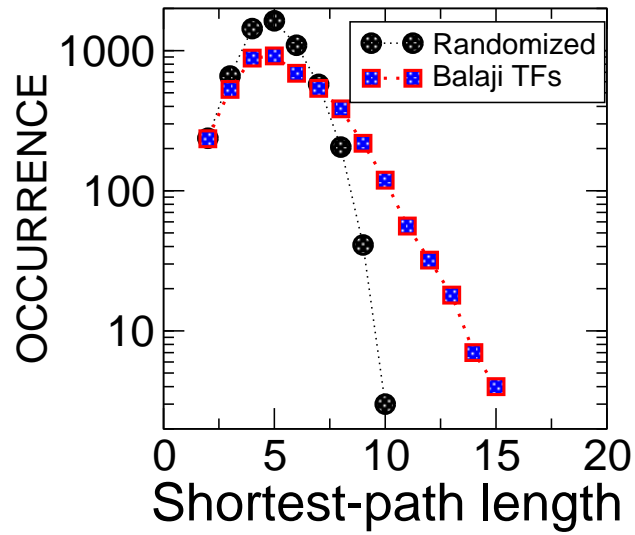


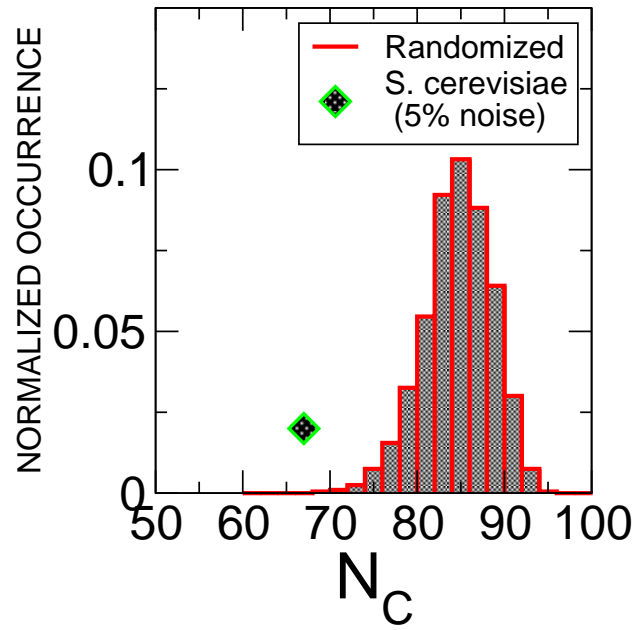
**“A comparative evolutionary study of transcription networks”**

**SUPPLEMENTARY MATERIAL**

A. L. Sellerio, B. Bassetti, H. Isambert, and M. Cosentino Lagomarsino



Supplementary Figure S1: Histogram of shortest-path lengths in the yeast TF network (Balaji data set), compared to a typical randomized instance. The graph plots the number of shortest paths of a given length. The two distributions have similar qualitative features, in particular showing exponential decay and the same location of the maximum. However, they differ significantly in their tails, where the empirical graph shows a higher abundance of longer shortest-paths.



Supplementary Figure S2: Role of noise in the data. Number of nodes  $N_C$  in the feedback core and histogram of degree conserving randomizations (see Figure 1 of the main text) for an instance of the Balaji network for yeast where 5% of the interactions were substituted by random ones to simulate noise in the data. The result of the analysis is tolerant to this operation.

<i>E. coli</i> (RDB 5.5)			
Subgraph type	# Empirical	Randomization	P-value
<b>(i)</b>	1	$7.877 \pm 3.605$	0.0017
<b>(ii)</b>	3	$0.412 \pm 0.640$	0.0087
<b>(iii)</b>	6	$4.089 \pm 2.639$	0.2497
<b>(iv)</b>	1	$0.1072 \pm 0.319$	0.1042
<b>(v)</b>	0	$0 \pm 0$	<i>n.a.</i>
<b>(vi)</b>	179	$81.55 \pm 15.24$	$< 10^{-4}$

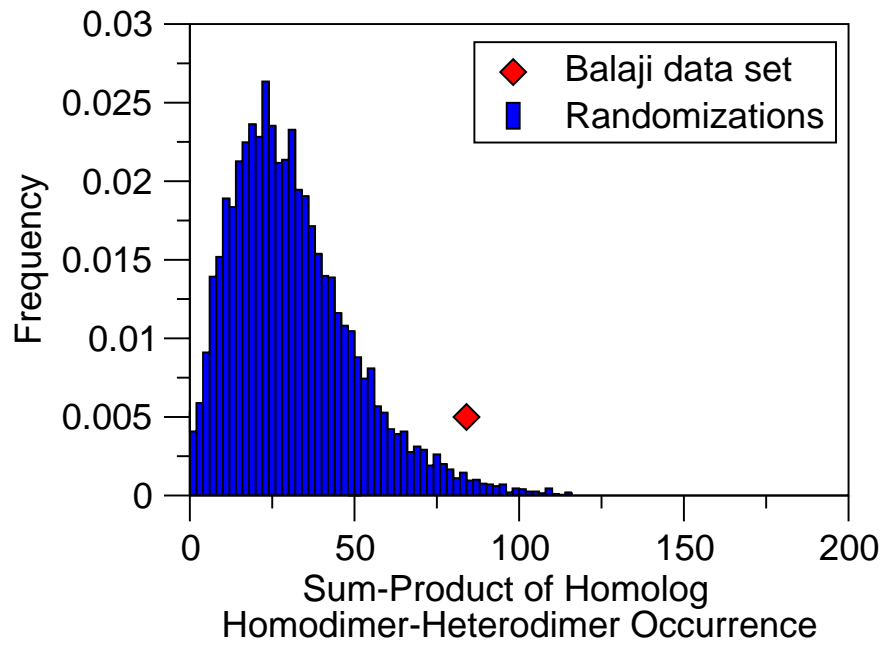
<i>B. subtilis</i> (DBTBS)			
Subgraph type	# Empirical	Randomization	P-value
<b>(i)</b>	1	$1.2472 \pm 1.0994$	0.2824
<b>(ii)</b>	2	$1.7739 \pm 1.8284$	0.4327
<b>(iii)</b>	1	$0.2370 \pm 0.4872$	0.2101
<b>(iv)</b>	1	$0.0755 \pm 0.2642$	0.0755
<b>(v)</b>	0	$0.0758 \pm 0.2646$	0.0758
<b>(vi)</b>	55	$29.237 \pm 10.672$	0.0242

<i>S. cerevisiae</i> (Balaji)			
Subgraph type	# Empirical	Randomization	P-value
<b>(i)</b>	7	$8.756 \pm 3.991$	0.3074
<b>(ii)</b>	6	$2.373 \pm 1.435$	0.0243
<b>(iii)</b>	0	$0.209 \pm 0.414$	$< 10^{-4}$
<b>(iv)</b>	0	$0 \pm 0$	<i>n.a.</i>
<b>(v)</b>	1	$0.199 \pm 0.399$	0.1995
<b>(vi)</b>	16	$17.49 \pm 6.51$	0.4308

Supplementary Table S1: Subgraphs compatible with AR duplications, containing crosstalks. The tables report the occurrence of different network subgraphs which can stem from duplications of nodes and edges and subsequent partial loss of regulatory interactions for the three transcription networks, compared to randomizations of the homology classes. These subgraphs are displayed in Fig. 3 of the main text.

Path	Data Set	Empirical	Randomization	Z-score
A - Long	<i>E. coli</i> (Shen-Orr)	844	687.37 ± 48.18	+3.25
	<i>E. coli</i> (RDB 5.5)	1171	420 ± 34.7	+21.6
	<i>B. subtilis</i>	377	353.43 ± 22.81	+1.03
	<i>S. cerevisiae</i> (Guelzim)	290	260 ± 24.7	+1.21
	<i>S. cerevisiae</i> (Balaji / TF)	299	262.88 ± 20.80	1.74
B - Short	<i>E. coli</i> (Shen-Orr)	1422	1367.3 ± 15.79	+3.5
	<i>E. coli</i> (RDB 5.5)	1193	453.69 ± 40.88	+18
	<i>B. subtilis</i>	676	654.04 ± 27.15	+0.19
	<i>S. cerevisiae</i> (Guelzim)	365	296.35 ± 25.76	+2.67
	<i>S. cerevisiae</i> (Balaji / TF)	572	543.97 ± 18.25	+1.54

Supplementary Table S2: *Long* and *short* path hierarchy and gene duplications. The model predicts that both hierarchies are conserved, in absence of rewiring and multiple AR duplications. In particular, we expect to find many members of the same homology class falling in the same computational layer, defined by longest paths to a root or shortest paths. Preservation of the layer structure, measured by the number of homolog pairs occupying the same layer. (A) Longest path. (B) Shortest path (data for the Balaji data-set refer to the TF-TF interaction subnetwork). In *E. Coli* both data sets indicate a strong tendency to have duplicates in the same layer, with both definitions of hierarchy. In *B. subtilis* we find a similar trend only for the long-path hierarchy (the signal is much weaker, possibly due to the small size of the sample). Finally, in yeast we also find evidence for hierarchy conservation all data sets. In the larger dataset, the long-path hierarchy was evaluated for the hierarchical component only.



Supplementary Figure S3: Heterodimer / homodimer TF duplication in yeast. The x axis reports the observable quantifying the co-occurrence of homodimers and heterodimers in the same homology class (the sum over classes of the product of number of heterdimers and homodimers within each class), for the empirical network of yeast (Balaji data-set), compared to randomizations (histogram). The histogram corresponds to the data shown in Table 3 of the main text.