Published in final edited form as:
Science. 2013 May 10; 340(6133): 730–734. doi:10.1126/science.1233465.

Networks of bZIP protein-protein interactions diversified over a billion years of evolution*

Aaron W. Reinke[1], **Jiyeon Baek**[1], **Orr Ashenberg**[1], and **Amy E. Keating**[1,*]

[1]MIT Department of Biology, Cambridge, MA 02139, USA

Abstract

Differences in biomolecular sequence and function underlie dramatic ranges of appearance and behavior among species. We studied the basic region-leucine zipper (bZIP) transcription factors and quantified bZIP dimerization networks for 5 metazoan and 2 single-cell species, measuring interactions *in vitro* for 2,891 protein pairs. Metazoans have a higher proportion of heteromeric bZIP interactions and more network complexity than the single-cell species. The metazoan bZIP interactomes have broadly similar structures, but there has been extensive rewiring of connections compared to the last common ancestor, and each species network is highly distinct. Many metazoan bZIP orthologs and paralogs have strikingly different interaction specificities, and some differences arise from minor sequence changes. Our data show that a shifting landscape of biochemical functions related to signaling and gene expression contributes to species diversity.

Differences in transcriptional regulation between species contribute to developmental and functional outcomes (1). Both changes in *cis* regulatory elements and coding mutations in transcription factors affecting protein-DNA and protein-protein interactions can influence gene regulation (2-5). The basic leucine-zipper (bZIP) proteins, a large class of multi-functional transcription factors, provide an opportunity to study the evolution of biomolecular interactions. bZIPs can be identified in eukaryotic genomes by a basic DNA binding region followed by a leucine-zipper coiled-coil motif. bZIP proteins can form homodimers and heterodimers via the coiled-coil region, and the dimer that forms influences the DNA sites that can be bound (6). Because bZIP proteins interact with other bZIPs, it is possible to compile a comprehensive list of candidate dimerization partners for each protein. Near-complete pair-wise bZIP interactions have been cataloged for human proteins, and many determinants of bZIP dimerization are understood (7, 8). Although bZIP proteins are conserved throughout metazoan evolution, it is unclear if their dimerization preferences are also conserved.

We measured the bZIP protein-protein interaction networks of 5 metazoan species that diverged approximately one billion years ago: human (*Homo sapiens*), sea squirt (*Ciona intestinalis*), fruit fly (*Drosophila melanogaster*), nematode (*Caenorhabditis elegans*), and

*Correspondence to: keating@mit.edu.

sea anemone (*Nematostella vectensis*). We also investigated two single-cell organisms, a choanoflagellate (*Monosiga brevicollis*), which belongs to the closest sister group of metazoans, and the yeast *Saccharomyces cerevisiae*. We defined 21 bZIP families in humans, containing 0 - 4 paralogs each (9-11) (Fig. 1A). 18 families are present in *C. intestinalis*, and 14 can be traced to the last common ancestor of human and sea anemone; we refer to these as the ancestral bZIP families. A smaller number of these 14 families are recognizable in *M. brevicollis* and *S. cerevisiae* (8 and 4, respectively). Each species also has a number of families that lack clear orthologs in the other 6 species; we refer to these as novel families (Fig. 1A, table S1).

We quantified interactions between bZIP proteins *in vitro* using a solution FRET assay (Fig. 1B) (11). Of 53 human bZIP proteins, 36 were selected to cover the observed sequence diversity (8, 11). For the 6 remaining species, all pairwise interactions between all identified bZIPs were measured (figs. S1-S7, table S2). We confirmed that the data were of high quality and reproducible through repeated measurements of interactions and comparisons with previous studies (figs. S8-S12 and table S3)(8, 11). Interactions in each species were observed over a range of affinities, and interactions conserved among five metazoans were stronger than those that were partially conserved or occurred for only one species (Fig. 1C and figs. S1-S7, S13A, table S2). The majority of the proteins in each network (∼50 - 90%) were capable of forming homodimers, and only 5 - 30% of all heterodimer combinations tested were observed to interact. However, because the number of possible heterodimers in each network is greater than the number of possible homodimers, the networks of the 5 metazoan species are composed primarily of heterodimers. For yeast, ∼90% of the observed interactions are homodimers. The choanoflagellate network includes an approximately equal number of homodimers and heterodimers (fig. S13B).

We compared the interaction networks of each pair of metazoan species, considering only those proteins with an ortholog in each species. The overlap ranged from 94% of *C. elegans* interactions occurring in human, to 33% of human interactions occurring in *D. melanogaster* (table S4)(11). Using the numbers of interactions gained or lost between species to calculate rewiring rates, we estimated that $\sim 0.7 - 2.6 \times 10^{-4}$ changes per interaction have occurred per million years (table S5)(11). Comparisons to previously reported evolutionary rates are complicated by differences in methodology, but we observed that changes in bZIP interactions occur faster than most estimated protein-protein interaction changes (12, 13). The sequences of bZIP domains have evolved at an average rate compared to other conserved metazoan proteins. Within the domain, the leucine-zipper regions of the bZIPs have evolved more rapidly and the DNA-binding region is more conserved (fig. S14).

To examine how metazoan interactomes have changed over time, we used parsimony to reconstruct an interaction network for the last common ancestor (11). This network included interactions among proteins in 14 conserved families and contained at least 10 homodimeric and 10 heterodimeric interactions (Fig. 2A); there is ambiguity about additional interactions due to the limited availability of metazoan binding data outside of the 5 species studied.

Using an interaction cutoff of $K_d < 1000$ nM, we tracked the gain and loss of interactions among proteins in 14 conserved families (Fig. 2, Fig. S15, and table S6). Many interactions

were lost in *C. elegans* and *D. melanogaster* (Fig. 2C, D), whereas many new interactions are observed in *C. intestinalis* and human. In the chordates, new interactions were introduced with CI9 (ENSCINP00000010446) in the XBP1 family in *C. intestinalis* and with ATF4 in human (Fig. 2E, F). Although only a few metazoan proteins have identifiable orthologs in choanoflagellates or yeast, several homodimeric interactions in the inferred ancestral network were observed in these pre-metazoan species (e.g. ATF6 and ATF2, figs. S6 and S7).

Each of the metazoans studied here contains bZIPs in families not found in the last common ancestor (Fig. 1A). Four bZIP families originated from gene duplication events involving the ancestral families, e.g. CEBP-CEBPG. Proteins from such pairs often show differences in their interaction profiles (fig. S16). For example, the CEBP protein in flies maintained two of the interactions found in the last common ancestor that were lost by CEBPG (fig. S16 and S17E). Finally, novel families are found in each species that lack orthologs in the other species studied. We observed novel protein interactions with proteins in both novel and conserved families (fig. S1-S7). Together, rewiring of interactions among ancestral proteins, the addition of conserved duplicated families, and the introduction of novel families (table S6) has allowed each species to evolve a highly distinct bZIP interaction network.

To pinpoint differences in the interaction properties of bZIP orthologs and paralogs by analyzing binding to a common set of proteins, we determined the cross-species interaction network of 56 human and *C. intestinalis* bZIPs. This revealed 6 families containing members with moderate-to-highly conserved interaction specificities between human and *C. intestinalis*, e.g. CEBP, and 3 families with specificities that were less than 25% similar between species, e.g. LMAF and BACH (Fig. 3A and fig. S18) (11). Proteins in the ATF4 family have widely varying interaction profiles (Fig. 3A). Human ATF4 has many more interactions than its paralog ATF5. Interactions of *C. intestinalis* ATF4 resemble those of human ATF5, whereas interactions for ATF4 proteins from *Danio rerio* resemble those for human ATF4, indicating the dramatic expansion in ATF4 vs. ATF5 interactions occurred at least ∼350-400 MYA (fig. S19).

There is very weak correlation between bZIP sequence identity and interaction specificity (fig. S20 and S21). Therefore, we investigated certain interaction changes in light of known coiled-coil specificity determinants (9). For example, asparagines at coiled-coil **a** positions are destabilizing when positioned across from hydrophobic amino acids, compared to when they are paired with asparagines (14). *C. elegans* PAR protein CE23 (Zip-7) contains an asparagine at an **a** position and does not interact with CE14 (Atf-2) or CE18 (Cebp-2), whereas other PAR paralogs contain an asparagine at this site and bind these proteins tightly (Fig. 3E). We mutated the asparagine in CE23 to alanine, which is the residue found in PAR protein CE12 (Ces-2), and made the reverse mutation in the CE12 protein. These changes were sufficient to recapitulate the different binding to CE14 and 18 (Fig. 3B). A similar result was observed for PAR family proteins in *D. melanogaster*, on the basis of the same mechanism (Fig. 3C). bZIP interactions can also be destabilized by non-optimal packing of beta-branched residues (e.g. valines or isoleucines) in the core of the coiled-coil interface (15). Human ATF5 has two consecutive coiled-coil **d** position valines, which are leucines in ATF4 (Fig. 3E). We mutated the valines to leucines in ATF5, and made the reverse

mutations in ATF4. This conferred an ATF4-like interaction profile on ATF5, and the ATF4 mutant also became much more ATF5-like (Fig. 3D). These examples highlight the plasticity of the bZIP interactome, which can be dramatically rewired with changes to just one or two amino acids.

We caution that the interactions observed *in vitro* in this study do not necessarily occur *in vivo*. mRNA profiling of human bZIPs shows that most of these proteins are ubiquitously expressed and almost all pairs of bZIPs are co-expressed at measurable levels in at least one tissue (fig. S22) (16), but other factors contribute to whether a bZIP pair will alter gene expression. To begin to investigate the functional consequences of bZIP interaction/non-interaction, we tested DNA binding *in vitro* for bZIP protein-protein interactions that are not conserved among metazoans (ATF2 with FOS, JUN and CEBPG and homodimers of PAR and CEBPG). For all families, loss of protein interaction in a species corresponded to loss of DNA binding (fig. S17). In contrast, protein interactions conserved in all five species (ATF4-CEBPG, FOS-JUN, and ATF2 and PAR homodimers) were functional for DNA binding in each of the species tested (fig. S17). These observations support changes in bZIP protein interactions as a factor in the evolution of gene regulation.

There is considerable interest in using interactions measured in one species to annotate other organisms (17), but our data suggest a conservative approach to inter-species interaction transfer, at least for large paralogous families. Changes in bZIP protein-protein interactions are common, making them a likely contributor to species diversity. This work provides rich information to guide the study of bZIP homo and heterodimer functions, and a resource for investigating the consequences of bZIP network rewiring.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by NIH award GM067681. We thank the MIT BioMicro center for use of TECAN liquid handling robot, J.J. Sims for guidance with intein protein labeling chemistry, L. Reich for analytical models for curve fitting, and A.M. Safer for the *C. elegans* used for making cDNA. All interaction data are available in the supplementary on-line material, in table S2.

References and Notes

1. Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 2008; 134:25–36. [PubMed: 18614008]
2. Lynch VJ, May G, Wagner GP. Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. Nature. 2011; 480:383–386. [PubMed: 22080951]
3. Brayer KJ, Lynch VJ, Wagner GP. Evolution of a derived protein-protein interaction between HoxA11 and Foxo1a in mammals caused by changes in intramolecular regulation. Proc Natl Acad Sci U S A. 2011; 108:E414–20. [PubMed: 21788518]
4. Kuo D, et al. Coevolution within a transcriptional network by compensatory *trans* and *cis* mutations. Genome Res. 2010; 20:1672–8. [PubMed: 20978140]
5. Baker CR, Tuch BB, Johnson AD. Extensive DNA-binding specificity divergence of a conserved transcription regulator. Proc Natl Acad Sci U S A. 2011; 108:7493–8. [PubMed: 21498688]

6. Hai T, Curran T. Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. Proc Natl Acad Sci U S A. 1991; 88:3720–4. [PubMed: 1827203]
7. Grigoryan G, Keating AE. Structure-based Prediction of bZIP Partnering Specificity. Journal of Molecular Biology. 2006; 355:1125–1142. [PubMed: 16359704]
8. Newman JRS, Keating AE. Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science. 2003; 300:2097–2101. [PubMed: 12805554]
9. Vinson C, Acharya A, Taparowsky EJ. Deciphering B-ZIP transcription factor interactions in vitro and in vivo. Biochimica Et Biophysica Acta (BBA) - Gene Structure and Expression. 2006; 1759:4–12.
10. Amoutzias GD, et al. One Billion Years of bZIP Transcription Factor Evolution: Conservation and Change in Dimerization and DNA-Binding Site Specificity. Mol Biol Evol. 2007; 24:827–835. [PubMed: 17194801]
11. Materials and methods are available as supplementary materials on Science Online.
12. Shou C, et al. Measuring the evolutionary rewiring of biological networks. PLoS Comput Biol. 2011; 7:e1001050. [PubMed: 21253555]
13. Qian W, He X, Chan E, Xu H, Zhang J. Measuring the evolutionary rate of protein-protein interaction. Proc Natl Acad Sci U S A. 2011; 108:8725–30. [PubMed: 21555556]
14. Acharya A, Rishi V, Vinson C. Stability of 100 Homo and Heterotypic Coiled-Coil a-a' Pairs for Ten Amino Acids (A, L, I, V, N, K, S, T, E, and R). Biochemistry. 2006; 45:11324–11332. [PubMed: 16981692]
15. Grigoryan G, Reinke AW, Keating AE. Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature. 2009; 458:859–864. [PubMed: 19370028]
16. Ravasi T, et al. An atlas of combinatorial transcriptional regulation in mouse and man. Cell. 2010; 140:744–752. [PubMed: 20211142]
17. Yu H, et al. Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. Genome Res. 2004; 14:1107–18. [PubMed: 15173116]
18. Eddy SR. Profile hidden Markov models. Bioinformatics. 1998; 14:755–763. [PubMed: 9918945]
19. Ellenberger TE, Brandl CJ, Struhl K, Harrison SC. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted [alpha] Helices: Crystal structure of the protein-DNA complex. Cell. 1992; 71:1223–1237. [PubMed: 1473154]
20. Waterhouse RM, Zdobnov EM, Tegenfeldt F, Li J, Kriventseva EV. OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011. Nucleic Acids Res. 2011; 39:D283–8. [PubMed: 20972218]
21. Ostlund G, et al. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. Nucleic Acids Res. 2010; 38:D196–203. [PubMed: 19892828]
22. Powell S, et al. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. Nucleic Acids Res. 2012; 40:D284–D289. [PubMed: 22096231]
23. Chen F, Mackey AJ, Stoeckert CJ Jr, Roos DS. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 2006; 34:D363–8. [PubMed: 16381887]
24. Hoover DM, Lubkowski J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucl Acids Res. 2002; 30:e43. [PubMed: 12000848]
25. Reinke AW, Grigoryan G, Keating AE. Identification of bZIP interaction partners of viral proteins HBZ, MEQ, BZLF1, and K-bZIP using coiled-coil arrays. Biochemistry. 2010; 49:1985–1997. [PubMed: 20102225]
26. Li MZ, Elledge SJ. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods. 2007; 4:251–256. [PubMed: 17293868]
27. Ashenberg O, Rozen-Gagnon K, Laub MT, Keating AE. Determinants of homodimerization specificity in histidine kinases. J Mol Biol. 2011; 413:222–35. [PubMed: 21854787]
28. Hedges SB, Dudley J, Kumar S. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics. 2006; 22:2971–2972. [PubMed: 17021158]
29. Luz H, Vingron M. Family specific rates of protein evolution. Bioinformatics. 2006; 22:1166–1171. [PubMed: 16510497]

30. Li H, et al. TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic Acids Res. 2006; 34:D572–80. [PubMed: 16381935]
31. Fu X, Apgar JR, Keating AE. Modeling backbone flexibility to achieve sequence diversity: the design of novel alpha-helical ligands for Bcl-xL. J Mol Biol. 2007; 371:1099–1117. [PubMed: 17597151]
32. Dragan AI, et al. Thermodynamic signature of GCN4-bZIP binding to DNA indicates the role of water in discriminating between the AP-1 and ATF/CREB sites. J Mol Biol. 2004; 343:865–878. [PubMed: 15476806]
33. Wendt H, Baici A, Bosshard HR. Mechanism of assembly of a leucine zipper domain. J Am Chem Soc. 1994; 116:6973–6974.
34. Bosshard HR, Durr E, Hitz T, Jelesarov I. Energetics of coiled coil folding: the nature of the transition states. Biochemistry. 2001; 40:3544–3552. [PubMed: 11297420]
35. Zitzewitz JA, Bilsel O, Luo J, Jones BE, Matthews CR. Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. Biochemistry. 1995; 34:12812–12819. [PubMed: 7548036]
36. Moll JR, Olive M, Vinson C. Attractive interhelical electrostatic interactions in the proline- and acidic-rich region (PAR) leucine zipper subfamily preclude heterodimerization with other basic leucine zipper subfamilies. J Biol Chem. 2000; 275:34826–34832. [PubMed: 10942764]
37. Krylov D, Olive M, Vinson C. Extending dimerization interfaces: the bZIP basic region can form a coiled coil. EMBO J. 1995; 14:5329–5337. [PubMed: 7489722]
38. Ahn S, et al. A dominant-negative inhibitor of CREB reveals that it is a general mediator of stimulus-dependent transcription of c-fos. Mol Cell Biol. 1998; 18:967–977. [PubMed: 9447994]
39. Olive M, et al. A dominant negative to activation protein-1 (AP1) that abolishes DNA binding and inhibits oncogenesis. J Biol Chem. 1997; 272:18586–18594. [PubMed: 9228025]
40. Kohler JJ, Schepartz A. Kinetic Studies of Fos.Jun.DNA Complex Formation: DNA Binding Prior to Dimerization. Biochemistry. 2001; 40:130–142. [PubMed: 11141063]
41. Carrillo RJ, Dragan AI, Privalov PL. Stability and DNA-binding ability of the bZIP dimers formed by the ATF-2 and c-Jun transcription factors. J Mol Biol. 2010; 396:431–440. [PubMed: 19944700]
42. Patel LR, Curran T, Kerppola TK. Energy transfer analysis of Fos-Jun dimerization and DNA binding. Proc Natl Acad Sci U S A. 1994; 91:7360–7364. [PubMed: 8041795]
43. Seldeen KL, McDonald CB, Deegan BJ, Farooq A. Thermodynamic analysis of the heterodimerization of leucine zippers of Jun and Fos transcription factors. Biochem Biophys Res Commun. 2008; 375:634–638. [PubMed: 18725194]
44. Pernelle C, Clerc FF, Dureuil C, Bracco L, Tocque B. An efficient screening assay for the rapid and precise determination of affinities between leucine zipper domains. Biochemistry (N Y). 1993; 32:11682–11687.
45. Acharya A, Rishi V, Moll J, Vinson C. Experimental identification of homodimerizing B-ZIP families in Homo sapiens. Journal of Structural Biology. 2006; 155:130–139. [PubMed: 16725346]
46. O'Shea EK, Rutkowski R, Stafford WF 3rd, Kim PS. Preferential heterodimer formation by isolated leucine zippers from fos and jun. Science. 1989; 245:646–648. [PubMed: 2503872]
47. Schneider TL, Schepartz A. Hepatitis B Virus Protein pX Enhances the Monomer Assembly Pathway of bZIP·DNA Complexes. Biochemistry. 2001; 40:2835–2843. [PubMed: 11258894]
48. Kohler JJ, Metallo SJ, Schneider TL, Schepartz A. DNA specificity enhanced by sequential binding of protein monomers. Proc Natl Acad Sci U S A. 1999; 96:11735–11739. [PubMed: 10518519]

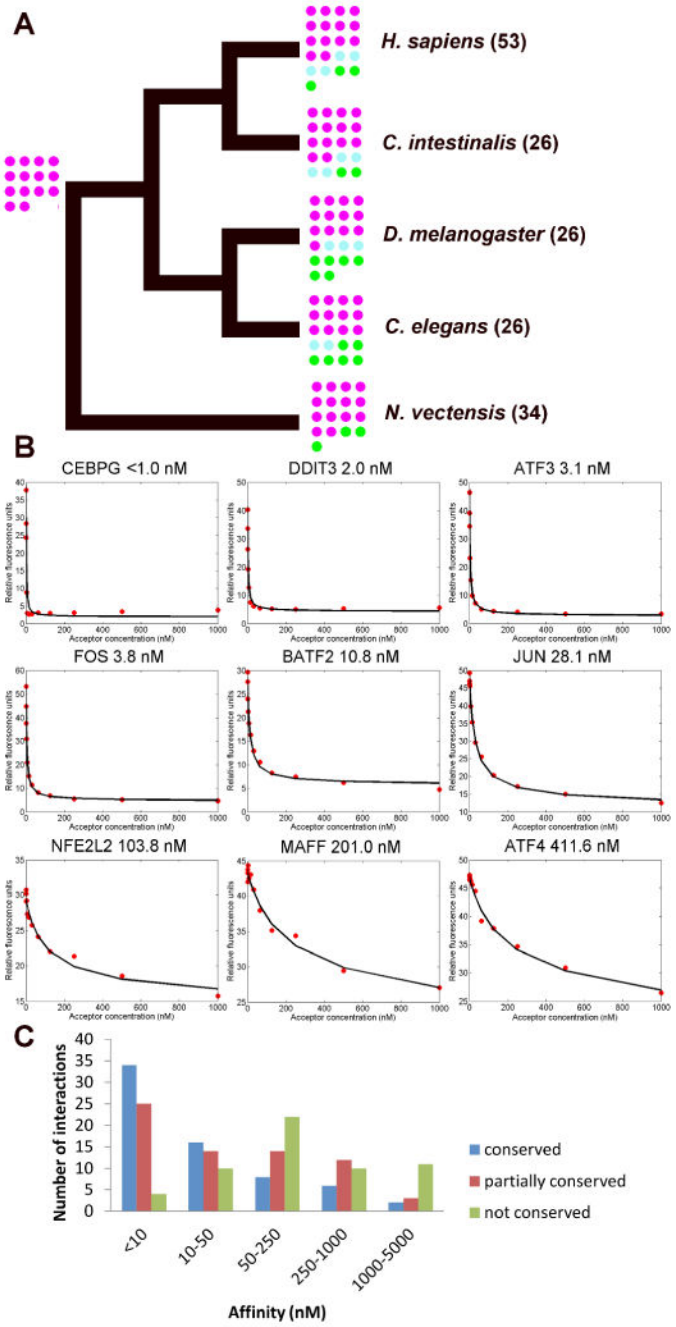

Fig. 1.

Quantitation of bZIP protein-protein interactions in 5 metazoan species. (**A**) Evolutionary tree for the 5 metazoan species studied; branch lengths are not to scale. Colored circles represent bZIP families: families in the last common ancestor of metazoans, magenta; families in two or more species, cyan; species-specific families, green. In parentheses is the total number of bZIPs in each organism. (**B**) Binding curves for interactions involving TAMRA-labeled ATF4 measured at 21° C, with K_d values indicated. (**C**) Relationship between conservation and interaction affinity (11).

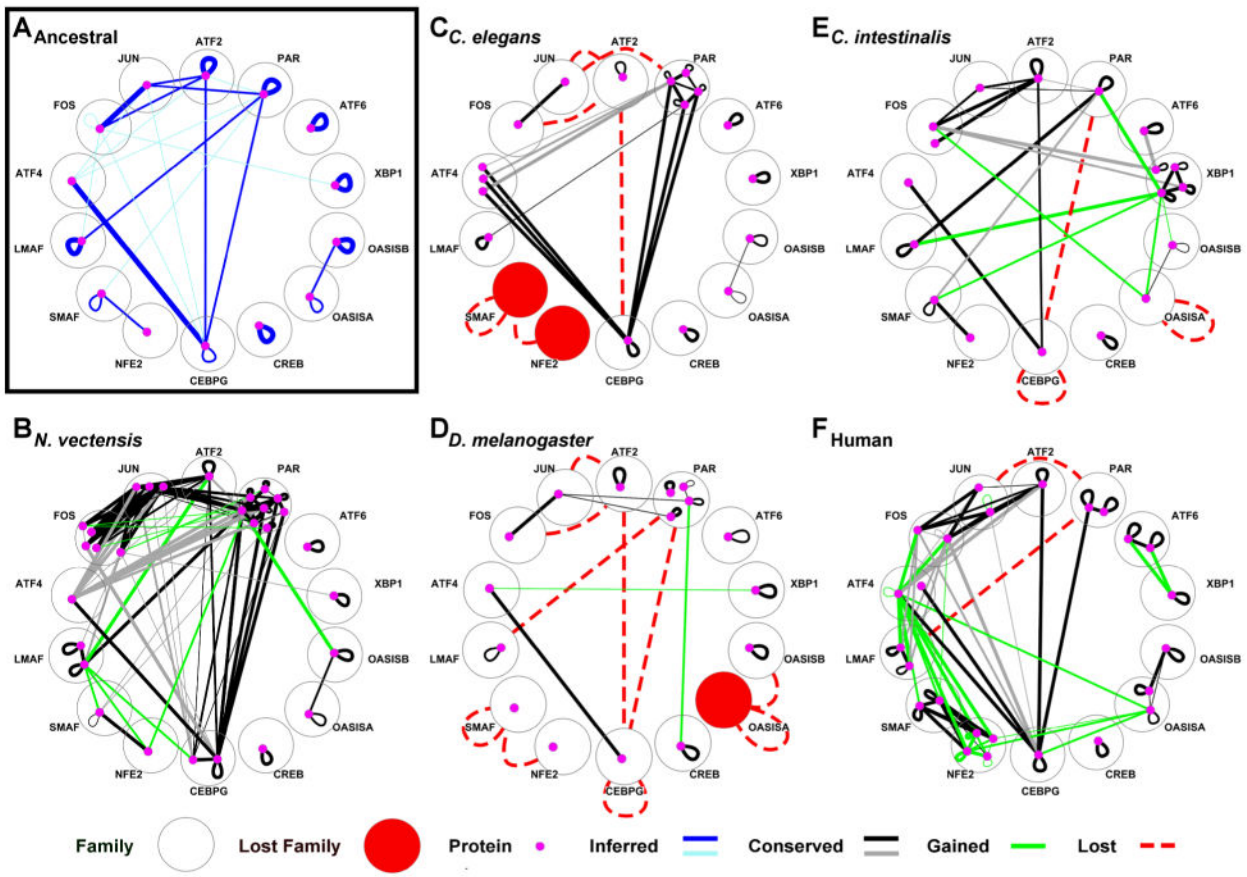

Fig. 2.

Changes in interactions between bZIPs in conserved families. Proteins (magenta nodes) are grouped into families (large circles). Edges in the graph represent interactions. In the inferred ancestral network (**A**), thick dark blue edges show interactions observed in 5 extant metazoans, narrow dark blue edges show interactions, light blue edges indicate interaction in the ancestor is ambiguous (11). In extant species (**B-F**), green edges show interactions gained compared to the last common ancestor, red circles highlight lost families and red dashed lines indicate lost interactions. Black interactions are conserved from the ancestor, and grey interactions may be conserved (status in the last common ancestor is ambiguous). For the extant species, three line thicknesses (widest to narrowest) indicate K_d <50 nM, 50 < K_d < 250 nM and 250 < K_d < 1000 nM. Graphs created using Cytoscape (http://www.cytoscape.org/).

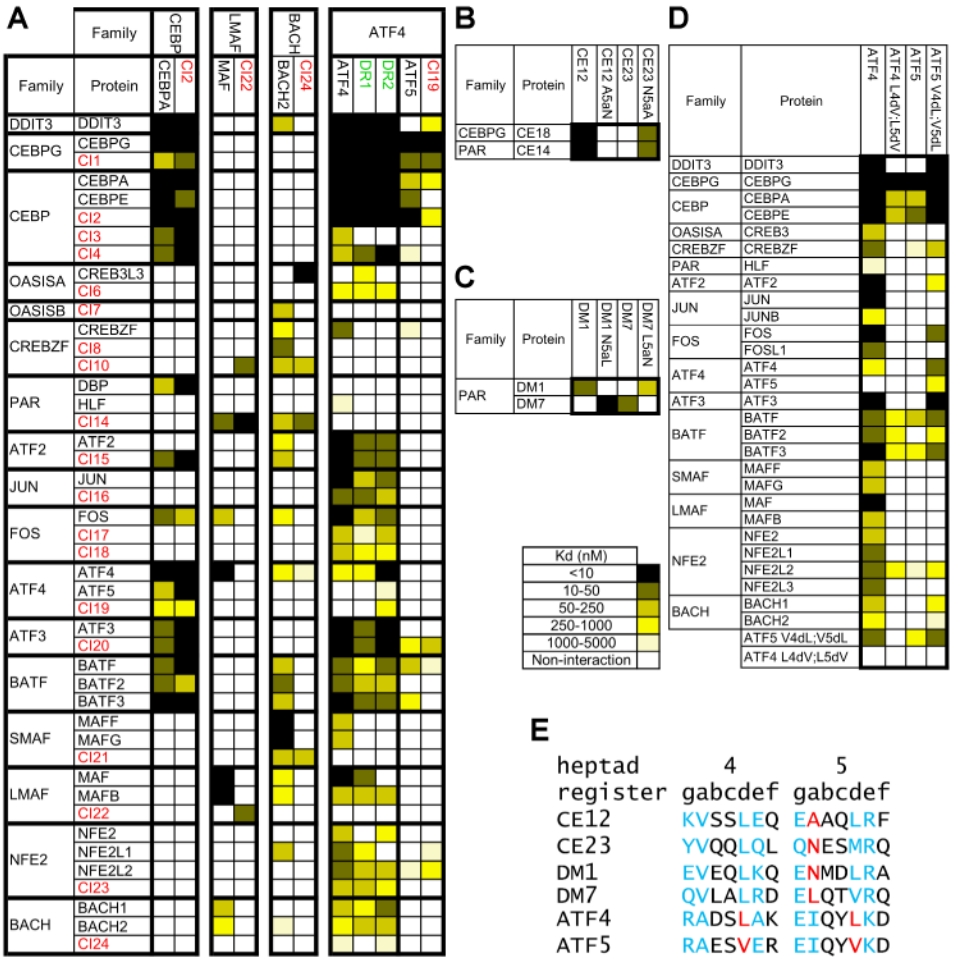

Fig 3.

Changes in the interaction specificities of bZIP proteins. **(A-D)** Binding affinities are presented as heat maps using the scale at bottom. (**A**) Interaction profiles for human and *C. intestinalis* orthologs are in columns, highlighting similar specificities (CEBP), diverged specificities (LMAF, BACH), and the differences among ATF4 family paralogs. Human proteins are black and *C. intestinalis* red; two *D. rerio* ATF4 paralogs (right column) are green. (**B-D**) Switching interaction profiles with mutations. Mutants are named by giving the wild-type residue, the coiled-coil heptad number and position as shown in panel **E**, then the mutant residue. (**B**) Point mutations in *C. elegans* PAR family paralogs CE12 and CE23 switch the interactions of these proteins with CE14 and CE18. (**C**) Point mutations in *D. melanogaster* proteins DM1 (Pdp1) and DM7 (CG7786) switch the homo vs. heterodimerization of these proteins. (**D**) Two mutations in human ATF4 or ATF5 change the interaction profiles of these proteins to resemble one another. (**E**) Sequences of PAR family proteins in *C. elegans* and *D. melanogaster* and ATF4 family proteins in human, highlighting specificity changing mutations. Interface positions are blue and mutated residues are red.