# Evolution of binding preferences among whole-genome duplicated transcription factors

Tamar Gera[1,3], Felix Jonas[1,3], Roye More[1], and Naama Barkai[1]*

[1]Department of Molecular Genetics, Weizmann Institute of Science; Rehovot 76100, Israel.

[3]Co-first authors

*Corresponding author. Email: naama.barkai@weizmann.ac.il

**Abstract**

Throughout evolution, new transcription factors (TFs) emerge by gene duplication, promoting growth and rewiring of transcriptional networks. How TF duplicates diverge is known for only a few studied cases. To provide a genome-scale view, we considered the 35% of budding yeast TFs, classified as whole-genome duplication (WGD)-retained paralogs. Using high-resolution profiling, we find that ~60% of paralogs evolved differential binding preferences. We show that this divergence results primarily from variations outside the DNA binding domains (DBDs), while DBD preferences remain largely conserved. Analysis of non-WGD orthologs revealed that ancestral preferences are unevenly split between duplicates, while new targets are acquired preferentially by the least conserved paralog (biased sub/neo-functionalization). Dimer-forming paralogs evolved mostly one-sided dependency, while other paralogs interacted through low-magnitude DNA-binding competition that minimized paralog interference. We discuss the implications of our findings for the evolutionary design of transcriptional networks.

**Main**

Transcription factors (TFs) bind at regulatory regions to activate or repress transcription. Cells express hundreds of TFs that together transit cells between different expression programs. Despite rapid advances, our understanding of transcriptional networks is still fragmented (*1*). For example, different TFs that bind to similar DNA sequences *in-vitro* (*2-10*) localize to different genomic sites *in-vivo* through mechanisms still poorly understood. Further, with increasing organism complexity, new TFs emerge, yet we know little about the means through which these emerging TFs adapt new targets and integrate into the existing transcriptional network.

Gene duplication is the major source of new TFs (*11-15*), with whole-genome duplications (WGDs) playing a particularly important role (*16-26*). In budding yeast, ~35% of all TFs are associated with a single WGD event dating ~100 million years ago (Fig. 1A) (*19, 20*). We reasoned that this set of TF duplicates, all generated at the same time and subjected to the same evolutionary history, provides a convenient platform for studying the fate of retained duplicated TF genes.

TF duplicates (paralogs) can diverge through changes in expression, regulation or function (*27-44*). Of particular interest is the selection of *in-vivo* binding sites, as these define potential regulatory targets. Mechanisms driving such divergence include changes in co-factor binding (*45*) or in DNA motif preferences (*46-51*). The prevalence of these different scenarios is still unclear, since studied cases considered only a few paralogs, of different ages and origins, and divergence was studied at individual targets. We therefore aimed to provide a genome-scale view, by comparing genome-wide binding preferences among the full set of WGD TFs in budding yeast.

## Results

### *Divergence of binding preferences among WGD TFs*

Within the *Saccharomyces cerevisiae* genome database (SGD) (*52*), 82 DNA binding domain (DBD)-containing proteins are classified as WGD-retained paralogs (Fig. 1A,B). We refined this list to include only pairs where both paralogs act as specific TFs (35 pairs, table S1), and defined the binding locations of these TFs across the genome using ChEC-seq (*53*). A total of 30 pairs (60 TFs) were successfully profiled, as verified by data reproducibility (Pearson's r>0.95 in promoter binding preferences, Fig. 1C,D) and manual literature survey (fig. S1). A large fraction of TFs were bound at their own or their paralog's promoters, potentially forming regulatory circuits (*15*) (Fig. 1E).

Perhaps unexpectedly, binding preferences were conserved (Pearson's r>0.8) among ~40% of paralogs, most notably these of the C2H2 zinc-finger family (6/10 pairs, e.g. Met31/Met32, Fig. 1C,D). Furthermore, most diverging paralogs still shared promoter targets (Fig. 1F). In some cases, the two duplicates localized to the same promoters but with different relative strengths (e.g. Gzf3/Dal80, Fig. 1C). In other cases, some promoters were bound by both paralogs, while others by just one (e.g. Ace2/Swi5, Fig. 1C). Therefore, binding preferences diverge at a rate that differs between pairs, and, within each pair, differs between individual promoters.

### *Paralog diverge through variations outside their DBDs*

TFs localize to genomic sites containing short motif sequences bound by their DBDs. The *in-vivo* binding could therefore diverge through DBD variations that modify motif preferences. To compare DBD sequences among paralogs, we aligned each pair and classified residues into these

defining the DBD family (e.g. C/H residues in the C2H2 zinc-finger domains), those that contribute to DNA motif preference (*54*), and the remaining ones (Fig. 2A,B and fig. S2).

Sequence conservation varied between DBD families (Fig. 2B). In particular, specificity-conferring residues often varied between paralogs of the fungal-specific zinc-cluster family (Fig. 2B), but remained invariant between paralogs of the C2H2 zinc-finger family and, to a lesser extent, in other families (e.g. Rph1/Gis1, Fig. 2A,B). Examining motif preferences derived from *in-vitro* data, we noted that reported preferences (*2, 54*) are often (although not always) similar amongst paralogs (Fig. 2A and fig. S2).

DBD variations may therefore contribute to the divergence of zinc-cluster paralogs, but appear to play a lesser role in paralogs of the C2H2 zinc finger and perhaps other families. To test this, we swapped DBDs between paralogs (Fig. 2C), reasoning that if DBD variations drive divergence, swapping DBDs would swap promoter preferences. Conversely, if critical variations are located outside the DBD, swapping will be of little effect. Consistent with our sequence analysis, DBD swapping perturbed binding for three of the four zinc-cluster TFs tested, although in none of these cases was DBD swapping sufficient for swapping promoter preferences (Fig. 2D,E). The fourth TF (Yrr1) remained largely invariant (Pearson's $r>0.8$) to DBD swapping, as did the eleven additionally tested TFs, taken from six different families. Of note, this invariance to DBD swapping characterizing most TFs was observed not only when comparing promoter preferences, but also when comparing *in-vivo* preferences to DNA *7*-mers (Fig. 2D and fig. S2). We conclude that, for most duplicate pairs, the variations driving divergence in promoter binding preferences are located outside the DBDs.

### *Dependencies and competitions between TF paralogs*

Evolved interactions between paralogs could affect binding preferences. TFs that bind DNA as dimers, for example, may bind as heterodimers to a subset of sites. Paralogs may also compete for binding, either directly or by interacting with a shared co-factor. In the broader context, cooperative interactions, where a TF depends on its paralog, may increase mutation fragility whereas binding competition, which allows a TF to access its perturbed paralog's sites, may increase mutation robustness (Fig. 3A). Both effects were observed in the context of protein-protein interactions (*44*).

To test the prevalence of cooperative or competitive interactions, we measured TF binding upon paralog deletion, testing 55/60 TFs in our dataset (Fig. 3B and fig. S3). Two TFs completely lost binding signals (Pip2, Hms2), and additional two lost binding to their respective paralogs' targets (Dal80, Tbs1). These large-scale effects, however, were infrequent (Fig. 3B,C). Cooperative interactions were generally minor (e.g. Stp2), as were compensatory interactions (e.g. Pdr3 or Ecm22; Fig. 3C,D). Therefore, strong interactions between TFs paralogs are rare and existing ones tend to increase mutation fragility.

### *Classifying paralogs' evolutionary paths by analyzing non-WGD orthologs*

Two prevailing models explain paralog divergence (*55-63*): In the first, ancestral functions split between duplicates (sub-functionalization) while, in the second, one duplicate retains ancestral functions, while the second adapts a new role (neo-functionalization, Fig. 4A). As a first approach to distinguish these two scenarios, we used phylogenetic sequence analysis to compare the evolutionary rates of the regions outside the conserved DBD (Fig. 4B). This analysis is informative, since a neo-functionalizing paralog undergoes a period of relaxed selection, followed by rapid evolution, and therefore evolves at an accelerated rate (*64, 65*). Diverging paralogs of the

C2H2 zinc-finger family evolved symmetrically, at rates that did not distinguish between paralogs, consistent with sub-functionalization. By contrast, diverging paralogs of other families evolved asymmetrically, suggesting dominant neo-functionalization (Fig. 4B).

To test experimentally for sub- and neo-functionalization, we compared binding preferences of paralogs in our set to that of a corresponding non-WGD ortholog (*Kluyveromyces lactis* TF, expressed within *S. cerevisiae;* Fig. 4C). We reasoned that, in terms of binding preferences, this non-WGD ortholog might serve as a good proxy for the ancestor (*15, 66, 67*). This was the case in pairs with clear expectations: binding of the *K. lactis* ortholog of Ixr1/Abf2 was indistinguishable from that of Ixr1 (Fig. 4C), consistent with Abf2 adopting a mitochondria role (*68*) (note Abf2's accelerated sequence evolution, Fig. 4B). Similarly, *K. lactis* orthologs of non-diverging duplicates retained highly similar binding preferences (Rsf2/Tda9 Fig. 4D and fig. S4 for additional pairs). Together, these results support the use of non-WGD orthologs to approximate ancestral preferences.

We next extended the analysis to divergent duplicates. We observed cases of clear sub- and neo-functionalization (Rph1/Gis1 and Vhr1/Vhr2, respectively Fig. 4D,E), but most pairs showed a combination of the two scenarios. Of note, in all 11 diverging cases, at least one paralog was more similar to the ortholog than to the other *S. cerevisiae* paralog and in 9 of them, the conserved paralog was also slower to evolve (Fig. 4E). Together, these results suggest that the prevalent route for diversification is through biased neo/sub-functionalization: ancestral preferences split unevenly between the duplicates, coupled to biased acquisition of novel targets (Fig. 4F and fig. S4).

***Divergence of zinc-cluster paralogs: motif preferences and paralog interactions***

To gain molecular insights into the divergence of specific duplicates, we examined individual pairs, focusing first on the fungal-specific zinc-cluster family. TFs of this family bind DNA as dimers, recognizing a composite DNA motif that includes two spaced CGG sites (*69*). Binding specificity depends on the orientation of the DBD-bound CGG triplets, and the spacer length, which may depend on an unstructured linker flanking the DBD (Fig. 5A). As described above, when compared to other families, zinc-cluster paralogs showed a greater dependence on DBD variations (cf. Fig. 2) and were more likely to interact (cf. Fig. 3).

We searched for CGG-containing motifs in regions bound by the zinc-cluster TFs in our dataset. In three of the five diverging pairs, differences in binding preferences correlated with differential motif preferences (Fig. 5A: differential spacing/orientation - Hal9/Tbs1 and Oaf1/Pip2 or variation within the CGG - Pdr1/Pdr3). All three pairs interacted: Hal9/Tbs1 recruited each other to their preferred sites, likely through heterodimerization, and Pip2 localized exclusively to a subset of Oaf1-preferred targets in an Oaf1-dependent manner, consistent with obligatory heterodimerization (*70*). Notably, binding preferences of both heterodimers correlated well with that of the *K. lactis* ortholog. In the case of Pdr1/3, interaction took the form of competition, with Pdr1 outcompeting Pdr3 from accessing its preferred motif.

In these three cases described above, binding preferences evolved through a combination of DNA-motif preferences and protein-protein interactions. Contrasting these, CGG-containing motifs at Yrr1/Pdr8- or Upc2/Ecm22-bound sites did not distinguish between paralogs (Fig. 5A,C and fig. S5). Rather, paralogs localized to the same motifs, but within different promoters. This divergence of promoter preferences was largely DBD-independent in the case of Yrr1/Pdr8, but greatly benefitted from DBD-dependent competition in the case of Upc2/Ecm22 (Fig. 5C). In fact, upon

UPC2 deletion or DBD swapping, Ecm22 gained access to strong Upc2 sites (Fig. 5C; note correspondence to *K. lactis* sites). Of note, while DBD-swapping retained binding at the TF-specific sites, it also allowed increased access to non-specific sites, suggesting co-evolution of the DBD and the linker domain, both of which varied substantially between the paralogs. We conclude that zinc-cluster paralogs evolved largely, but not exclusively, through changes in motif preferences or affinity, resulting from combined effects of variations within, and outside, their DBDs.

***Resolution of paralog interference through competitive binding***

Our analysis above described plausible evolutionary scenarios governing zinc-cluster paralog divergence, where motif preferences appear to play an important role. In most other paralogs, changes in motif preference, if exist, appear secondary to the major effective variations located outside the DBD. Still, even in such cases, DBD variations may play a role in resolving paralog interference (*45, 71*). Thus, since the two paralogs are initially redundant, divergence of binding preferences requires not only the acquisition of differential preferences, but also limiting residual, and possibly interfering paralog's cross binding. In the case of the non-WGD pair Mcm1/Arg80, for example, such interference was resolved by weakening direct Arg80-DNA binding (*45, 71*). We asked whether similar effects can be observed for WGD paralogs.

DBD variations contributed little to the divergence of Rph1/Gis1, the most diverged C2H2 zinc-finger paralogs. We noted, however, that Rph1 gained residual access to Gis1-binding sites upon GIS1 deletion or DBD swapping (Fig. 6A,B), and Rph1-exclusive targets contained a specific variant of the common Gis1/Rph1 motif, lacking from Gis1 binding sites (Fig. 6C,D). Phylogenetic analysis revealed two variations between the paralogs: first, Gis1 has lost a conserved demethylase activity (*72*), an event that occurred quite early following WGD (Fig. 6E and fig. S6). Second,

Rph1's DBD experienced a more recent (minor) variation altering a conserved residue. This suggests the following evolutionary scenario: first, a loss of demethylase activity allowed Gis1 to specialize towards a subset of (weaker) ancestral targets and acquire new sites, through a primarily DBD-independent evolution. Second, mutations within Rph1's DBD prevented its binding to Gis1-specialized sites, thereby reducing paralog interference (Fig. 6E).

Finally, extending the analysis to other diverging paralogs suggested additional cases conforming to this same design showing limited binding competition (Fig. 6F and fig. S6). Together, these results suggest a common path, whereby DBD-independent divergence is complemented by asymmetric competition, limiting residual paralog interference.

**Discussion**

The binding of TFs at individual regulatory regions can vary through mutations that alter the DNA sequence or mutations that change TF binding preferences. While promoter mutations are gene specific, changes in TF binding preferences will affect multiple genes, and are therefore less likely to occur (*15, 66, 67*). Evolution of binding preferences, however, is possible through TF duplication, which generates two redundant copies and relaxes selection. The evolutionary paths through which TF paralogs diverge their binding preferences may therefore hold the key to understanding the principles through which TFs select their binding locations across the genomes.

Studying a comprehensive set of WGD-retained TF paralogs, we found that the majority of pairs still share a large fraction of overlapping targets. In fact, even diverged paralogs still localized to common targets, although at different relative strengths. This gradual divergence was not explained by variations in the DBDs. In fact, we presented evidence that, with the exception of the zinc-cluster family, variations within the DBD contributed little to binding divergence. DBD preferences may play a minor role in setting promoter binding preferences but are primarily

important for stabilizing binding within selected promoters. Further, cooperation and competition act to adjust binding profiles and limit paralog interference but, with few exceptions, are not the major factors guiding divergence. In this context, gradual, and promoter-specific divergence is harder to explain within prevailing models for TF specificity, but may be more consistent with the paradigm of distributed specificity we recently proposed (*1, 73*).

At the more global level, since duplication is the major means through which new TFs emerge, the evolutionary trajectories of retained paralogs shape the transcriptional network's design. Duplicates that diverge through sub-functionalization, for example, will confer a hierarchical design, where regulatory modules are gradually refined as the network expands. By contrast, neo-functionalization may support a distributed design, where new regulatory modules can emerge largely independent of previous connectivity. Our results reveal that neo-functionalization is quite common, although it is often combined with a biased sub-functionalization, namely uneven splitting of ancestral targets.

Finally, we focused our analysis on diverging paralogs, but it is notable that a significant fraction of paralog pairs (~40%) still binds at practically identical sites. Retention of these paralogs is therefore due to other properties. Duplicates' tendency to cross-bind their own promoters may suggest that interactions between duplicates have evolved to confer beneficial dynamic properties not achieved by a single TF (*15, 74, 75*). Further studies are required to examine the potential advantages provided by such circuit-forming duplicates.

## References and Notes


1. T. Jana, S. Brodsky, N. Barkai, Speed-Specificity Trade-Offs in the Transcription Factors Search for Their Genomic Binding Sites. *Trends Genet* **37**, 421-432 (2021).
2. Matthew T. Weirauch *et al.*, Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell* **158**, 1431-1443 (2014).
3. G.-H. Wei *et al.*, Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *The EMBO Journal* **29**, 2147-2160 (2010).
4. A. Jolma *et al.*, DNA-Binding Specificities of Human Transcription Factors. *Cell* **152**, 327-339 (2013).
5. V. Matys *et al.*, TRANSFAC® and its module TRANSCompel®: transcriptional gene regulation in eukaryotes. *Nucleic Acids Research* **34**, D108-D110 (2006).
6. T. Siggers, J. Reddy, B. Barron, M. L. Bulyk, Diversification of transcription factor paralogs via noncanonical modularity in C2H2 zinc finger DNA binding. *Mol Cell* **55**, 640-648 (2014).
7. S. Nakagawa, S. S. Gisselbrecht, J. M. Rogers, D. L. Hartl, M. L. Bulyk, DNA-binding specificity changes in the evolution of forkhead transcription factors. *Proc Natl Acad Sci U S A* **110**, 12349-12354 (2013).
8. J. M. Rogers, M. L. Bulyk, Diversification of transcription factor-DNA interactions and the evolution of gene regulatory networks. *Wiley Interdiscip Rev Syst Biol Med*, e1423 (2018).
9. M. F. Berger *et al.*, Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266-1276 (2008).
10. N. Shen *et al.*, Divergence in DNA Specificity among Paralogous Transcription Factors Contributes to Their Differential In Vivo Binding. *Cell systems* **6**, 470-483.e478 (2018).
11. S. Ohno, U. Wolf, N. B. Atkin, Evolution from fish to mammals by gene duplication. *Hereditas* **59**, 169-187 (1968).
12. A. Rosanova, A. Colliva, M. Osella, M. Caselle, Modelling the evolution of transcription factor binding preferences in complex eukaryotes. *Sci Rep* **7**, 7596 (2017).
13. J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, N. M. Luscombe, A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics* **10**, 252-263 (2009).
14. M. Levine, R. Tjian, Transcription regulation and animal diversity. *Nature* **424**, 147-151 (2003).
15. S. A. Teichmann, M. M. Babu, Gene regulatory network growth by duplication. *Nat Genet* **36**, 492-496 (2004).
16. G. Blanc, K. H. Wolfe, Functional Divergence of Duplicated Genes Formed by Polyploidy during Arabidopsis Evolution. *The Plant Cell* **16**, 1679-1691 (2004).

17. S. Maere *et al.*, Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci U S A* **102**, 5454-5459 (2005).

18. H. V. S. Rody, G. J. Baute, L. H. Rieseberg, L. O. Oliveira, Both mechanism and age of duplications contribute to biased gene retention patterns in plants. *BMC genomics* **18**, 46-46 (2017).

19. K. H. Wolfe, D. C. Shields, Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708-713 (1997).

20. M. Marcet-Houben, T. Gabaldón, Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol* **13**, e1002220 (2015).

21. J. A. Birchler, N. C. Riddle, D. L. Auger, R. A. Veitia, Dosage balance in gene regulation: biological implications. *Trends Genet* **21**, 219-226 (2005).

22. M. Freeling, B. C. Thomas, Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res* **16**, 805-814 (2006).

23. P. P. Edger, J. C. Pires, Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res* **17**, 699-717 (2009).

24. L. G. Lundin, Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* **16**, 1-19 (1993).

25. P. Dehal, J. L. Boore, Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* **3**, e314 (2005).

26. T. Blomme *et al.*, The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol* **7**, R43 (2006).

27. T. Vavouri, J. I. Semple, B. Lehner, Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. *Trends Genet* **24**, 485-488 (2008).

28. A. DeLuna *et al.*, Exposing the fitness contribution of duplicated genes. *Nat Genet* **40**, 676-681 (2008).

29. G. Diss, D. Ascencio, A. DeLuna, C. R. Landry, Molecular mechanisms of paralogous compensation and the robustness of cellular networks. *J Exp Zool B Mol Dev Evol* **322**, 488-499 (2014).

30. R. Kafri, M. Levy, Y. Pilpel, The regulatory utilization of genetic redundancy through responsive backup circuits. *Proceedings of the National Academy of Sciences* **103**, 11653-11658 (2006).

31. I. Wapinski, A. Pfeffer, N. Friedman, A. Regev, Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449**, 54-61 (2007).

32. Z. Gu *et al.*, Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63-66 (2003).

33. B. VanderSluis *et al.*, Genetic interactions reveal the evolutionary trajectories of duplicate genes. *Molecular systems biology* **6**, 429-429 (2010).

34. V. Charoensawan, D. Wilson, S. A. Teichmann, Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucleic Acids Research* **38**, 7364-7377 (2010).

35. D. L. Des Marais, M. D. Rausher, Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454**, 762-765 (2008).

36. T. L. Hsiao, D. Vitkup, Role of duplicate genes in robustness against deleterious human mutations. *PLoS Genet* **4**, e1000014 (2008).

37. J. Ihmels, S. R. Collins, M. Schuldiner, N. J. Krogan, J. S. Weissman, Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. *Mol Syst Biol* **3**, 86 (2007).

38. E. Kuzmin *et al.*, Exploring whole-genome duplicate gene retention with complex genetic interaction analysis. *Science* **368**, eaaz5667 (2020).

39. I. M. Ehrenreich, Evolution after genome duplication. *Science* **368**, 1424-1425 (2020).

40. J. L. Payne, A. Wagner, Mechanisms of mutational robustness in transcriptional regulation. *Front Genet* **6**, 322 (2015).

41. W.-H. Chen, X.-M. Zhao, V. van Noort, P. Bork, Human Monogenic Disease Genes Have Frequently Functionally Redundant Paralogs. *PLOS Computational Biology* **9**, e1003073 (2013).

42. A. Burga, M. O. Casanueva, B. Lehner, Predicting mutation outcome from early stochastic variation in genetic interaction partners. *Nature* **480**, 250-253 (2011).

43. L. T. MacNeil, A. J. M. Walhout, Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Research* **21**, 645-657 (2011).

44. G. Diss *et al.*, Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. *Science* **355**, 630-634 (2017).

45. C. R. Baker, V. Hanson-Smith, A. D. Johnson, Following Gene Duplication, Paralog Interference Constrains Transcriptional Circuit Evolution. *Science* **342**, 104-108 (2013).

46. J. C. Perez *et al.*, How duplicated transcription regulators can diversify to govern the expression of nonoverlapping sets of genes. *Genes Dev* **28**, 1272-1277 (2014).

47. Victoria M. Blake, S. Barolo, Genome Evolution: How Sister Genes Grow Apart. *Current Biology* **24**, R695-R697 (2014).

48. A. N. McKeown *et al.*, Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. *Cell* **159**, 58-68 (2014).

49. J. T. Bridgham, J. E. Brown, A. Rodríguez-Marí, J. M. Catchen, J. W. Thornton, Evolution of a New Function by Degenerative Mutation in Cephalochordate Steroid Receptors. *PLOS Genetics* **4**, e1000191 (2008).

50. M. V. Humbert, R. M. Rasia, S. K. Checa, F. C. Soncini, Protein signatures that promote operator selectivity among paralog MerR monovalent metal ion regulators. *J Biol Chem* **288**, 20510-20519 (2013).

51. K. Pougach *et al.*, Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. *Nature Communications* **5**, 4868 (2014).

52. J. M. Cherry *et al.*, Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* **40**, D700-705 (2012).

53. G. E. Zentner, S. Kasinathan, B. Xin, R. Rohs, S. Henikoff, ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. *Nature Communications* **6**, 8733 (2015).

54. S. A. Lambert *et al.*, Similarity regression predicts evolution of transcription factor sequence specificity. *Nat Genet* **51**, 981-989 (2019).

55. K. Voordeckers *et al.*, Reconstruction of Ancestral Metabolic Enzymes Reveals Molecular Mechanisms Underlying Evolutionary Innovation through Gene Duplication. *PLOS Biology* **10**, e1001446 (2012).

56. M. Lynch, J. S. Conery, The Evolutionary Fate and Consequences of Duplicate Genes. *Science* **290**, 1151-1155 (2000).

57. R. P. Sugino, H. Innan, Selection for more of the same product as a force to enhance concerted evolution of duplicated genes. *Trends in Genetics* **22**, 642-644 (2006).

58. M. Lynch, A. Force, The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459-473 (2000).

59. G. C. Conant, K. H. Wolfe, Functional Partitioning of Yeast Co-Expression Networks after Genome Duplication. *PLOS Biology* **4**, e109 (2006).

60. A. Force *et al.*, Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* **151**, 1531-1545 (1999).

61. F. A. Kondrashov, I. B. Rogozin, Y. I. Wolf, E. V. Koonin, Selection in the evolution of gene duplications. *Genome Biol* **3**, Research0008 (2002).

62. X. He, J. Zhang, Rapid Subfunctionalization Accompanied by Prolonged and Substantial Neofunctionalization in Duplicate Gene Evolution. *Genetics* **169**, 1157-1164 (2005).

63. K. Nowick, L. Stubbs, Lineage-specific transcription factors and the evolution of gene regulatory networks. *Brief Funct Genomics* **9**, 65-78 (2010).

64. C. Pegueroles, S. Laurie, M. M. Albà, Accelerated Evolution after Gene Duplication: A Time-Dependent Process Affecting Just One Copy. *Molecular Biology and Evolution* **30**, 1830-1842 (2013).

65. K. P. Byrne, K. H. Wolfe, Consistent Patterns of Rate Asymmetry and Gene Loss Indicate Widespread Neofunctionalization of Yeast Genes After Whole-Genome Duplication. *Genetics* **175**, 1341-1350 (2007).

66. C. C. Hsia, W. McGinnis, Evolution of transcription factor function. *Current Opinion in Genetics & Development* **13**, 199-206 (2003).

67. S. B. Carroll, Evolution at Two Levels: On Genes and Form. *PLOS Biology* **3**, e245 (2005).

68. F. Caron, C. Jacq, J. Rouvière-Yaniv, Characterization of a histone-like protein extracted from yeast mitochondria. *Proc Natl Acad Sci U S A* **76**, 4265-4269 (1979).

69. S. MacPherson, M. Larochelle, B. Turcotte, A fungal family of transcriptional regulators: the zinc cluster proteins. *Microbiology and molecular biology reviews : MMBR* **70**, 583-604 (2006).

70. H. Rottensteiner, A. J. Kal, B. Hamilton, H. Ruis, H. F. Tabak, A heterodimer of the Zn2Cys6 transcription factors Pip2p and Oaf1p controls induction of genes encoding peroxisomal proteins in Saccharomyces cerevisiae. *Eur J Biochem* **247**, 776-783 (1997).

71. E. Kaltenegger, D. Ober, Paralogue Interference Affects the Dynamics after Gene Duplication. *Trends in Plant Science* **20**, 814-821 (2015).

72. R. J. Klose, E. M. Kallin, Y. Zhang, JmjC-domain-containing proteins and histone demethylation. *Nature Reviews Genetics* **7**, 715-727 (2006).

73. B. S *et al.*, Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Molecular cell* **79**, (2020).

74. M. Chapal, S. Mintzer, S. Brodsky, M. Carmi, N. Barkai, Resolving noise–control conflict by gene duplication. *PLOS Biology* **17**, e3000289 (2019).

75. B. Lehner, Conflict between noise and plasticity in yeast. *PLoS Genet* **6**, e1001185 (2010).

76. B. Mészáros, G. Erdos, Z. Dosztányi, IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res* **46**, W329-W337 (2018).

77. R. a. M. G. a. H. J. Anand, Cas9-mediated gene editing in Saccharomyces cerevisiae. *Protocol Exchange*, (2017).

78. K. Labun *et al.*, CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research* **47**, (2021).

79. B. CB *et al.*, Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast (Chichester, England)* **14**, (1998).

80. G. RD, S. RH, W. AR, W. RA, Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast (Chichester, England)* **11**, (1995).

81. M. Meurer *et al.*, Genome-wide C-SWAT library for high-throughput yeast genome tagging. *Nat Methods* **15**, 598-600 (2018).

82. U. Weill *et al.*, Genome-wide SWAp-Tag yeast libraries for proteome exploration. *Nat Methods* **15**, 617-622 (2018).

83. O. Lupo *et al.*, Accumulation of cis- and trans-regulatory variations is associated with phenotypic divergence of a complex trait between yeast species. *G3 Genes|Genomes|Genetics* **11**, (2021).

84. H. DW, D. TN, M. EG, Fluorescence resonance energy transfer using color variants of green fluorescent protein. *Methods in enzymology* **351**, (2002).

85. C. Janke *et al.*, A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21**, 947-962 (2004).

86. P. J. Skene, S. Henikoff, An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, (2017).

87. R. Blecher-Gonen *et al.*, High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nature Protocols* **8**, 539-554 (2013).

88. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, 10 (2011).

89. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

90. D. Park, A. R. Morris, A. Battenhouse, V. R. Iyer, Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. *Nucleic Acids Res* **42**, 3736-3749 (2014).

91. V. Pelechano, W. Wei, L. M. Steinmetz, Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497**, 127-131 (2013).

92. R. A. Policastro, R. T. Raborn, V. P. Brendel, G. E. Zentner, Simple and efficient profiling of transcription initiation and transcript levels with STRIPE-seq. *Genome Res* **30**, 910-923 (2020).

93. R. V. Chereji, S. Ramachandran, T. D. Bryson, S. Henikoff, Precise genome-wide mapping of single nucleosomes and linkers in vivo. *Genome Biol* **19**, 19 (2018).

94. J. Mistry *et al.*, Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412-D419 (2021).

95. F. Madeira *et al.*, The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**, W636-W641 (2019).

96. T. J. Wheeler, J. Clements, R. D. Finn, Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics* **15**, 7 (2014).

97. K. P. Byrne, K. H. Wolfe, The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res* **15**, 1456-1461 (2005).

98. S. Moretti *et al.*, The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. *Nucleic Acids Res* **35**, W645-648 (2007).

99. L. T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-274 (2015).

100. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* **35**, 518-522 (2018).

**Acknowledgments**: We like to thank Sagie Brodsky, Tamar Jana and Offir Lupo for their strains and all Barkai lab members for fruitful scientific discussions and careful reading of the manuscript.

**Funding:** This work was funded by the Israel Science Foundation, ERC and the Minerva Foundation.

**Author contributions:**

Conceptualization: NB, TG; Methodology: TG, FJ, NB; Investigation: TG, RM; Visualization: TG, FJ; Formal analysis: TG, FJ; Data curation and Software: TG, FJ; Funding acquisition: NB; Supervision: NB; Writing – original draft: NB, TG, FJ.

**Competing interests:** The authors declare no competing interests.

**Data and materials availability:** All yeast strains are available upon request, raw and processed data is available at GEO (GSE179430). MATLAB scripts for analysis and visualization are available upon request.

**Supplementary Materials**

Materials and Methods

Figs. S1 to S6

Tables S1 to S4

References (*77–100*)

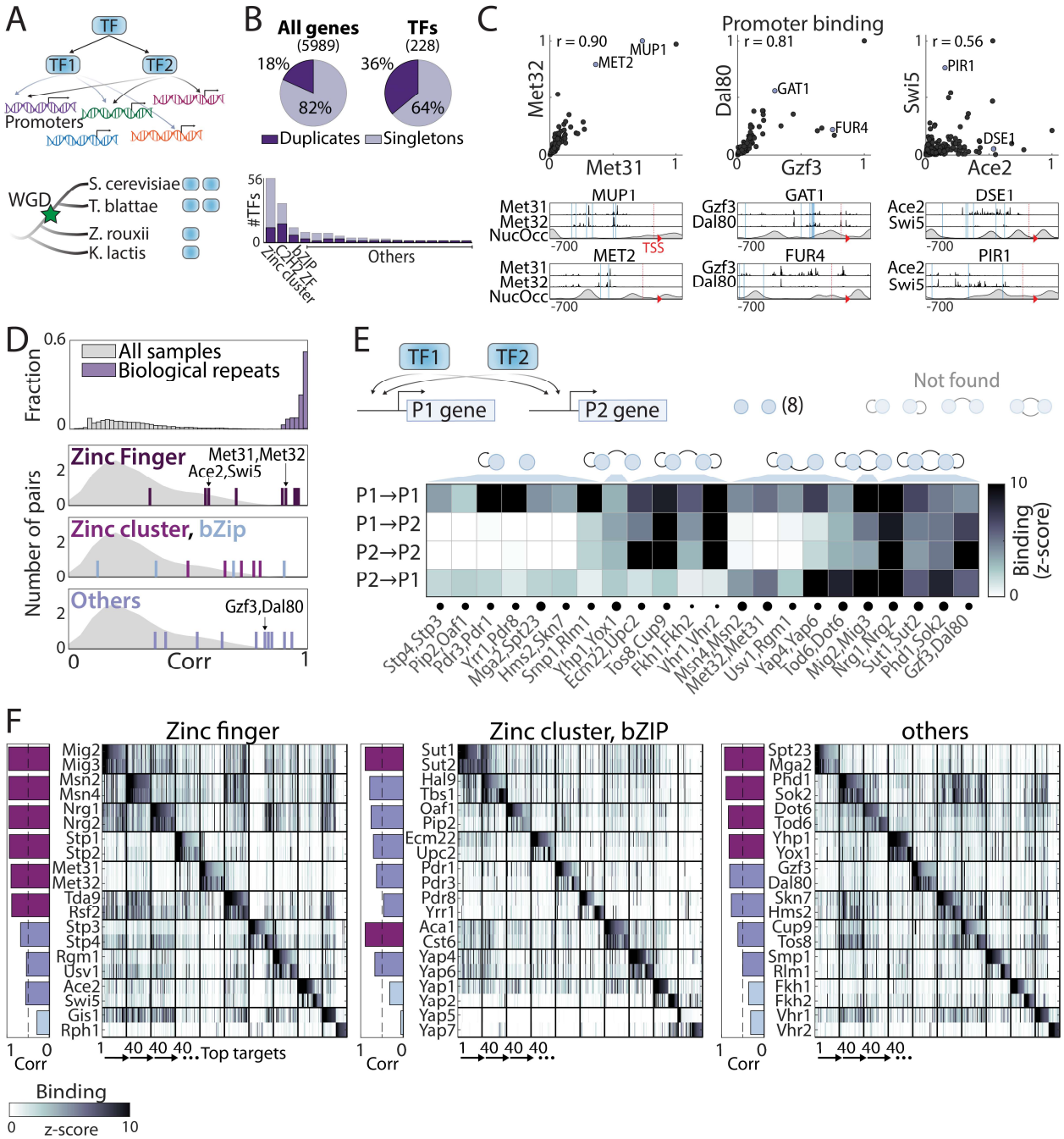

A
TF
TF1
TF2
Promoters
WGD
S. cerevisiae
T. blattae
Z. rouxii
K. lactis
B
All genes
(5989)
TFs
(228)
18%
82%
36%
64%
Duplicates
Singletons
#TFs
Zinc cluster
C2H2 ZF
bZIP
Others
C
Promoter binding
r = 0.90
MUP1
MET2
Met32
Met31
r = 0.81
GAT1
FUR4
Dal80
Gzf3
r = 0.56
PIR1
DSE1
Swi5
Ace2
Met31
Met32
NucOcc
TSS
-700
Gzf3
Dal80
Ace2
Swi5
D
Fraction
All samples
Biological repeats
Number of pairs
Zinc Finger
Met31,Met32
Ace2,Swi5
Zinc cluster, bZip
Others
Gzf3,Dal80
Corr
E
TF1
TF2
P1 gene
P2 gene
(8)
Not found
P1→P1
P1→P2
P2→P2
P2→P1
Binding (z-score)
Stp4,Stp3
Pip2,Oaf1
Pdr3,Pdr1
Yrr1,Pdr8
Mga2,Spt23
Hms2,Skn7
Smp1,Rlm1
Yhp1,Yox1
Ecm22,Upc2
Tos8,Cup9
Fkh1,Fkh2
Vhr1,Vhr2
Msn4,Msn2
Met32,Met31
Usv1,Rgm1
Yap4,Yap6
Tod6,Dot6
Mig2,Mig3
Nrg1,Nrg2
Sut1,Sut2
Phd1,Sok2
Gzf3,Dal80
F
Zinc finger
Zinc cluster, bZIP
others
Mig2
Mig3
Msn2
Msn4
Nrg1
Nrg2
Stp1
Stp2
Met31
Met32
Tda9
Rsf2
Stp3
Stp4
Rgm1
Usv1
Ace2
Swi5
Gis1
Rph1
Sut1
Sut2
Hal9
Tbs1
Oaf1
Pip2
Ecm22
Upc2
Pdr1
Pdr3
Pdr8
Yrr1
Aca1
Cst6
Yap4
Yap6
Yap1
Yap2
Yap5
Yap7
Spt23
Mga2
Phd1
Sok2
Dot6
Tod6
Yhp1
Yox1
Gzf3
Dal80
Skn7
Hms2
Cup9
Tos8
Smp1
Rlm1
Fkh1
Fkh2
Vhr1
Vhr2
Corr
Top targets
Binding
z-score

**Fig. 1. Promoter binding preferences of WGD TF-duplicates.**

**(A-B)** Whole-genome duplication (WGD) shaped the budding yeast transcription network: (A) TF duplicates (paralogs) can diverge to bind different targets. (B) In *S. cerevisiae*, ~35% (82) of all present day TFs are retained WGD paralogs (see fig. S1). **(C-D)** Binding profiles of TF paralogs: Shown in (C) are promoter binding preferences of the indicated TFs (top, each dot is a promoter, r: Pearson's correlation). Bottom plots show binding signal over individual promoters, with lines indicating transcription start site (TSS, red dashed) and locations of *in-vitro* motifs (blue). Nucleosome positioning shown in grey. (D) Distributions of correlations between promoter binding preferences of different TFs (grey), between biological repeats (purple, top) and between paralogs (colored lines, bottom). **(E)** Auto- and cross-promoter binding by TF paralogs: Promoter binding signal is shown as z-score, with potentially formed circuits shown on top. Note that 22/30 pairs form six of nine possible circuits. **(F)** Binding to top-scoring promoters: The 40 top-bound promoters by each TF pair were selected (see table S2), ordered along the x-axis, and color-coded according to binding strength by each TF (y-axis). Bars on the left show Pearson's correlations of binding preferences between the indicated pairs.

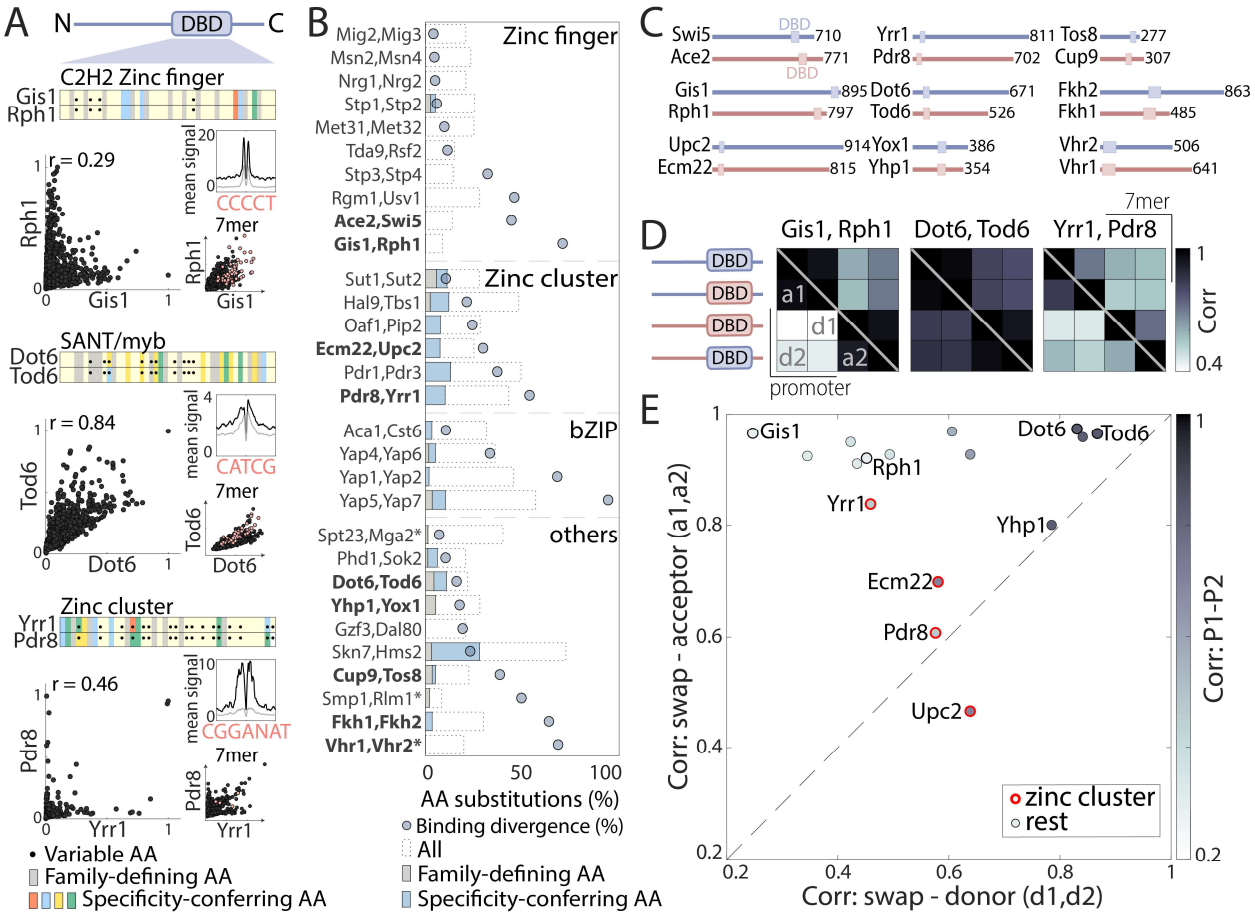

A
N
DBD
C
C2H2 Zinc finger
Gis1
Rph1
r = 0.29
mean signal
CCCCT
7mer
SANT/myb
Dot6
Tod6
r = 0.84
CATCG
Zinc cluster
Yrr1
Pdr8
r = 0.46
CGGANAT
Variable AA
Family-defining AA
Specificity-conferring AA
B
Mig2,Mig3
Msn2,Msn4
Nrg1,Nrg2
Stp1,Stp2
Met31,Met32
Tda9,Rsf2
Stp3,Stp4
Rgm1,Usv1
Ace2,Swi5
Gis1,Rph1
Zinc finger
Sut1,Sut2
Hal9,Tbs1
Oaf1,Pip2
Ecm22,Upc2
Pdr1,Pdr3
Pdr8,Yrr1
Zinc cluster
Aca1,Cst6
Yap4,Yap6
Yap1,Yap2
Yap5,Yap7
bZIP
Spt23,Mga2*
Phd1,Sok2
Dot6,Tod6
Yhp1,Yox1
Gzf3,Dal80
Skn7,Hms2
Cup9,Tos8
Smp1,Rlm1*
Fkh1,Fkh2
Vhr1,Vhr2*
others
AA substitutions (%)
Binding divergence (%)
All
Family-defining AA
Specificity-conferring AA
C
Swi5
Ace2
Yrr1
Pdr8
Tos8
Cup9
Gis1
Rph1
Dot6
Tod6
Fkh2
Fkh1
Upc2
Ecm22
Yox1
Yhp1
Vhr2
Vhr1
DBD
D
Gis1, Rph1
Dot6, Tod6
Yrr1, Pdr8
7mer
a1
d1
d2
a2
promoter
Corr
E
Corr: swap - acceptor (a1,a2)
Corr: swap - donor (d1,d2)
Corr: P1-P2
zinc cluster
rest

**Fig. 2. The contribution of DBDs to the divergence of TF paralogs.**

**(A-B)** Sequence variations between paralogs' DBDs: For all pairs, Pfam-defined DBDs were aligned and residues classified into those conserved among all family members (grey) and specificity-conferring ones (colored; blue, red, yellow and green denoting basic, acidic hydrophobic and hydrophilic residues respectively, (*54*)), as shown in (A) for the three examples (black dots indicate positions varying between paralogs). The fraction and class of varying residues is summarized in (B) (*:specificity-conferring residues not defined). Note the little correspondence between DBD sequence variations and binding profile divergence (blue dots). Paralogs chosen for further analysis are highlighted in bold. Also shown in (A) are the average signal around *in-vitro*-motif occurrences (right, top), the respective binding signals across promoters (left) and at all 7-mer sequences (right bottom, red dots indicate 7-mers containing the *in-vitro* motif). **(C-E)** DBD swapping has a minor effect on binding preferences: DBDs of the indicated paralog pairs (C, protein length indicated) were swapped, and their binding profiles mapped. Shown in (D) are the correlations between binding preferences of the indicated TFs and their swapped variants (bottom triangle promoters, top triangles 7-mers; see also fig. S2). These correlations are summarized in (E) for all tested pairs (red indicates zinc-cluster TFs; shading depicts correlation between wild type paralogs). Note that outside the zinc-cluster family, DBD swapping is of little consequence for promoter binding preferences, even among highly divergent paralogs. Within the zinc-cluster family, DBD swapping affected binding profiles, but did not recover donor profiles.

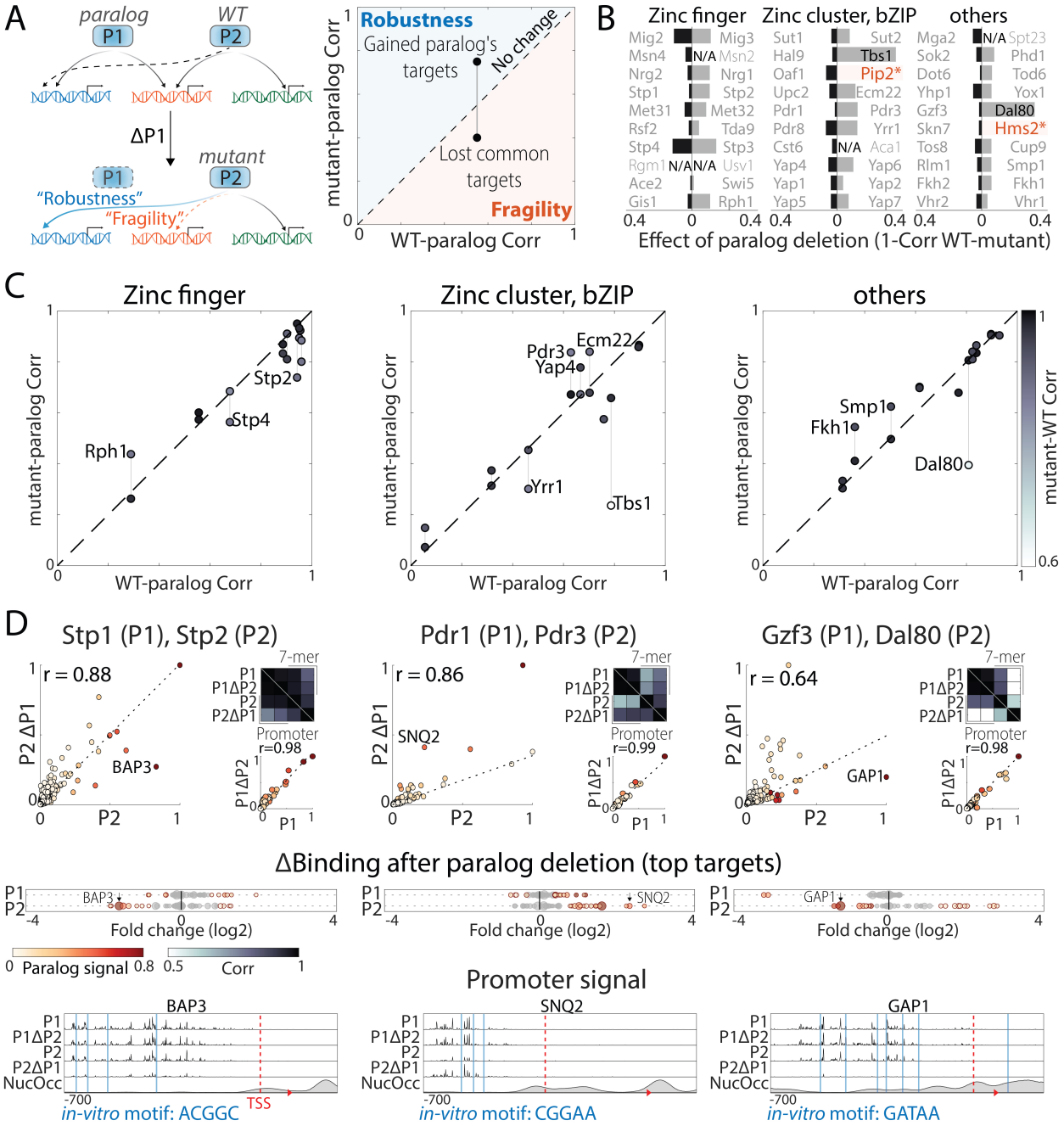

A
paralog
WT
P1
P2
ΔP1
mutant
P1
P2
"Robustness"
"Fragility"
Robustness
Gained paralog's targets
No change
Lost common targets
Fragility
mutant-paralog Corr
WT-paralog Corr
B
Zinc finger
Zinc cluster, bZIP
others
Mig2 Mig3
Msn4 N/A Msn2
Nrg2 Nrg1
Stp1 Stp2
Met31 Met32
Rsf2 Tda9
Stp4 Stp3
Rgm1 N/A N/A Usv1
Ace2 Swi5
Gis1 Rph1
Sut1 Sut2
Hal9 Tbs1
Oaf1 Pip2*
Upc2 Ecm22
Pdr1 Pdr3
Pdr8 Yrr1
Cst6 N/A Aca1
Yap4 Yap6
Yap1 Yap2
Yap5 Yap7
Mga2 N/A Spt23
Sok2 Phd1
Dot6 Tod6
Yhp1 Yox1
Gzf3 Dal80
Skn7 Hms2*
Tos8 Cup9
Rlm1 Smp1
Fkh2 Fkh1
Vhr2 Vhr1
Effect of paralog deletion (1-Corr WT-mutant)
C
Zinc finger
Zinc cluster, bZIP
others
mutant-paralog Corr
WT-paralog Corr
Stp2
Stp4
Rph1
Pdr3
Ecm22
Yap4
Yrr1
Tbs1
Smp1
Fkh1
Dal80
mutant-WT Corr
D
Stp1 (P1), Stp2 (P2)
Pdr1 (P1), Pdr3 (P2)
Gzf3 (P1), Dal80 (P2)
r = 0.88
r = 0.86
r = 0.64
P2 ΔP1
P2
BAP3
SNQ2
GAP1
7-mer
P1
P1ΔP2
P2
P2ΔP1
Promoter
r=0.98
r=0.99
r=0.98
P1ΔP2
P1
ΔBinding after paralog deletion (top targets)
Fold change (log2)
Paralog signal
Corr
Promoter signal
NucOcc
TSS
in-vitro motif: ACGGC
in-vitro motif: CGGAA
in-vitro motif: GATAA

**Fig. 3. Interactions between TF paralogs may increase network fragility.**

**(A)** Paralogs' contribution to mutation robustness or fragility: Following paralog deletion, a TF may gain access to its paralog's unique sites, potentially compensating for the loss ("robustness", blue line). Alternatively, paralogs may become interdependent ("Fragility", orange line). These interactions can be visualized by comparing the TF's binding preferences in wild type (x-axis) or paralog-deleted (y-axis) backgrounds to those of the paralog. **(B)** Paralog interactions are rare: the effect of paralog deletion on promoter binding preferences was measured for 55 TFs of 60 paralogs in our dataset. Shown is the effect of paralog deletion on binding preference for each TF. Note that most deletions were of little effect, or asymmetric in cases of large effects. Also indicated are substantial effects (TFs highlighted in black), and TFs that completely lost binding specificity (colored in orange, N/A: not profiled, see materials and methods). **(C-D)** Tendency for paralog interactions varies between TF families: (C) robustness/fragility analysis, as in (A) for all tested paralog pairs, divided into families. (D) For individual examples, binding is shown over all (top) or individual promoters (bottom), and as relative change following paralog deletion (middle; dot color and size indicate a TF's and its paralog's binding signal, respectively). Note that Stp2 and Dal80 lose binding to some of the targets shared with their respective paralog following paralog deletion ("fragility"), whereas Pdr3 gains binding to Pdr1 targets (SNQ2) upon the latter's deletion ("robustness").

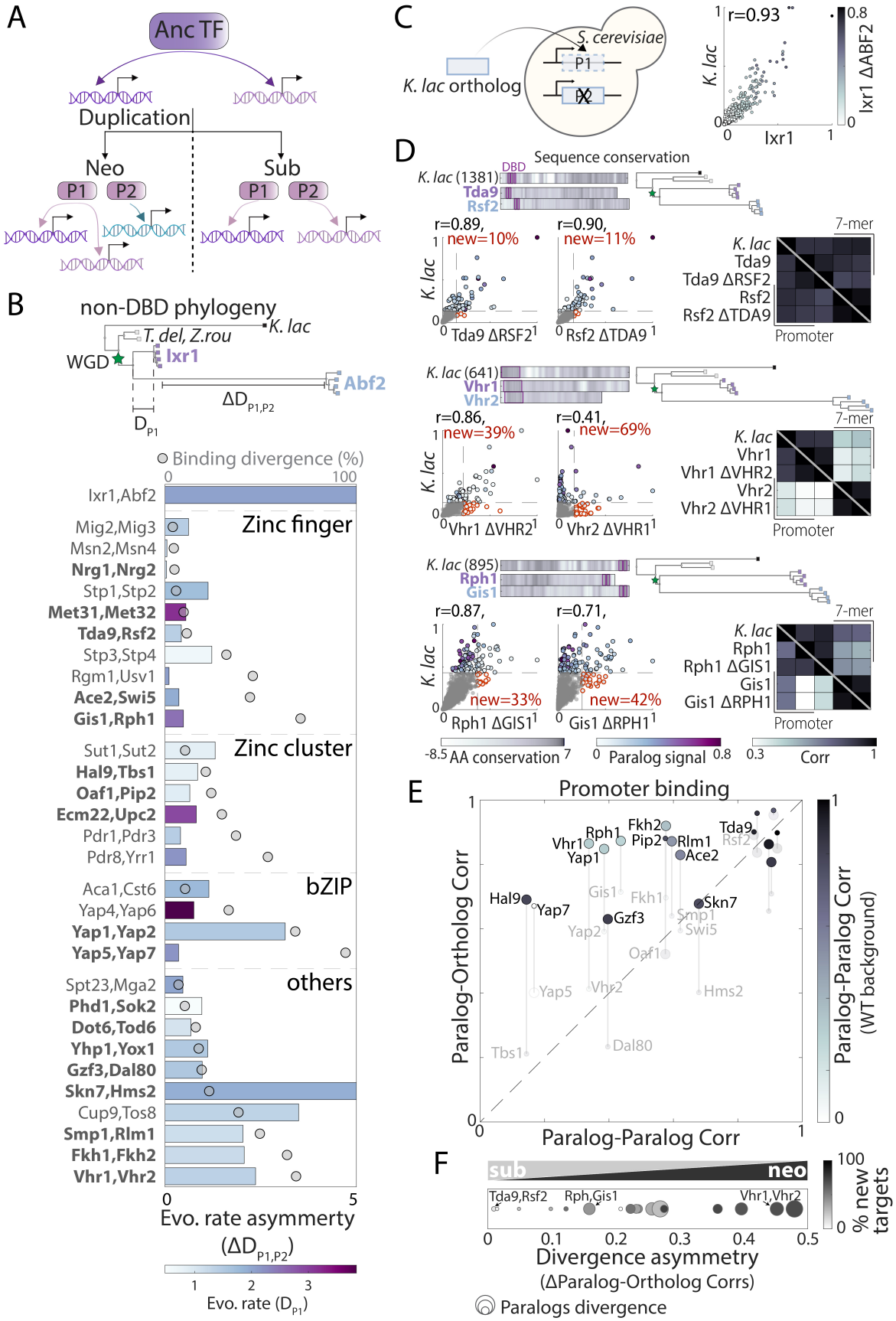

A
Anc TF
Duplication
Neo
Sub
P1
P2
B
non-DBD phylogeny
K. lac
T. del, Z.rou
WGD
Ixr1
Abf2
ΔD P1,P2
D P1
Binding divergence (%)
0
100
Ixr1,Abf2
Zinc finger
Mig2,Mig3
Msn2,Msn4
Nrg1,Nrg2
Stp1,Stp2
Met31,Met32
Tda9,Rsf2
Stp3,Stp4
Rgm1,Usv1
Ace2,Swi5
Gis1,Rph1
Zinc cluster
Sut1,Sut2
Hal9,Tbs1
Oaf1,Pip2
Ecm22,Upc2
Pdr1,Pdr3
Pdr8,Yrr1
bZIP
Aca1,Cst6
Yap4,Yap6
Yap1,Yap2
Yap5,Yap7
others
Spt23,Mga2
Phd1,Sok2
Dot6,Tod6
Yhp1,Yox1
Gzf3,Dal80
Skn7,Hms2
Cup9,Tos8
Smp1,Rlm1
Fkh1,Fkh2
Vhr1,Vhr2
0
5
Evo. rate asymmerty
(ΔD P1,P2)
1
2
3
Evo. rate (D P1)
C
S. cerevisiae
K. lac ortholog
P1
P2
r=0.93
K. lac
Ixr1
Ixr1 ΔABF2
0.8
D
Sequence conservation
K. lac (1381)
DBD
Tda9
Rsf2
r=0.89,
new=10%
r=0.90,
new=11%
Tda9 ΔRSF2
Rsf2 ΔTDA9
7-mer
K. lac
Tda9
Tda9 ΔRSF2
Rsf2
Rsf2 ΔTDA9
Promoter
K. lac (641)
Vhr1
Vhr2
r=0.86,
new=39%
r=0.41,
new=69%
Vhr1 ΔVHR2
Vhr2 ΔVHR1
Vhr1
Vhr1 ΔVHR2
Vhr2
Vhr2 ΔVHR1
K. lac (895)
Rph1
Gis1
r=0.87,
r=0.71,
new=33%
new=42%
Rph1 ΔGIS1
Gis1 ΔRPH1
Rph1
Rph1 ΔGIS1
Gis1
Gis1 ΔRPH1
-8.5
AA conservation
7
0
Paralog signal
0.8
0.3
Corr
1
E
Promoter binding
Paralog-Ortholog Corr
Paralog-Paralog Corr
Paralog-Paralog Corr
(WT background)
Vhr1
Rph1
Fkh2
Pip2
Rlm1
Ace2
Tda9
Rsf2
Yap1
Hal9
Yap7
Gis1
Fkh1
Gzf3
Yap2
Skn7
Smp1
Swi5
Oaf1
Yap5
Vhr2
Hms2
Tbs1
Dal80
F
sub
neo
Tda9,Rsf2
Rph,Gis1
Vhr1,Vhr2
0
0.1
0.2
0.3
0.4
0.5
Divergence asymmetry
(ΔParalog-Ortholog Corrs)
100
% new
targets
Paralogs divergence

**Fig.4. Classifying paralog divergence through comparison to non-WGD orthologs.**

**(A)** Models of evolutionary paths: Paralogs could diverge by acquiring new preferences (neo-functionalization) or by splitting ancestral preferences (sub-functionalization). **(B)** Sequence evolution: Top: Phylogenetic comparison of non-DBD sequences of the Ixr1/Abf2 paralogs, indicating distances from last common ancestor (LCA). Bottom: differences in sequence evolution rate between paralogs (evolutionary rate asymmetry, $\Delta D_{P1,P2}$, see methods for details). Color indicates rate of the conserved paralog ($D_{P1}$), and grey dots indicate divergence of promoter binding preferences. Paralogs chosen for further analysis are highlighted in bold. **(C-F)** Binding preferences of non-WGD orthologs: Binding profiles of *K. lactis* orthologs were mapped within *S. cerevisiae* and compared to both paralogs (left). Ixr1 binding is tightly conserved between *S. cerevisiae* and *K. lactis* (right). Shown are the correlations of binding preference between the *K. lactis* and *S. cerevisiae* orthologs for specific examples (D) and for all tested paralogs (E; bigger spot size corresponds to the paralog with slower sequence evolution; see also fig. S4). Note the high similarity of binding preferences between each *K. lactis* TF with at least one of its *S. cerevisiae* paralogs, commonly the one experiencing slower sequence divergence. Divergence asymmetry is quantified in (F), with colors indicating the fraction of 'new' targets acquired by the less conserved paralog.

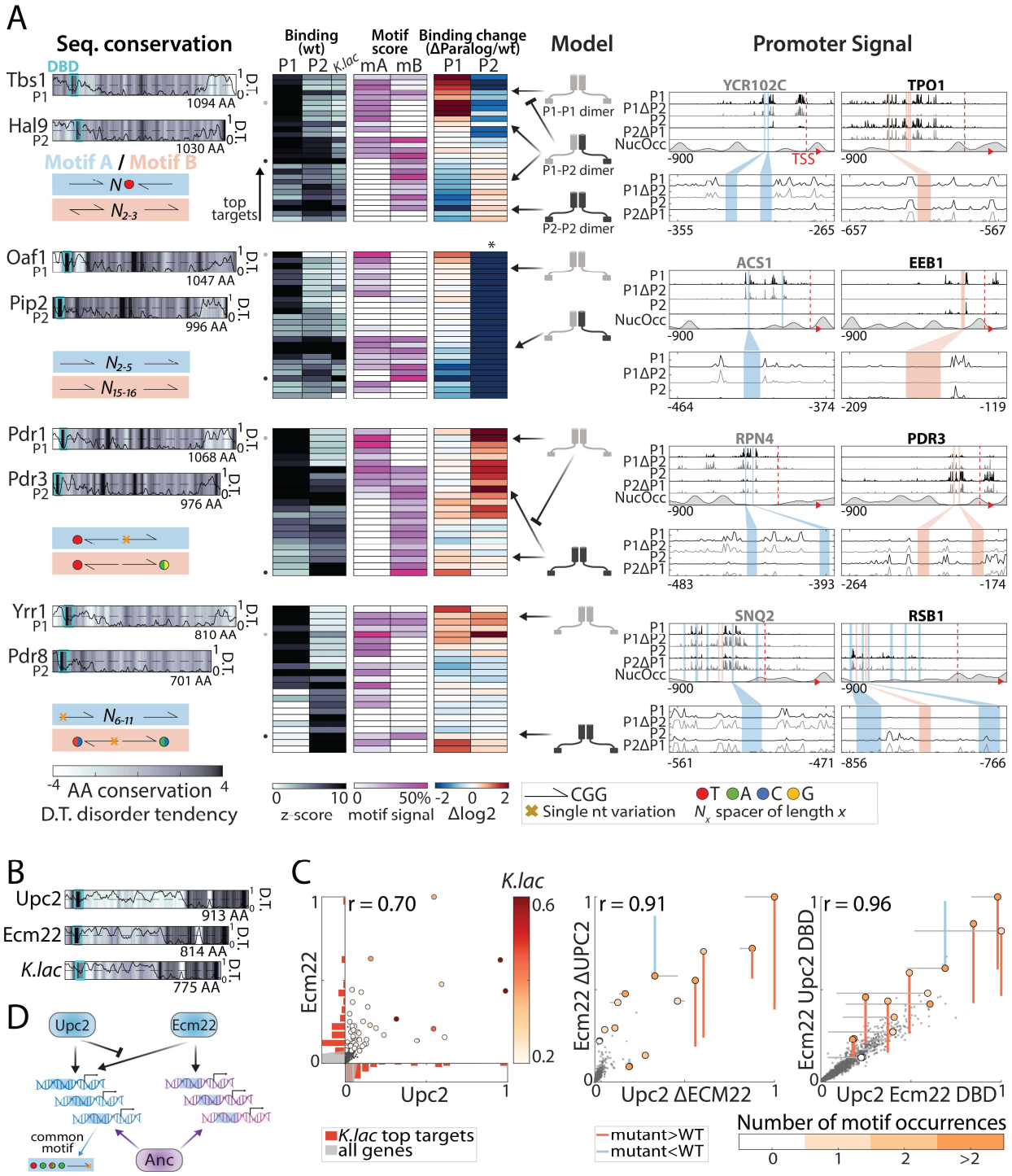
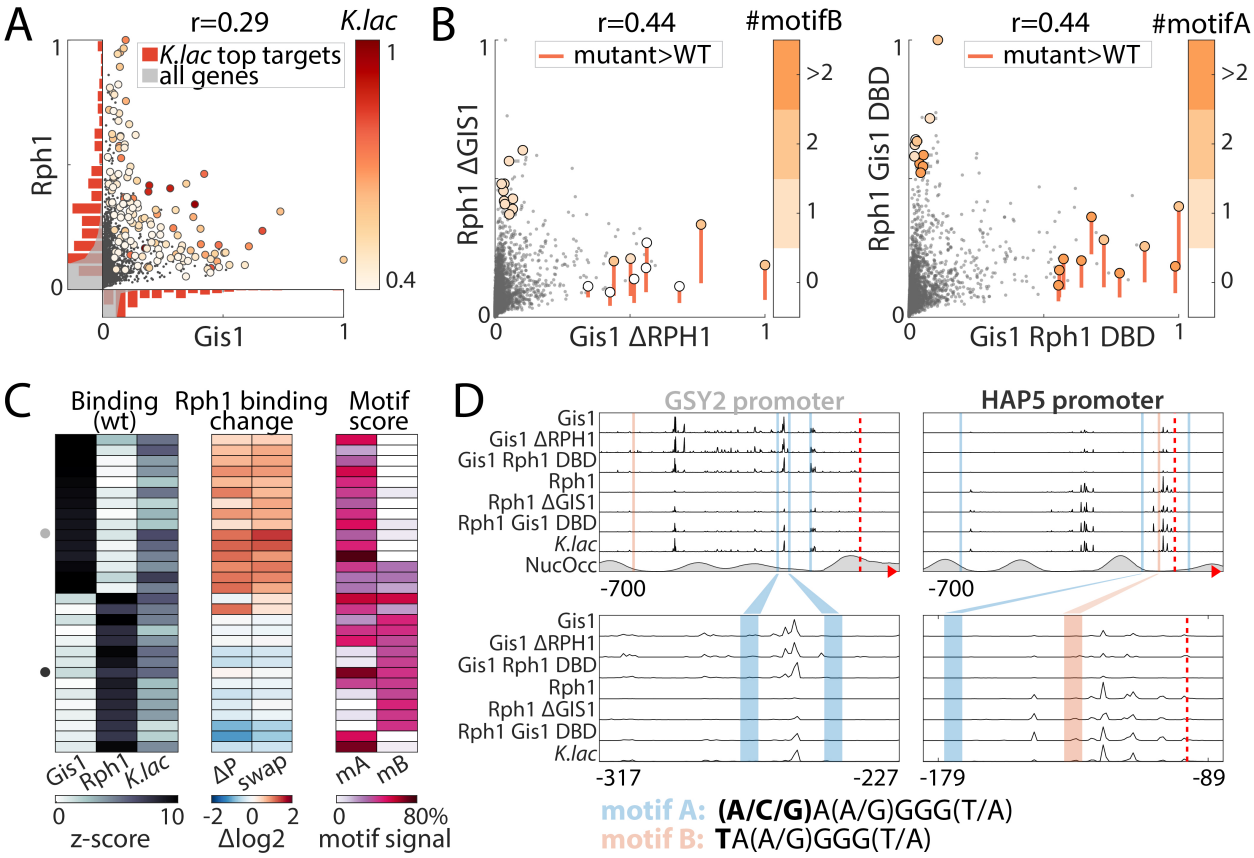
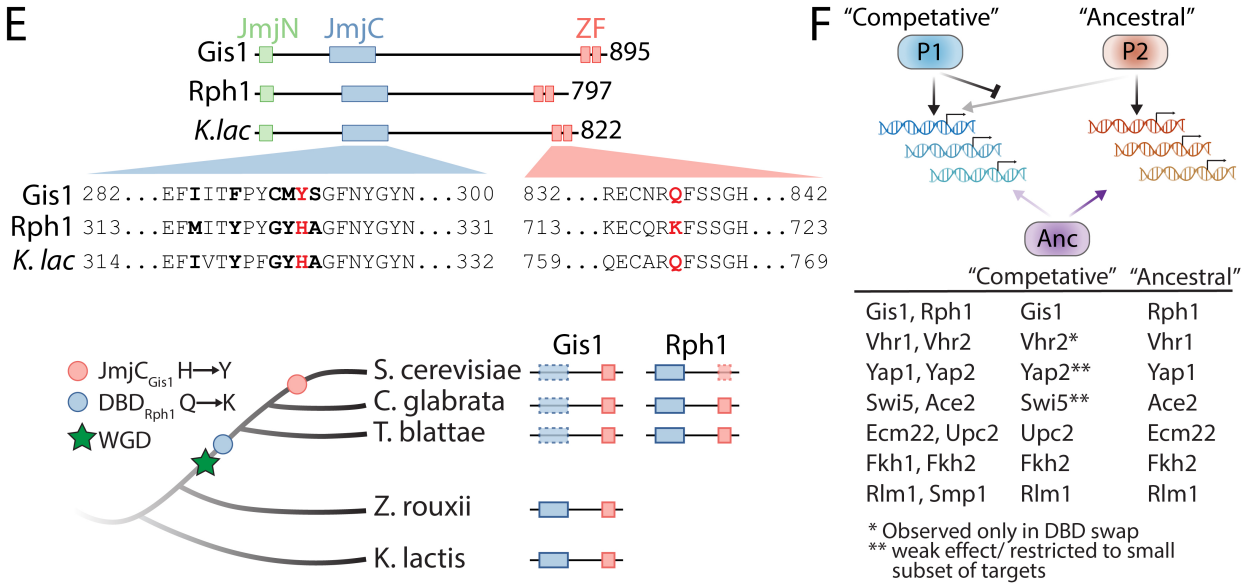

A
Seq. conservation
Binding (wt)
Motif score
Binding change (ΔParalog/wt)
Model
Promoter Signal
DBD
Tbs1
Hal9
Oaf1
Pip2
Pdr1
Pdr3
Yrr1
Pdr8
Motif A / Motif B
top targets
P1-P1 dimer
P1-P2 dimer
P2-P2 dimer
YCR102C
TPO1
ACS1
EEB1
RPN4
PDR3
SNQ2
RSB1
TSS
NucOcc
AA conservation
D.T. disorder tendency
z-score
motif signal
Δlog2
CGG
Single nt variation
spacer of length x
B
Upc2
Ecm22
K.lac
C
r = 0.70
r = 0.91
r = 0.96
Ecm22 ΔUPC2
Upc2 ΔECM22
Ecm22 Upc2 DBD
Upc2 Ecm22 DBD
K.lac top targets
all genes
mutant>WT
mutant<WT
Number of motif occurrences
D
common motif
Anc

**Fig. 5. Divergence of zinc-cluster TF paralogs correlates with changes in motif preferences.**

**(A)** Dimerization and changes in motif-preferences may explain divergence of zinc-cluster paralogs: Zinc-cluster paralogs vary in sequence and localized at different variants of their characteristic motif, shown on the left (AA sequence conservation in color-code, DBD indicated as cyan box, and disorder tendency (*76*) shown as black line; motif symbols indicated on the bottom, see fig. S5 for motif sequences). For each pair, top-bound promoters were selected, and peak-proximal motifs defined. Shown, as indicated, are promoter binding (z-score), percentage of total promoter signal 50 bases around the indicated motifs, and binding change upon paralog deletion (log2). Suggested models explaining divergence and signal on exemplary promoters (indicated by small dots next to the binding panel) are also shown. **(B-D)** Upc2/Ecm22 diverge through DNA-binding competition: shown in (B) is the disorder tendency and pair-wise sequence conservation of Upc2-Ecm22 and the *K. lactis* ortholog with Upc2. Promoter binding preferences in the indicated backgrounds are shown in (C). Large dots indicate top 50 *K. lactis* targets, color-coded by binding strength. Distribution of these targets across the Upc2/Ecm22 binding preference are shown as histograms (red, grey: all genes). Note that Upc2 and Ecm22 bind comparably to strong *K. lactis* targets, while Ecm22 dominates on the low-intermediate targets. Colors in mid/right panels indicate the number of occurrences of the known *in-vitro* motif (TA(T/A)ACGA) and lines show change in binding relative to the wild type. A suggested model summarizing these results is depicted in (D).




| "Competative" | "Ancestral" | |
|---|---|---|
| Gis1, Rph1 | Gis1 | Rph1 |
| Vhr1, Vhr2 | Vhr2* | Vhr1 |
| Yap1, Yap2 | Yap2** | Yap1 |
| Swi5, Ace2 | Swi5** | Ace2 |
| Ecm22, Upc2 | Upc2 | Ecm22 |
| Fkh1, Fkh2 | Fkh2 | Fkh2 |
| Rlm1, Smp1 | Rlm1 | Rlm1 |

* Observed only in DBD swap
** weak effect/ restricted to small subset of targets

**Fig. 6. Resolution of paralog interference through competitive binding.**

**(A-D)** Gis1 limits Rph1 binding through DBD-dependent binding competition: shown are promoter binding preferences of Gis1 and Rph1 in wild type backgrounds (A, as in Fig. 5C) and following mutual paralog-deletion and DBD swapping (B, colored by the number of occurrences of the two known motif variants specified in D). The analysis of all top-bound promoters is summarized in (C) and binding signals on exemplary promoters are shown (D, as described in Fig. 5A). Note the increased binding of Rph1 to Gis1 target promoters upon GIS1 deletion or DBD swapping (e.g. GSY2), and reduced Gis1 binding to its target promoter after DBD swapping (e.g. HAP5). **(E)** Gis1's loss of demethylase activity preceded variation in Rph1's DBD: The conserved JmjC domain providing Rph1 a histone demethylase activity is mutated in Gis1 orthologs of all post-WGD species. The respective DBDs differ in only four positions, one of which replaced a conserved glutamine by lysine specifically in Rph1 and its closest orthologs (fig. S6). This suggests that the divergence was triggered by the loss of demethylase function and DBD-independent acquisition of new targets by Gis1, and finally mutation in Rph1-DBD to reduce residual Rph1 binding interference at the newly acquired Gis1 sites. **(F)** Resolution of paralog interference among diverging TF paralogs: A model for the resolution of paralog interference through competitive binding. The TF inhibiting its paralog's binding is denoted as 'competitive', while the TF whose binding preference better resembles that of the *K. lactis* ortholog is denoted as 'ancestral'. In addition to Gis1/Rph1, other diverging paralogs whose *K. lactis* orthologs were profiled appear to conform to this general model (fig. S6). Note that in most cases (indicated), divergence in promoter binding is driven by variations outside the DBD, with competition only refining, but not driving this divergence of target preferences.