Networks of bZIP Protein-Protein Interactions Diversified Over a Billion Years of Evolution

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1126/science.1233465">http://dx.doi.org/10.1126/science.1233465</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>American Association for the Advancement of Science (AAAS)</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sun May 06 08:01:34 EDT 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/85989">http://hdl.handle.net/1721.1/85989</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Networks of bZIP protein-protein interactions diversified over a billion years of evolution

Authors: Aaron W. Reinke,1 Jiyeon Baek,1 Orr Ashenberg,1 Amy E. Keating1*

Affiliations:
1 MIT Department of Biology, Cambridge, MA 02139, USA.
*Correspondence to: keating@mit.edu

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 re-wiring 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.

Main Text:
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 pairwise 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 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 ~0.7 - 2.6 x 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 C19 (ENSCIP0000010446) in the XB1 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-CEBP, 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.

References and Notes:


11. Materials and methods are available as supplementary materials on Science Online.


32. A. I. Dragan 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.* **343**, 865-878 (2004).


Acknowledgements:

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.
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.