Supplementary Material: Evolution and topology in the yeast protein interaction network

Stefan Wuchty[†]

[†]Department of Physics, University of Notre Dame, Notre Dame, IN 46556, U.S.A.

	r_D (P-value)	$r_{\langle ESC_k \rangle}$ (P-value)	r_{ER_k} (P-value)
H.sapiens	-0.0459 (0.0789)	-0.8446 (1.8×10^{-5})	0.9212 (6×10^{-6})
M.musculus	-0.0619 (0.0185)	-0.7400 (1.6×10^{-3})	0.9632 (1×10^{-8})
D.melanogaster	-0.0249 (0.3391)	-0.9239 (8.4×10^{-8})	0.8230 (9×10^{-4})
C.elegans	-0.0347 (0.2253)	-0.6464 (1.0×10^{-2})	0.8561 (3×10^{-4})
A.thaliana	-0.0609 (0.0234)	-0.7229 (1.5×10^{-3})	0.8979 (4×10^{-5})

Table 1. Pearsons correlation coefficients r and corresponding P-values of the individual distributions of the evolutionary distance D, mean sequence excess conservation $\langle ESC_k \rangle$ and orthologous excess retention ER_k as presented in Fig. 1a,b,c of the main paper.

	ρ_D (P-value)	$\rho_{\langle ESC_k \rangle}$ (P-value)	ρ_{ER_k} (P-value)
H.sapiens	-0.0584 (0.0258)	-0.7417 (0.0101)	1.0000 (0.0047)
M.musculus	-0.0777 (0.0031)	-0.7552 (0.0122)	1.0000 (0.0047)
D.melanogaster	-0.0241 (0.3559)	-0.7552 (0.0122)	0.9833 (0.0054)
C.elegans	-0.0599 (0.0363)	-0.5454 (0.0704)	0.9833 (0.0054)
A.thaliana	-0.0558 (0.0381)	-0.8022 (0.0055)	1.0000 (0.0047)

Table 2. Spearmans rank coefficients ρ and corresponding P-values of the individual distributions of the evolutionary distance D, mean sequence excess conservation $\langle ESC_k \rangle$ and orthologous excess retention ER_k as presented in Fig. 1a,b,c of the main paper.

	$D \sim$	$\langle ESC_k \rangle \sim$	$ER_k \sim$
H.sapiens	$-0.0079k$	$-0.226 \log k$	$0.831 \log k$
M.musculus	$-0.0091k$	$-0.180 \log k$	$0.930 \log k$
D.melanogaster	$-0.0039k$	$-0.129 \log k$	$0.741 \log k$
C.elegans	$-0.0048k$	$-0.117 \log k$	$0.857 \log k$
A.thaliana	$-0.0094k$	$-0.106 \log k$	$0.702 \log k$

Table 3. Regression statistics of the individual distributions of the evolutionary distance D, mean sequence excess conservation $\langle ESC_k \rangle$ and orthologous excess retention ER_k as presented in Fig. 1a,b,c of the main paper.

	$ER_k \sim$
essential	$+1.386 \log k$
non-essential	$-0.371 \log k$

Table 4. Regression statistics of the individual distributions of (non-)essential excess retention ER_k as presented in Fig. 2a of the main paper.

	$ER_k \sim$
H.sapiens	$1.992 \log k$
M.musculus	$2.201 \log k$
D.melanogaster	$2.146 \log k$
C.elegans	$2.256 \log k$
A.thaliana	$2.042 \log k$

Table 5. Regression statistics of the individual distributions of the excess retention ER_k of essential proteins having orthologs in *H. sapiens*, *M. musculus*, *D. melanogaster*, *C.elegans* and *A. thaliana* as presented in Fig. 2b of the main paper.

	r_{ER_k} (P-value)	ρ_{ER_k} (P-value)
essential	0.8681 (4.5×10^{-4})	0.9761 (9.8×10^{-3})
non-essential	-0.8481 (9.5×10^{-4})	0.8681 (4.5×10^{-3})

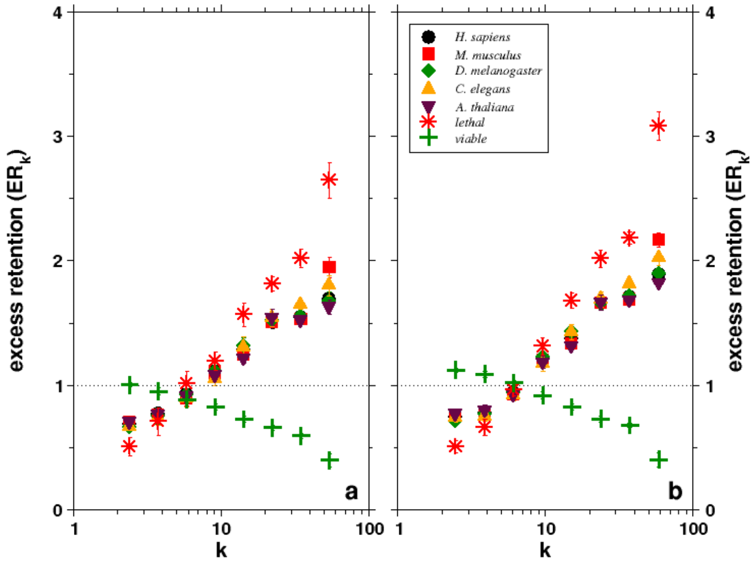
Table 6. Pearsons correlation coefficients r and Spearmans rank coefficients ρ with corresponding P-values of the individual distributions of the (non-)essential excess retention ER_k presented in Fig. 2a in the main paper.

	r_{ER_k} (P-value)	ρ_{ER_k} (P-value)
H.sapiens	0.8884 (1.8×10^{-4})	0.9524 (1.1×10^{-2})
M.musculus	0.9121 (4.6×10^{-5})	0.9524 (1.1×10^{-2})
D.melanogaster	0.9049 (7.3×10^{-5})	0.9762 (9.8×10^{-3})
C.elegans	0.9171 (3.3×10^{-5})	0.9762 (9.8×10^{-3})
A.thaliana	0.9212 (2.4×10^{-5})	1.0000 (8.2×10^{-3})

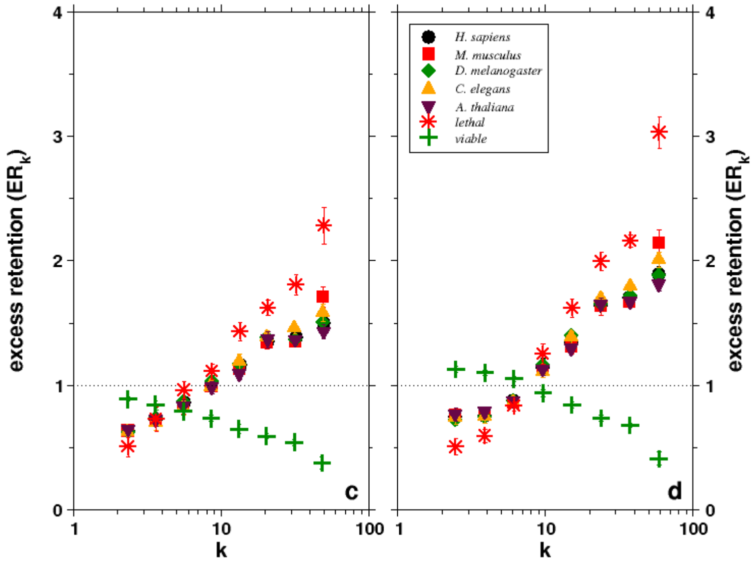
Table 7. Pearsons correlation coefficients r and Spearmans rank coefficients ρ with corresponding P-values of the individual distributions of the excess retention ER_k of essential proteins which have orthologs in *H. sapiens*, *M. musculus*, *D. melanogaster*, *C.elegans* and *A. thaliana* as presented in Figure 2b in the main paper.

	r_{ER_k} (P-value)	ρ_{ER_k} (P-value)
H.sapiens	-0.292 (0.426)	0.142 (0.705)
M.musculus	0.276 (0.464)	0.476 (0.208)
D.melanogaster	-0.154 (0.681)	0.071 (0.850)
C.elegans	0.704 (0.020)	0.667 (0.078)
A.thaliana	-0.209 (0.574)	0.167 (0.659)

Table 8. Pearsons correlation coefficients r and Spearmans rank coefficients ρ with corresponding P-values of the individual distributions of the excess retention ER_k of non-essential proteins which have orthologs in *H. sapiens*, *M. musculus*, *D. melanogaster*, *C.elegans* and *A. thaliana* as presented in Figure 2c in the main paper.

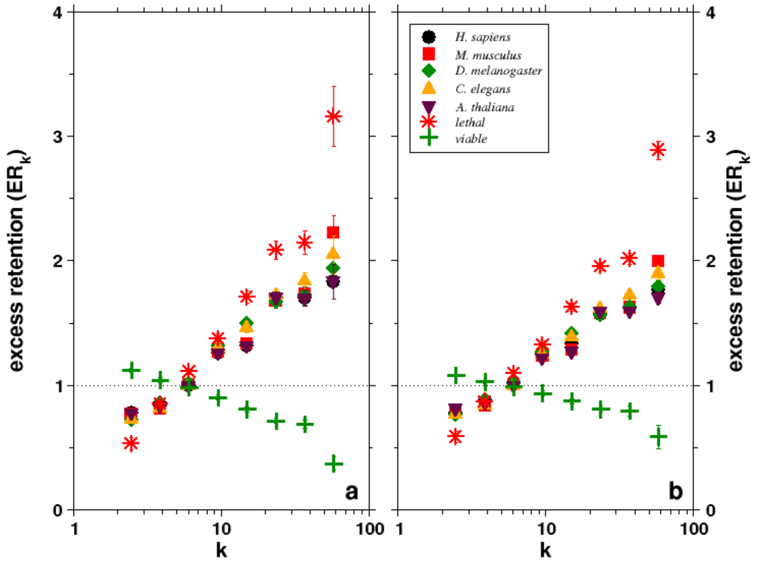


-10% interactions +10%

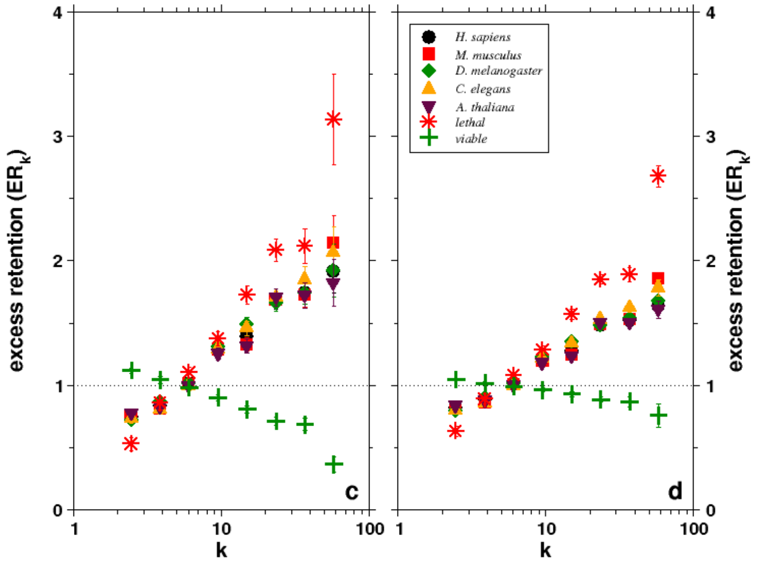


-20% interactions +20%

Figure 1. Dependence of the excess retention of (non-)essential and orthologous proteins from noisy and incomplete interaction data: (a),(c) We removed 10 and 20% of interactions between randomly selected protein pairs, mimicking false negatives. **(b),(d)** Simulating the effects of false positives, we added 10 and 20% of interactions which were previously absent in the network. In each case, 1,000 samples were generated. Excess retentions with respect to orthologs in *H. sapiens*, *M. musculus*, *D. melanogaster*, *C.elegans* and *A. thaliana*, essential and non-essential proteins were determined. Although the correlations between evolutionary excess retention and connectivity are shifted to lower/higher values, the basic trends remain unchanged.

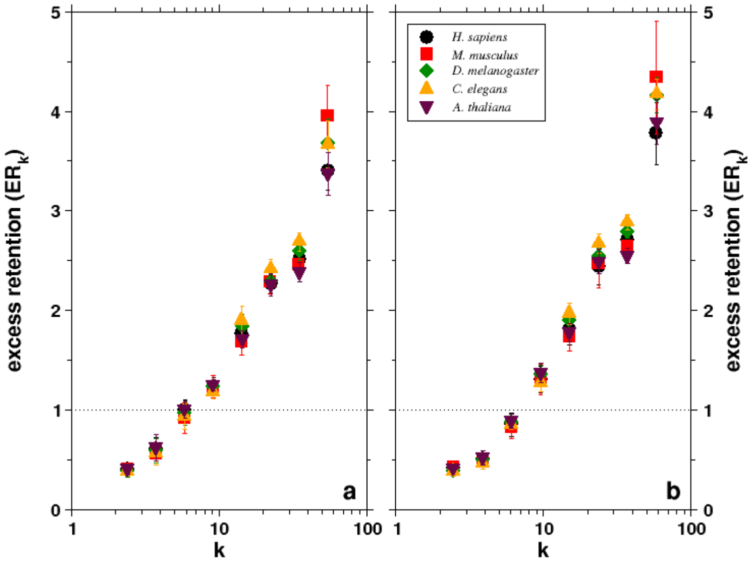


-10% orthologs +10%

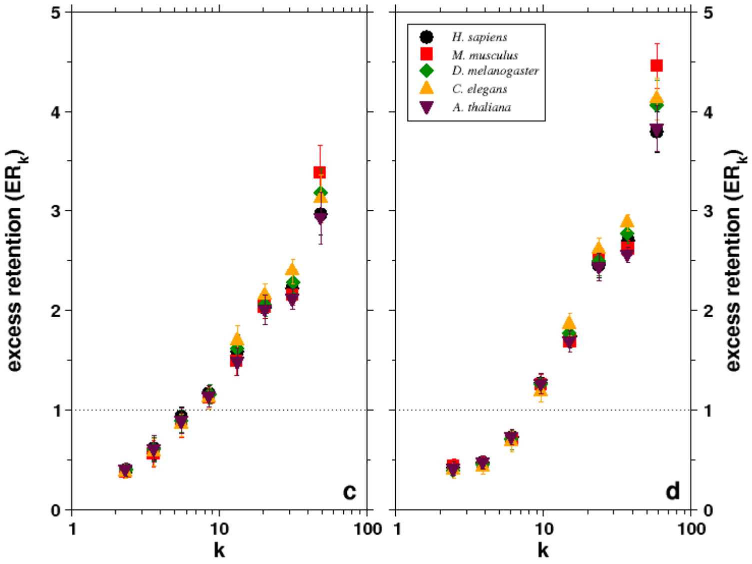


-20% orthologs +20%

Figure 2. Dependence of the excess retention from incomplete and inconsistent ortholog and (non-)essential protein data: (a),(c) In order to account for incompleteness, 10 and 20% of the proteins present in each set of orthologs, essential and non-essential proteins were eliminated. **(b),(d)** Accounting for the effects of inconsistent ortholog, essential and non-essential data we added 10 and 20% new, previously absent proteins to each set of orthologous, essential and non-essential proteins. In each case, we generated 1,000 samples and determined the excess retention with respect to orthologs in the reference organisms, essential and non-essential proteins as outlined in the paper. Similarly to Suppl. Fig.1, we do not find any qualitative changes.

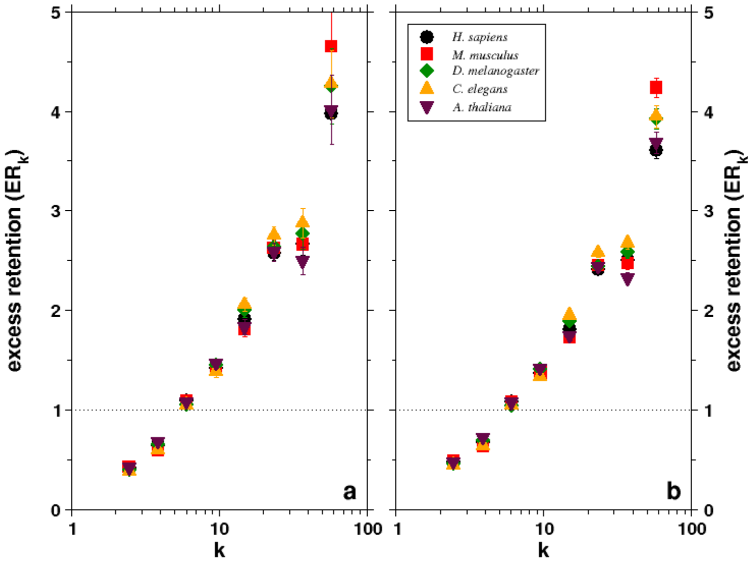


-10% interactions +10%

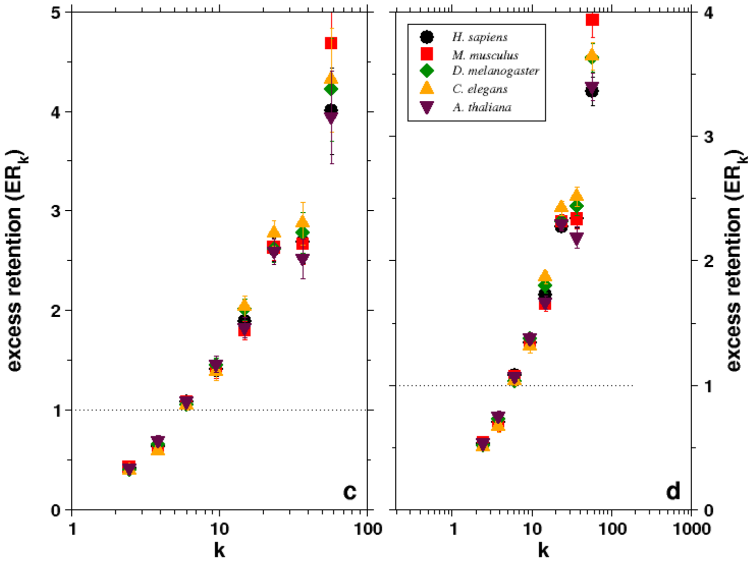


-20% interactions +20%

Figure 3. Dependence of the excess retention of essential proteins which have orthologs in *H. sapiens*, *M. musculus*, *D. melanogaster*, *C.elegans* **and** *A. thaliana* **from noisy and incomplete interaction data: (a),(c)** We removed 10 and 20% of interactions between randomly selected protein pairs, mimicking false negatives. **(b),(d)** Simulating the effects of false positives, we added 10 and 20% of interactions which were previously absent in the network. In each case, 1,000 samples were generated. Excess retentions with respect to essential proteins which have orthologs in the reference organisms were determined. Similarly to Suppl. Fig.1, we do not find any interference of data incompleteness with the presented trends.



-10% orthologs +10%



-20% orthologs +20%

Figure 4. Dependence of the excess retention from incomplete and inconsistent sets of essential data which have orthologs in *H. sapiens*, *M. musculus*, *D. melanogaster*, *C.elegans* **and** *A. thaliana***:** **(a),(c)** In order to account for incompleteness, 10 and 20% of the proteins present in each set were eliminated. **(b),(d)** Accounting for the effects of inconsistent data we added 10 and 20% new, previously absent proteins to each set. In each case, we generated 1,000 samples and determined the excess retentions of essential proteins which have orthologs in the reference organisms. Similarly to Suppl. Fig.1, we do not find any interference of data incompleteness with the presented trends.