the youngest group (highlighted in red in table 2) that is likely due to lack of data.

Table 3 – The correlation between likelihood of protein interactions and average age of proteins is not due to yeast-two-hybrid interactions, for *H. sapiens* proteins. Group A contains the predicted most ancient proteins. Groups B to E contain proteins with predicted decreasing average age.

Reference group	Proteins in group A	Proteins in reference group	Prot. interactions between groups	Likelihood for prot. interactions
B	437	2281	772	0.000774
C	437	755	182	0.000552
D	437	486	106	0.000499
E	437	79	18	0.000521

Likelihood for protein interactions within group A = 0.004241

Caveats and possible continuations

The protein-protein interactions used here for *S. cereivisae* also contain the high-throughput studies and therefore the interactome used should be considered with caution. It would interesting to redo this analysis with a recent set of interactions compiled from the literature [5] but this will also introduce some bias into the interactome.

To validate these results it would be crucial to test the statistical significance of the observations. If the differences are significant it could be useful to try to correlate the likelihood of interactions with a quantitative measure like average protein identity.

One possible use of observation reported in this preliminary result, if it holds to further scrutiny, would be to use the likely time of origin of the proteins as information to include in protein-protein prediction algorithms.

References

[1] Qin H, Lu HH, Wu WB, Li WH. Evolution of the yeast protein interaction network. Proc Natl Acad Sci U S A. 2003 Oct 28;100(22):12820-4. Epub 2003 Oct 13

[2] Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK *et al.* Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Research. (2003). 13:2363-2371.

[3] Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A *et al.* Towards a proteome-scale map of the human protein-protein interaction network. Nature. 2005 Oct 20;437(7062):1173-8. Epub 2005 Sep 28.

[4] Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH *et al.* A human protein-protein interaction network: a resource for annotating the proteome.Cell. 2005 Sep 23;122(6):957-68.

[5] Reguly T, Breitkreutz A, Boucher L, et al. Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. J Biol. 2006 Jun 8;5(4):11 [Epub ahead of print]