ANCESTRAL COMPLEXITY AND LINEAGE SPECIFIC EXPANSIONS OF THE ANIMAL BZIP INTERACTOME

by

Eric Gregory Kane

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science
Biological Sciences

Charlotte

2017

Approved by:

_____
Dr. Adam Reitzel

_____
Dr. Matthew Parrow

_____
Dr. Kenneth Pillar

ABSTRACT

ERIC GREGORY KANE. Ancestral complexity and lineage-specific expansions of the animal bZIP interactome.
(Under the direction of DR. ADAM REITZEL)

Protein-protein interactions are central to the regulation and function of transcription factors in time- and cell-specific regulation of the genome. Increases in interactome complexity may serve as a mechanistic driver for evolutionary novelty. This novelty could arise as temporal and spatial gene expression becomes differentially regulated. Here, we used a combinatory phylogenetic and algorithmic approach to characterize the diversification of interactomes for the bZIP family across 18 metazoans and three unicellular outgroups. Our bZIP gene tree analyses identified two new subfamilies (CREBL2 and CREBZF) across animals and reinforce previous bZIP family classifications, indicating that diversification occurred in two waves, prior to the animal ancestor (12 of 18 subfamilies) and then again after the cnidarian-bilaterian ancestor (6 of 18 subfamilies). Expansions and losses of bZIP subfamilies with different dimerization capacities drive substantial variations in interactomes of ctenophores, placozoans, cnidarians, and vertebrates. While these expansions and losses result in a large majority of the observed rewiring of the metazoan interactome, variability of per transcription factor connectedness itself was also a factor, with total dimers shifting by two-fold while bZIP count remains consistent between invertebrate bilaterians. Heterodimeric potential for particular bZIP orthologs spanning more than 1 billion years, suggesting an evolutionary constraint on protein-protein interactions. Collectively, the interactome connectedness for bZIP transcription factors has undergone large and uneven changes across animal evolution driven principally by gene duplication and loss, and changes in the interaction behavior of the proteins themselves plays a less significant but impactful role.

ACKNOWLEDGMENTS

I would like to acknowledge Dr. Adam Reitzel for contributing significant assistance in interpretation of data. Also, I would like to acknowledge Dr. Jason Macrander for assisting with data analysis. I would like to acknowledge the Binational Science Foundation and National Science Foundation for providing funding through Dr. Adam Reitzel.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

bZIP	Basic Leucine Zipper
GRN	Genetic Regulatory Network
MUSCLE	Multiple Sequence Comparison by Log-Expectation
LG+G	Le and Gascuel Substitution Matrix + Gamma
FRET	Fluorescence Resonance Energy Transfer
K_d	Dissociation Constant
BLAST	Basic Local Alignment Search Tool

INTRODUCTION

Molecular mechanisms underlying phenotypic diversification across metazoans remain a primary focus in evolutionary and developmental biology. Early hypotheses regarding metazoan diversification suggested that organisms of seemingly high complexity – those with bilateral symmetry, a diverse array of cell types, three tissue layers, and perhaps organ systems – possess complex genomes containing by a large number of genes (Claverie 2001; Kusserow, et al. 2005; Miller, et al. 2005; Rivera, et al. 2010; Taylor and Raes 2004). Over the past decade, the addition of a number of metazoan genomes has provided a wealth of comparative data that did not support a direct relationship between genotypic diversity and perceived phenotypic complexity (Putnam, et al. 2007; Srivastava, et al. 2008; Srivastava, et al. 2010) . The combination of conserved gene families and extensive gene loss throughout Metazoa, and more specifically Bilateria, suggests that gene content was not the primary factor of phenotypic diversification among metazoans and alternative mechanisms are responsible.

The development of multicellular organisms is determined through gene regulatory networks (GRNs) and the expansion and diversification of these molecular pathways may drive phenotypic evolution (Davidson and Erwin 2006). GRNs are modular networks of interacting transcription factors and cis-regulatory modules that regulate metazoan development. GRNs in development appear to balance deep conservation and flexibility. For example, kernels are core GRN circuitry that remains evolutionarily constrained over millions of years (Davidson and Erwin 2006), while other GRN components appear to be “flexible” with regards to signaling pathways or transcription factors (Rebeiz, et al. 2015). Increases in transcription factor number may

have altered GRN behavior and influenced the evolution of multicellular GRN circuitry, making it possible to evolve a developmental regimen that functioned separately from the strict environmental response GRN circuitry of unicellular organisms (Sebé-Pedrós and de Mendoza 2015). Transcription factors within GRNs may also change cis-regulatory DNA binding preference (Cheatle Jarvela, et al. 2014; Nakagawa, et al. 2013), modulate cofactor interactions (Lynch, et al. 2011; Miller, et al. 2003), and shift dimerization activity within transcription factor subfamilies (Cheatle Jarvela and Hinman 2015; Voordeckers, et al. 2015). These changes in the behavior of a particular transcription factor are in part dependent on the extent of specificity for protein-protein interactions and how these interactions change over time. Therefore, it is crucial to understand changes in the protein-protein interactions of transcription factors to uncover potential mechanisms for these shifting interactions over evolutionary time.

The explanatory power of changes in protein-protein interactions driving patterns of phenotype evolution has been difficult to assess in early branching metazoans. While the genomes of many metazoans have been sequenced and analyzed, annotation of protein-protein networks are typically constrained to *Caenorhabditis elegans*, *Drosophila melanogaster* and *Homo sapiens*. (Hart-Smith, et al. 2016; Yu, et al. 2011). The topologies of interactomes – which may change as quickly as genes themselves (Mosca, et al. 2012; Vo, et al. 2016) – influences the evolution of genes by constraining proteins with high interaction rates and increasing rates of evolution in others by limiting pleiotropy via compensatory networks (Luisi, et al. 2015). This mechanism can have profound effects on developmental outcomes (Mozgova and Hennig 2015; Sayou, et al. 2014). Annotating interactomes of non-model species is a next step toward understanding

molecular evolution and its contributions to phenotypic variation. Further, a broader taxonomic sampling of interactomic data could improve the quality of interactome prediction (Andreani and Guerois 2014) by reduced prediction noise (Hamer, et al. 2010), and elucidating mechanisms involved with constraining the evolution of interaction potential (Elcock and McCammon 2001).

A model transcription factor dimerization network is the basic leucine zipper (bZIP) family. bZIP proteins are highly conserved eukaryotic transcription factors that regulate a variety of central cellular and tissue-level functions (Dibner and Schibler 2015; Guo, et al. 2011; Zhang and Kaufman 2008) during phylotypic stages of development (Levin, et al. 2016). The phylotypic activity of bZIPs suggests that the superfamily plays a role in animal evolutionary development. However, there is no evidence that bZIP function is constrained within GRN kernels, suggesting that this family of transcription factors has a more flexible role within the animal. Many bZIPs are involved in highly conserved processes, including the role of XBP1 in the unfolded protein response that most likely traces back to an Opisthokont ancestor (Howell 2013). However, bZIPs are also pleiotropic, for example XBP1 also has a specific role in the regulation of brain development in mammals and *Drosophila* (Hayashi, et al. 2007; Sone, et al. 2013). The bZIP proteins are targeted by pathways commonly defined as plug-ins, suggesting that bZIP-related pathways are modular and can be rewired into diverse GRN circuits. Within Bilateria, bZIPs have been shown to perform a variety of functions, including immune cell and erythrocyte differentiation (Chan, et al. 1998; Gasiorek and Blank 2015), endomesoderm specification (An and Blackwell 2003), mandibular development (Veraksa, et al. 2000), regulation of circadian rhythm (Gachon, et al. 2006; Reitzel, et al.

2013), and regulation of imaginal disc formation through decapentaplegic (George and Terracol 1997).

bZIP proteins share two conserved domains: 1) the DNA-binding basic region, a highly conserved 35-40 amino acid sequence, and 2) the leucine zipper, a less conserved region comprised of heptad repeats that regulate bZIP dimerization activity. The leucine heptad repeats, defined as [**abcdefg**]$_n$, are alpha-helices where **a** and **d** represent residues buried within the hydrophobic core of the coiled-coil, **e** and **g** are generally exposed acidic or basic residues and **b** and **c** represent the hydrophilic backbone. The **d** position is often dominated by leucine residues, while other positions tend to vary, (Amoutzias, et al. 2007; Grigoryan and Keating 2006; Potapov, et al. 2015) bZIPs are able to homo- and heterodimerize through hydrophobic exclusion (commonly contributed by the residues at the **a** and **d** positions) and favorable electrostatic interactions (usually contributed by the residues at the **e** and **g** positions) (Acharya, et al. 2006; Grigoryan and Keating 2006; Thompson, et al. 1993). bZIP dimerization predictions are bolstered by empirical biochemical studies that have queried bZIP interactions in a variety of non-model animals that span a billion years of evolution (Reinke, et al. 2013). The simple coiled-coil structure and empirical confirmation have led to the development of high fidelity predictive algorithms that can define bZIP coiled-coil interactions with up to 90% accuracy (Grigoryan, et al. 2009). As bZIP gene families have gone through events of duplication or loss, dimerization rates vary between species (Deppmann, et al. 2006), changing bZIP interactomes and potentially changing bZIP-target gene expression (Lynch, et al. 2009).

Previous studies of metazoan bZIP evolution have identified deep conservation of subfamilies throughout the animal tree. The first bZIP gene tree that included animals outside the bilaterian lineage identified 19 bZIP subfamilies. Of these 19, 6 (CEBPg, OASISb, L-MAF, BATF, ATF3, BACH) were bilaterian-specific duplication events, with all other subfamilies originating before the emergence of Bilateria (Amoutzias, et al. 2007). The addition of genomes from closely related unicellular organisms- including choanoflagellates and the filasterian *Capsaspora owczarzaki* (King, et al. 2008; Sebé-Pedrós, et al. 2011)- resulted in a recent revisit to the bZIP gene tree, which showed that a majority of bZIP subfamilies predated the animal ancestor (Jindrich and Degnan 2016). A central question that follows these phylogenetic studies is how the diversity of genes then relates to protein-protein interactions that are critical for how bZIPs regulate the genome. Investigations into the diversification of bZIP interactomes have been taxonomically limited in animals. A broad phylogenetic approach would elucidate patterns for how networks of protein-protein interaction change over long evolutionary patterns and if these changes relate to the emergence of animal complexity. In this study, we have identified metazoan bZIP genes, characterized diversity across subfamilies, and evaluated the interactomes from 15 metazoans and closely related unicellular non-metazoan lineages to characterize changes in transcription factor content and dimerization patterns in a broad phylogenetic context. We identify multiple instances of expansion and loss of particular bZIP subfamilies across different specific lineages and determine the interactome of each species. Our analysis identified high levels of connectivity in a unicellular outgroup and in deep animal evolution, suggesting that animal lineages have rewired a well conserved ancestral interactome topology (transcription factor count and

protein-protein behavior) through expansions of different bZIP subfamilies, which are not necessarily correlated with our perceived notions of phenotypic complexity.

MATERIALS AND METHODS

Domain Identification

We identified candidate bZIP proteins using the bZIP_1, bZIP_2, and bZIP_MAF domain search functions in the Sanger Institute Pfam database (Finn, et al. 2014), and bZIP domains were confirmed with reciprocal searches using BLASTp against the non-redundant protein database in NCBI. Depending on the species, we also used BLASTp to query bZIP domains not identified with Pfam, or identify partial bZIP proteins in these first generation genome annotations at species-specific databases (Supplemental Table 1). Amino acid sequences for the candidate bZIP domains were aligned using MUSCLE (version 3.8) (Edgar 2004), visually inspected in GeneDoc (Nicholas, et al. 1997), and trimmed to include only the conserved bZIP domain (~60 amino acids) (Supplemental File 1). Sequences were removed from the data set if 4 or more amino acids in the highly conserved section of the basic region (residues 9, 12, 13, 16, 17, 20, 21, 23, 24, and 25 as defined by (Jindrich and Degnan 2016) were not present. Many residues in this region are diagnostically conserved and predicted to be essential for interactions with the DNA phosphate backbone and direct interaction with the Watson-Crick bases. Of the 414 bZIP candidates, 22 were deleted due to these criteria (Supplemental Table 2).

Gene Tree Reconstruction

We used a maximum likelihood approach to determine evolutionary relationships of bZIPs in Metazoa, with closely related unicellular species. The single bZIP from *Giardia lamblia* was used as an outgroup. bZIP gene tree reconstructions were completed using RaxML-HPC2 on XSEDE (version 8.1.24) (Stamatakis 2006) with 1000 bootstraps through the CIPRES portal (Miller, et al. 2010) following a LG+G substitution matrix (as

determined by ProTest (version 3.4) (Darriba, et al. 2011). In parallel, we reconstructed a bZIP gene tree using neighbor-joining methods to compare tree topology with the LG+G maximum likelihood tree (MEGA version 6) (Tamura, et al. 2013). Trees were visualized using the International Tree of Life website (Letunic and Bork 2007). All protein accession numbers, IDs, and the alignments are provided in Supplemental Table 3.

Genes were categorized as members of monophyletic groups using two criteria. First, we enumerated independently resolved groups of sequences when the basal node was ≥40% bootstrap support, hereon referred to as bZIP subfamilies. Second, we searched for conserved bZIP residues and non-bZIP domains shared by members of resolved groups. This strategy was employed for all queried bZIPs, but was especially important for three bZIP subfamilies that were difficult to resolve with bootstrap support.The FOS subfamilycontains acidic repeats N-terminal to the basic region. Proteins in the CEBP subfamily share a unique tyrosine residue within the basic region as well as a diagnostic asparagine pair. The PAR-L subfamily also has the diagnostic asparagine pair (Amoutzias et al. 2007, Sebe-Pedros, et al. 2012).

Dimerization Prediction

Due to variation in length of the heptad repeats, leucine zippers were manually trimmed and confirmed with the 2ZIP leucine zipper prediction server for dimerization prediction. Dimerization between bZIP proteins can be predicted with high efficiency using computation models (Grigoryan and Keating 2006). We used two methods to predict bZIP coiled-coil interactions: a model developed by (Fong, et al. 2004) trained in coupling energies and coiled-coil arrays (*http://compbio.cs.princeton.edu/bzip*) and a second model proposed by (Potapov, et al. 2015) trained on FRET data points spanning

choanoflagellates, Cnidaria, Nematoda, Arthropoda, and Chordata. Both algorithms use optimized electrostatic weights of pairs of interacting amino acids within the interaction interface of the bZIP coiled-coil and were trained on a dataset including thousands of FRET data points. However, the Potapov algorithm also weighs interactions between triplets of amino acids within the interaction interface and was trained on a larger dataset.

bZIP sequences were compared in a pairwise manner across species, with the model proposed by (Fong, et al. 2004) using an interaction cutoff value of 27, instead of a value of 30. According to (Fong, et al. 2004) a cutoff of 30 produces approximately a 7% false negative frequency. However, we attempted to reduce false negatives in our preliminary analysis by manually inspecting interaction cutoffs ranging between 27 and 30. Relatively lower scores are known to accurately represent conserved interactions, especially those involving the bZIP subfamilies JUN and FOS as these interactions occur at a score closer to 27 when compared with other subfamilies (Fong, et al. 2004). The Fong algorithm also fails to identify 90% of human NFE2 and MAF interactions. We complimented analyses using the Fong algorithm with the algorithm by (Potapov, et al. 2015) The algorithm developed by Potapov et al. (2015) with an interaction cutoff value of -6, or log (.001), represents a K_d of 1000 nM. Previous studies used 1000 nM as a cutoff value (Reinke, et al. 2013; Reinke, et al. 2010), and a majority of inferred interactions are under this simulated K_d value. Values for the calculated interactome were represented in a heatmap to summarize these interactions.

RESULTS

Gene Tree Analysis

In total, our searches for bZIP genes identified 414 total candidates, 118 bZIP genes were identified using BLASTp in NCBI sequences, 38 from the Joint Genome Institute, 118 from the UCSC Genome Browser, 28 from Metazome, and 13 sequences from the *Mnemiopsis* Genome Portal (Supplemental Table 3). Previous work identified subfamilies through trained HMMs (Amoutzias, et al. 2007) or through gene tree reconstructions of a single species bZIP repertoire against a subfamily reference sequence (Jindrich and Degnan 2016). In most cases, our bootstrap cutoff assessment was in complete agreement with these methods. Rare differences, including counts of PAR/PAR-L genes, are present between the datasets (Supplemental Table 4). We report that a bZIP repertoire consisting of at least 8 canonical bZIP subfamilies was present in *Capsaspora owczarzaki* and *Monosiga brevicollis*. These subfamilies include ATF4/5 (Sebé-Pedrós, et al. 2011) CREB1/ATF1, ATF2/7, ATF6, OASIS, PAR-L, C/EBP (Jindrich and Degnan 2016; Sebé-Pedrós, et al. 2011) and JUN (Jindrich and Degnan 2016).

We identified two previously unreported bZIP subfamilies in our study. CREBZF, or Zhangfei, and CREBL2 are omitted from recent bZIP gene tree reconstructions (Amoutzias, et al. 2007; Jindrich and Degnan 2016). However, these subfamilies have been functionally characterized in vertebrates (Ma, et al. 2011; Zhang, et al. 2013), and contain the holozoan bZIP consensus sequence (**bb**x**b**NxxAAxxxRx**bbb)**, where **b** is a basic residue and x is a variable site, as previously defined (Jindrich and Degnan 2016). Both subfamilies resolve independently in the bZIP gene tree with modest bootstrap

support, 44% support for CREBZF and 62% for CREBL2 (Figure 1). All identified CREBZF orthologs possess an extended leucine zipper (Supplemental Table 1). The CREBL2 family, with the exception of the *Capsaspora* ortholog, shares a conserved asparagine to serine substitution at the diagnostic fifth residue of the holozoan bZIP consensus sequence. Thus, based on criteria of monophyly and diagnostic amino acids, CREBZF and CREBL2 are *bona fide* bZIP subfamilies with independent evolutionary histories.

Our gene tree reconstruction supports a conclusion that 10 of 12 bZIP subfamilies — defined as any subfamily not including the bilaterian-specific orthologs CEBPg, BATF, ATF3, BACH, l-MAF, NFIL3 — were present prior to the emergence of Metazoa (Figures 1, 2B), with the families NFE2 and MAF only appearing in metazoans. Candidate ancestors of these two families remain unclear. In addition to our gene tree reconstruction, we evaluated some of the defining characteristics for these bZIP families, such as the diagnostic double glutamine at the third and fourth residues of the ATF6 subfamily and unique neighboring or expanded domains, in addition to the Cap ‘n’ Collar domain within the NFE2 orthologues or the expanded binding domain in MAF. Our tree failed to independently resolve CEBPg and BATF. These subfamilies are present due to bilaterian-specific duplication and divergence of subfamilies with ancestors, the CEBP and FOS family orthologs, present in *Capsaspora*. However, gene tree reconstructions of these subfamilies using the same parameters as used for the full data set alone increased bootstrap support to over 40%. When comparing these bootstrap values to the phylogeny containing all sequences

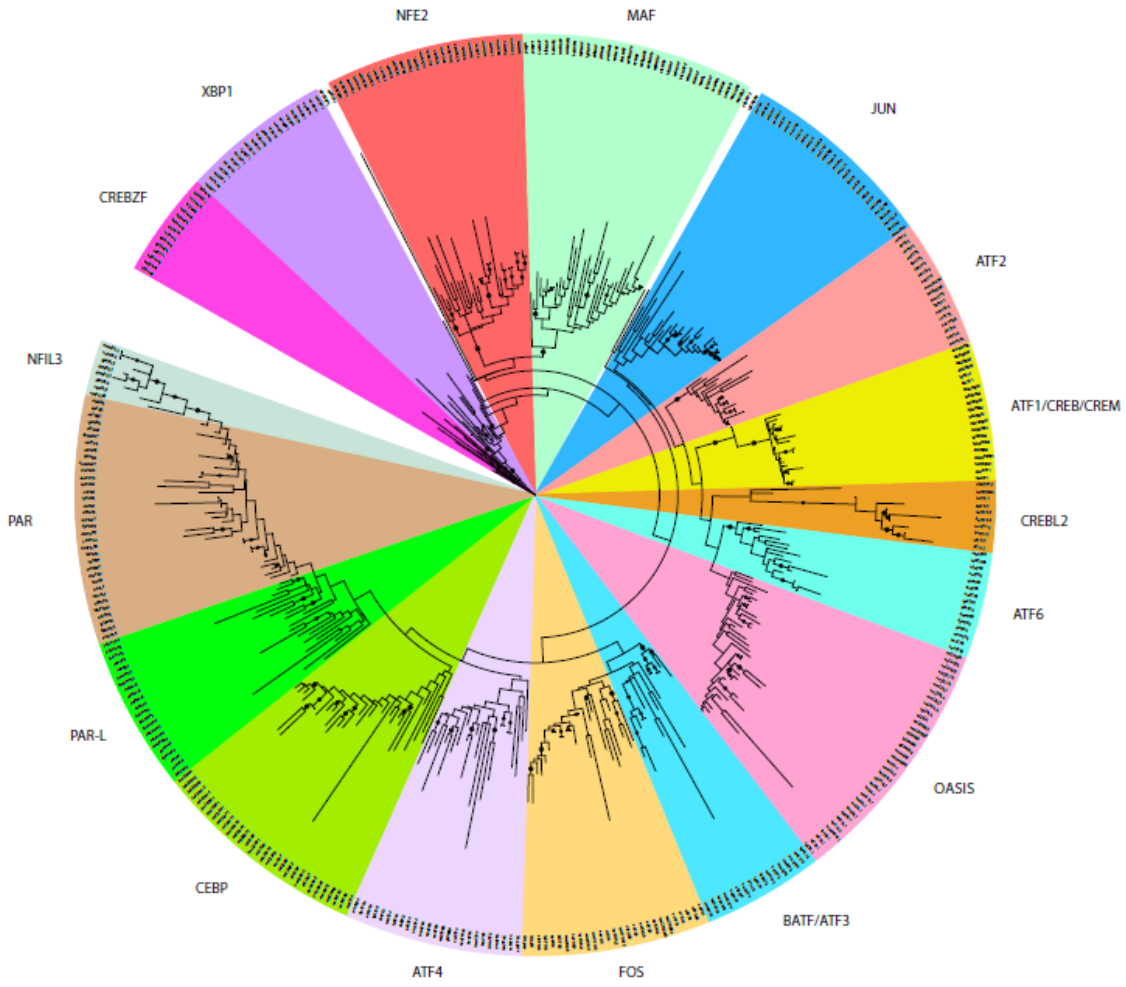


Figure 1: Phylogenetic analysis of bZIP sequences from 18 organisms reveals deep conservation of subfamily organization and dynamic evolution in animals and closely related unicellular outgroups. A) Maximum likelihood tree of 389 bZIP sequences. Created with the RaxML suite in CIPRES portal and analyzed with 1000 bootstraps. Black circles represent branches with >40% bootstrap support. Each color represents a putative subfamily, labeled on the edges of the tree. Uncolored leaves are sequences that failed to resolve into putative subfamilies.

(Figure 1), we found these bootstrap values to be sufficient. The overall results of manual inspection are listed in (Supplemental Figure 1).

The PAR proteins are bilaterian-specific PAR-like duplicates

PAR-L sequences (as defined by Jindrich and Degnan, 2016) are present in unicellular organisms and persist in all invertebrate species included in our analysis. We report a larger subset of PAR-L sequences than previous studies (Amoutzias, et al. 2007). Nine reported ecdysozoan-specific PAR-L proteins are highly divergent, possessing ~15 divergent residues within the basic region-leucine zipper interface, none of which are present within the residues used to define the subfamily (Supplemental Figure 2). The gene tree reconstructions of PAR, PAR-L and the closely related NFIL3 sequences revealed that true PAR proteins are nested within PAR-L nodes, comprising a subset of the subfamily (Figure 2B). While the true PARs resolve with >60% bootstrap support, the PAR-L subfamily exhibit lower support values, likely due to high sequence divergence. PAR-L sequences are enriched in some early diverging animals and ecdysozoans. *Mnemiopsis* lacks PAR-L proteins, but *Trichoplax* and *Amphimedon* possess 3 homologs, representing 20% of their bZIP repertoire. Duplication events are evident in cnidarians – *Nematostella* contains 7 PAR-L sequences – and ecdysozoans, but vertebrates contain only true PAR protein with a proline, acidic-rich region.

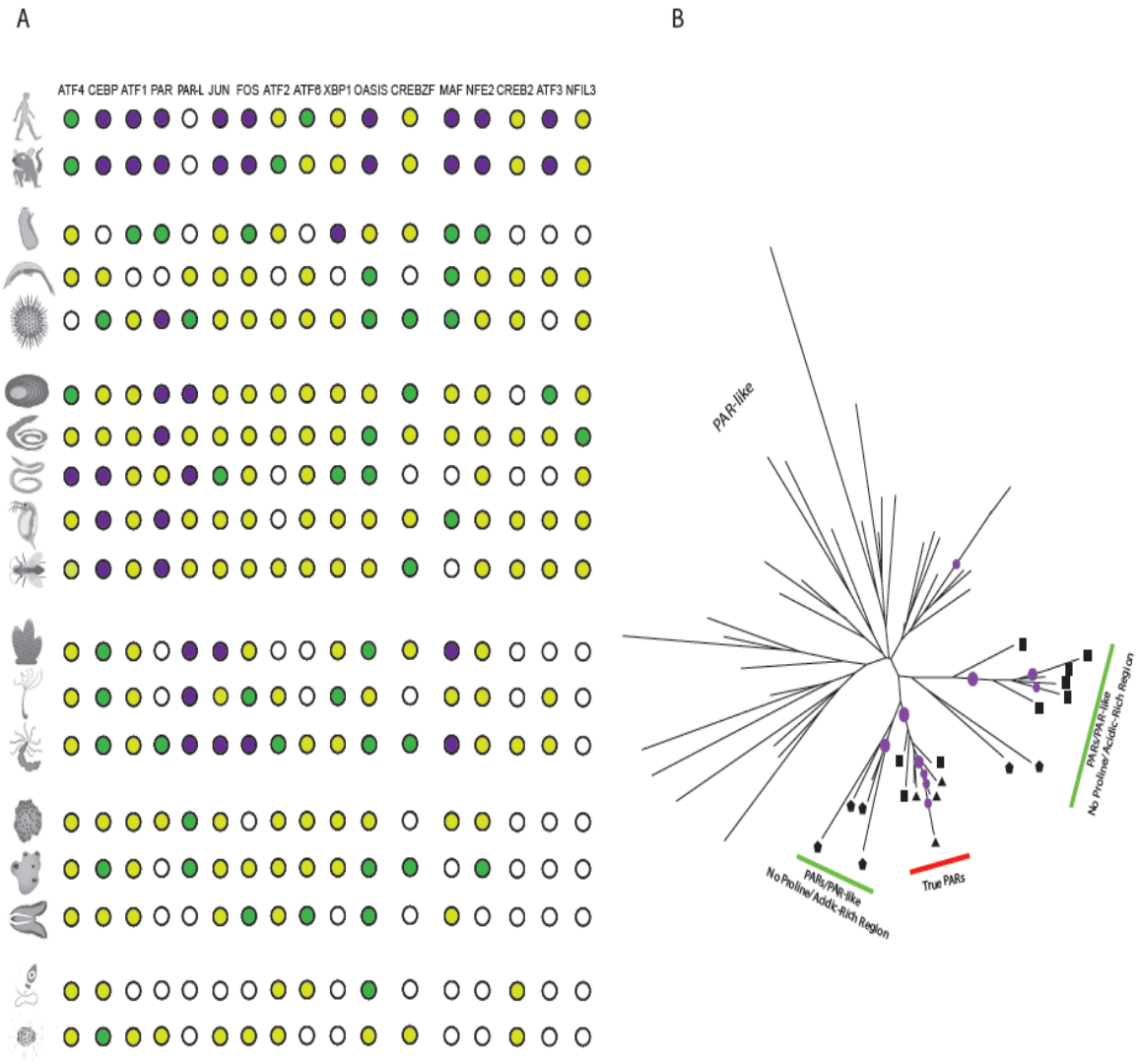


Figure 2: The bZIP Repertoire Shows Persistent Duplications of PAR-L and PAR Families. A) Visualization of the bZIP repertoire in each organism studied. White circles denote the absence of a family, yellow circle denote the organism possesses one orthologous copy, green denotes 2 paralogs, and purple denotes 3+ paralogs. B) The phylogenetic tree of the PAR-L/PAR subfamily, each purple circle denotes >70% bootstrap support. Black squares denote ecdysozoan sequences, triangles denote vertebrate sequences and pentagons denote cnidarians sequences. The bZIP domain of "True" PARs (those that contain Proline-Acidic Rich regions) nest within PAR-L families.

bZIP Dimerization Prediction

Heterodimeric activity of bZIP sequence dimerization in unicellular organisms

Interactome complexity in the three studied unicellular organisms varied widely. For yeast, we observed similar results to previous studies where bZIP interactions are limited to homodimers (Deppmann, et al. 2006). In total, there were 12 interactions predicted for the two unicellular organisms from the Holozoa. Heterodimerization activity was high in the filasterean *Capsaspora owczarzaki*, with 46 predicted dimerizations of 72 total possible unique interactions (65%) between 12 bZIP sequences. Changes in homodimerization topology between *Capsaspora* and all animals in this study was minimal (Supplemental Table 5). A majority of changes in interactome complexity were linked to expansion and contraction of heterodimerization. The complexity of the *Capsaspora* interactome was driven by a PAR-like ortholog, which interacted with 11 of the 12 *Capsaspora* bZIPs. Other promiscuous bZIPs include four ATF4-like paralogs and JUN, which interacted with 9 bZIPs (75% of total *Capsaspora* bZIPs) and ATF2 which interacts with 8 bZIPs (67% of total *Capsaspora* bZIPs). By contrast, the choanoflagellate *Monosiga brevicollis* possessed a relatively restricted interactome that formed only 25% of all possible dimers. Heterodimeric bZIPs included an ATF4 ortholog that interacted with 5 of the 8 bZIPs. The lack of promiscuous bZIPs such as ATF2, JUN, and PAR-L contributed to the stringency of this interactome.

Amphimedon, Mnemiopsis, *and* Trichoplax *retain consistently promiscuous bZIP sequences but differ in interactome density*

Animals from phyla emerging near the base of the animal tree had consistently dense interactomes regardless of the algorithm used (Figure 3, Supplemental Figure 3).

The *Amphimedon* interactome formed 88 total dimers, representing 61% of a total of 145 possible dimers within a 17 bZIP repertoire. Of the 17 *Amphimedon* bZIPs, 13 interacted with 5 or more partners. Dimerization density was driven by the PAR-L family (interacted with all 17 *Amphimedon* bZIPs), CEBP (both paralogs interacted with 15 of 17 bZIPs) and ATF2 (interacted with 15 of 17 bZIPs). The *Amphimedon* interactome consisted of the most dimerizations per bZIP queried of all species in this study. *Mnemiopsis* and *Trichoplax* have relatively reduced interactomes with 55 and 64 total dimers, respectively, representing 41% and 42% of the possible dimerization combinations. Interactome density was driven by CEBP (interacted with 8 of 12 *Mnemiopsis* bZIPs), ATF6a (interacts with 8 of 12), and JUN (also interacts with 8 of 12). The complexity of the *Trichoplax* interactome was also driven by CEBP, which formed dimers with 11 of 15 *Trichoplax* bZIPs, as well as ATF4 (10 dimers) and ATF (9 dimers). Of 15 *Trichoplax* bZIPs, 5 formed dimers with less than 6 partners.

Cnidarian bZIP repertoires form more dimers than other early diverging animals despite diminished dimerization potential

The dimerization density of bZIPs in the three cnidarians we evaluated was widely variable, but no queried cnidarian interactome forms more dimerizations per bZIP than species representing the three other early branching phyla from the metazoan tree. Each of the cnidarian repertoires produces more bZIP dimers than *Trichoplax* and *Mnemiopsis* due to higher numbers of bZIP proteins (Figure 4). The *Acropora digitifera* repertoire forms 64 bZIP dimers, while *Hydra magnipapillata* bZIPs form 104 total dimers. This is about 36% of all possible dimers for each species' respective interactome. Interactome connectivity in *Acropora* was driven by 3 CEBP proteins (all interact with

11 or more partners), JUN paralogs (9, 10 and 13 dimerization partners) and two PAR-L paralogs (9 and 11 dimerization partners). The *Acropora* bZIP repertoire includes four paralogs for MAF bZIPs, each with restrictive dimerization behavior (none interacted with more than 2 partners). In *Hydra*, promiscuous bZIPs included ATF2 (11 interactions), ATF4 (10 interactions), PAR-L (7, 10 and 9 interactions) and JUN (11 and 7 interactions). Of the 18 *Hydra* bZIPs, 13 interacted with less than 9 partners.

Nematostella vectensis was an outlier by possessing an interactome that includes 118 total predicted dimer pairs, nearly three times more than the other cnidarian. The *Nematostella* genome includes 30 total bZIPs, and explores 57% of 225 dimerization possibilities. Promiscuous bZIPs included ATF4 (21 interactions), CEBP and CEBPg (16 and 21 interactions, respectively) and five duplicated PAR-L bZIPs (all forming more than 18 different dimers). More stringent bZIPs account for 33% of the *Nematostella* repertoire, with 10 of 30 bZIPs interacting with less than 10 partners.

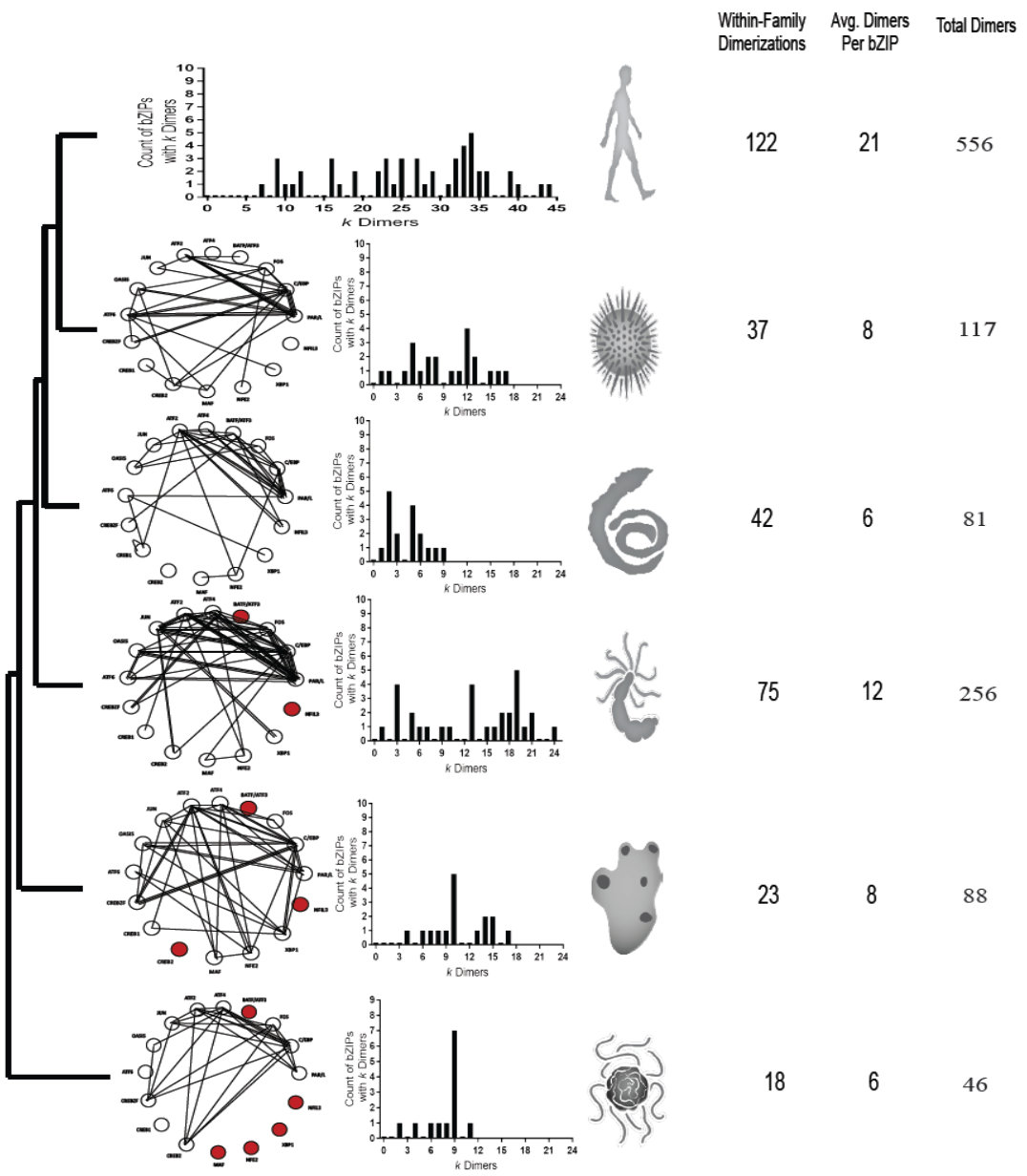


Figure 3: Complex Connectedness of closed bZIP interaction networks is present in unicellular relatives and frequently lost in invertebrate bilaterians. A) Closed network interaction network maps, circles represent orthologous bZIP groups; the absence of a subfamily is denoted in red. Each line represents an orthologous heterodimeric interaction. The wheels represent the interactome of a representative species. From bottom: *Capsaspora owczarzaki, Amphimedon queenslandica, Nematostella vectensis, Capitella teleta, Strongylocentrotus purpuratus,* and *Homo sapiens*. B) Histograms showing the count of bZIPs with a particular promiscuity in each represented genome. As the graph shifts to the right, connectedness of the interactome increases. C) Number of homodimeric interactions, average interactions per bZIP, and total predicted dimerization partners in each interactome.

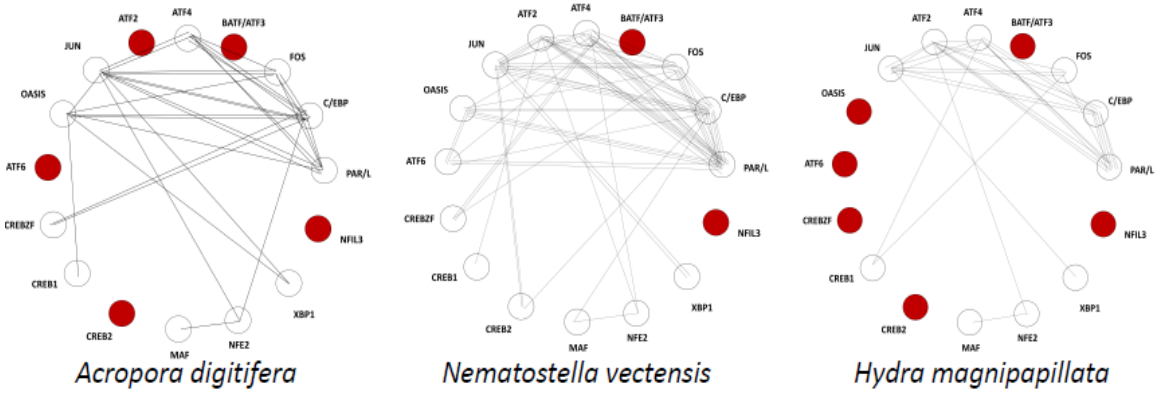


Figure 4: Dynamic Rewiring of bZIP Interactomes in Cnidaria. Interactome wheels of three cnidarians, reveal that both number of possible dimerizations and the topology of dimerization partners changes dramatically. Most rewiring is driven by expansion, contraction, and absence of orthologous families. However, topological changes also lead to different interactome characteristics, such as the CREB1-OASIS interaction that is unique to *Acropora*.

Fluctuations in bZIP gene count, bZIP promiscuity and variation of interactome complexity in bilaterian invertebrates

The bZIP repertoire of the annelid *Capitella teleta* includes 24 bZIPs, the most of the sampled bilaterian invertebrates. The *Capitella* interactome consisted of 81 total dimers, about 28% of total possible dimers. Despite lower bZIP gene counts, *Lottia gigantea* (22 bZIPs and 121 dimers), *Strongylocentrotus purpuratus* (22 bZIPs and 117 dimers) and *Daphnia pulex* (23 bZIPs and 101 dimers) possessed bZIP interactomes of higher density. The cephalochordate *Branchiostoma floridaae* bZIP count was the lowest among the bilaterian invertebrates, possessing 18 total bZIPs. The *Branchiostoma* interactome formed 60 total dimers, 37% of all possible dimers. Despite higher gene counts, *Drosophila melanogaster* (20 bZIPs, and 67 dimers) and *Ciona intestinalis* (19 bZIPs, and 63 dimers) formed less dimers. Even though the *Caenorhabditis elegans* bZIP repertoire consisted of 5 more bZIPs than that of *Branchiostoma*, only 9 more dimers were present in the *Caenorhabditis elegans* bZIP dimerization network.

While diversity in bZIP interactivity for individual proteins was generally consistent and thus had a lower impact on connectedness compared to bZIP count, changes in promiscuity of the JUN, FOS and ATF4 subfamilies show variation over the animal species. The JUN orthologs in *Capitella*, *Branchiostoma, Caenorhabditis*, and *Drosophila* interact with an average of 29% of within-species bZIPs while averaging interactions with 51% of bZIPs in *Lottia, Strongylocentrotus* and *Daphnia*. Similarly, FOS proteins interacted with an average of 25% of within-species bZIPs in the stringent group listed above while interacting with an average of 42% in *Lottia, Strongylocentrotus*

and *Daphnia*. The promiscuity of ATF4 orthologs also fluctuates, interacting with an average of 24% of within-species bZIP partners in the stringent group and 46% in the more well-connected group mentioned above.

Increase in gene count drives gains in interactome complexity in vertebrates

The proposed duplication of all subfamilies due to whole genome or single gene duplication events in the common ancestor of vertebrates- or the 2R hypothesis- contribute to increases in dimerization. The *Homo sapiens* interactome contains 51 bZIPs and 556 possible dimers of 1,302 total possible combinations. The subfamilies JUN (3 paralogs), FOS (5 paralogs), ATF2 (2 paralogs), CEBP (5 paralogs), ATF4 (2 paralogs), and PAR-L (4 paralogs) are highly promiscuous, and all duplicated paralogs retain the ability to form many dimers. For example, CEBPA, B, E and G interact with between 34 and 36 partners, a range of 68% to 72% of bZIPs in the repertoire. The FOS paralogs interact with no less than 24 partners and the JUN-dimerization protein 2 (JDP2) forming 44 potential dimers. Retention of promiscuity in vertebrate duplicates is more pronounced than in other bilaterians. For example, PAR-L proteins in *Branchiostoma* interact with a range from 28% of the interactome to 67% of the bZIP interactome. Similarly, the PAR-L paralogs in *Capitella* range from interacting with 6 partners to 14 partners, a range of 33% between the least and highest connected paralog. Duplication of non-promiscuous families such as OASIS (4 paralogs) and ATF1 (3 paralogs) also contributed to increases in dimerization without an increase in promiscuity. The OASIS paralogs averaged interactions with 20% of the repertoire while ATF1 paralogs averaged 23%. Both values are close to the average promiscuity of the subfamily in other species, ATF1 averaged 22% interaction rate across all species while OASIS averages 26%.

Interactome connectivity is related to conserved promiscuity of certain hub bZIPs

Certain bZIP subfamilies have conserved degrees of promiscuous interactions, independent of the connectedness of the overall bZIP interactome. These highly connected proteins can be considered "hub" proteins in interactome maps due to the large number of connections with other proteins. The PAR-L family interacts with an average of 49% of bZIPs in any given interactome. In each animal studied, a PAR-L ortholog interacts with more than 55% of the bZIP repertoire. Similarly, the CEBP subfamily averages dimerization with 52% of bZIPs in any repertoire and each animal studies possesses at least one ortholog that interacts with at least 60% of potential bZIP partners.

Other hubs show high promiscuity in a species-dependent manner where certain conserved bZIPs possess substantial differences in interactive capacity (Figure 5). The ATF4 subfamily interacted with an average of 56% of bZIPs within any given interactome. However, orthologs present in *Caenorhabditis*, *Drosophila*, *Capitella*, and *Branchiostoma* never exceeded dimerization rates of 44%. The *Capitella* ATF4 interacted with only 20% of potential partners. By contrast, the *Nematostella* ATF4 interacted with 70% of potential partners.

Stringent bZIPs retain low interaction rates in each animal studied. The OASIS and ATF1 families, as described above, interacted with an average of 22% and 26% of bZIPs within any interactome, respectively. Neither subfamily possesses an ortholog that interacted with over 30% of any interactome. The Zhangfei family averaged 23 % interaction rate within any species, and never interacted with more than 35% of any interactome.

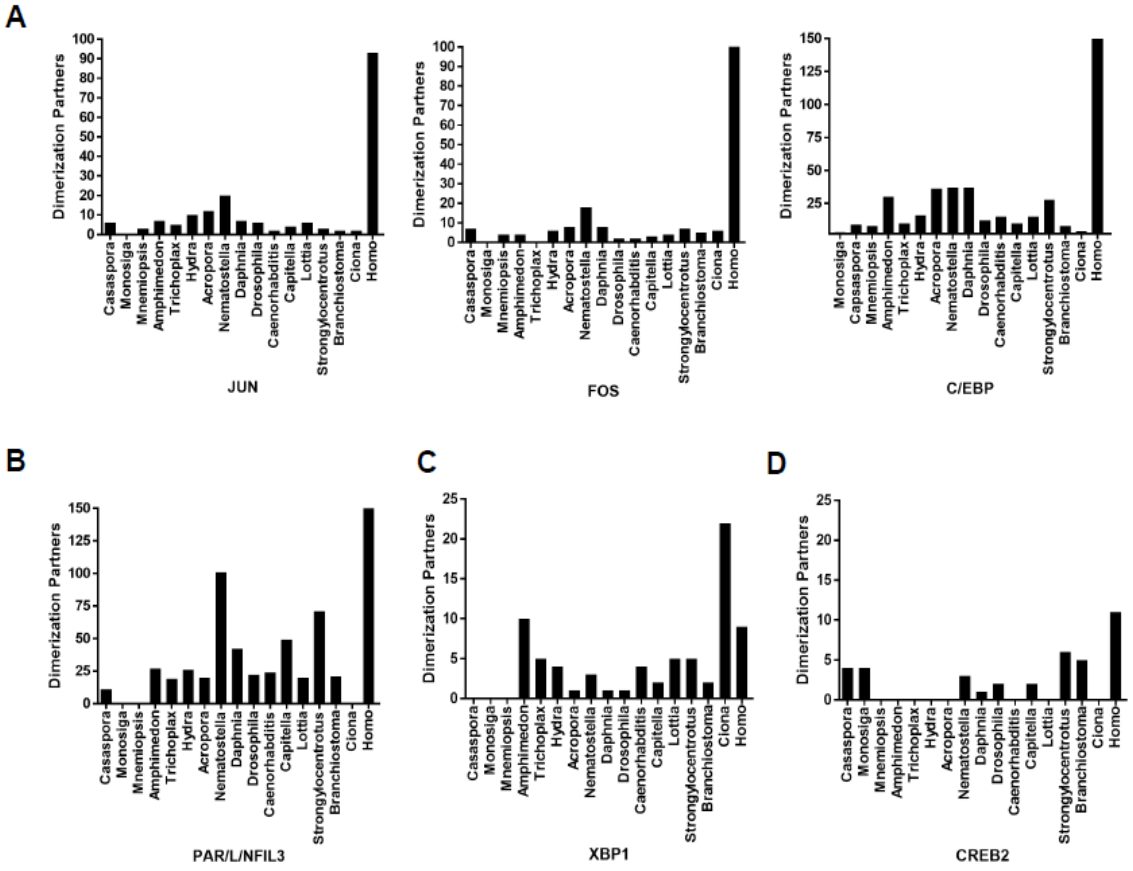


Figure 5: Promiscuity of hub bZIPs drives interactome diversity throughout the metazoan lineage. A) Dimerization activity of classical hub bZIPs is particularly enriched in cnidarians and vertebrates. B) PAR/L bZIPs are highly promiscuous dimerization partners in cnidarians, and promiscuity fluctuates dramatically in invertebrate bilaterians. C) *Amphimedon* and *Ciona* contain XBP1 orthologs with higher promiscuity than the vertebrate orthologue. D) The dimerization potential of the CREBL2 family is strictly constrained.

Between-Species bZIP Interactions

The interaction rate of bZIP sequences within species was similar to the interaction rate of bZIPs between species. For example, the *Amphimedon* PAR-L ortholog was predicted to interact with 100% of *Amphimedon* bZIPs and 79% of all bZIPs in the dataset. As a whole, the PAR-bZIPs interact with 57% of the bZIP dataset; and CEBPs interact with 53% of the entire dataset. Other hub proteins, ATF2 and ATF4 both interact with 52% of the dataset. Stringent bZIPs are also constrained across species. NFE2 and MAF sequences interact with 20% and 14% of all bZIPs, respectively and CREBL2 interacts with 16% of the full bZIP dataset.

Furthermore, subtle shifts in promiscuity of each subfamily led to wider patterns of between-species sequence promiscuity. The invertebrate bilaterian bZIPs are less interactive with other invertebrate bilaterian sequences than vertebrate sequences are with other vertebrate sequences (Figure 6). This is likely a result of duplication events and retention of promiscuity in duplicate hub orthologues. We observe multiplicative increases in interactome connectivity between vertebrates rather than squared (Figure 3). Sequences from *Capsaspora* and animals from early diverging phyla are more promiscuous and tend to interact with each other to a greater degree despite possessing less or comparable bZIP gene count. However, some bZIPs develop unique within-species increases in promiscuity without becoming broadly promiscuous between species. A unique duplication of XBP1 in the tunicate *Ciona intestinalis* led to an uncharacteristic expansion of heterodimeric activity (one paralog interacts with 50% of the *Ciona* repertoire). Despite this increase in promiscuity within-species, the *Ciona* XBP1

subfamily only interacts with 21% of the overall between-species dataset only 2% more interactions than an average XBP1 protein.

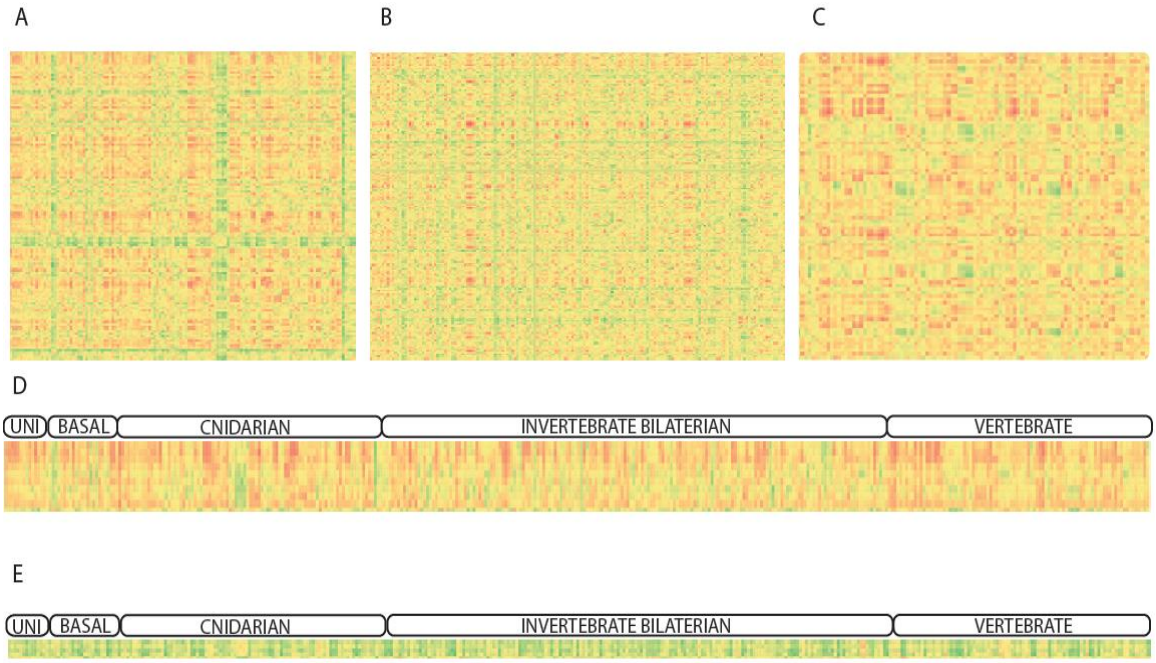


Figure 6: bZIP promiscuity is constrained between species. Heatmaps reveal the likelihood that a given bZIP interacts with another. Red colors denote a high interactive likelihood; yellow denotes moderate likelihood, and green denotes no interaction predicted A) Heatmap showing interaction scores between each unicellular, divergent metazoan, and cnidarian bZIP sequences. Promiscuous bZIPs retain connectedness between these species. B) Interactions scores between bZIP sequences from each invertebrate bilaterian show a constrained drop in interconnectedness. C) Heatmap of vertebrate sequences reveals high between species promiscuity. D) bZIP hub sequences queried from *Acropora digitifera* retain promiscuous interaction with all represented bZIPs, covering over half a billion years of evolution. E) MAF sequences from *Acropora digitifera* reveal between-species conservation of limited connectivity.

DISCUSSION

The majority of transcription factor families appear to have emerged prior to the divergence of metazoans and undergone duplication events throughout the animal kingdom. The bZIPs are among the most diverse and ancient eukaryotic families and have radiated into the second most diverse transcription factor family in metazoans (Amoutzias, et al. 2008; Dunn and Ryan 2015; Jindrich and Degnan 2016). An updated bZIP phylogeny reveals patterns of duplications and losses in subfamily content, but tracing phenotypic consequences of these patterns requires annotation of protein behavior. Protein-protein interactions are crucially important in molecular pathways and can evolve dynamically (Echave, et al. 2016; Fraser, et al. 2002). Prediction of bZIP protein-protein interactions in metazoans suggests that bZIP dimerization activity is dependent on expansion or loss of bZIP subfamily gene count itself. However, count-independent changes in promiscuity of certain bZIP subfamilies also play a significant, but less impactful, role in evolving interactomes.

The dependence of interactome connectedness on bZIP repertoire overall suggests a conclusion that particular interaction topologies are highly conserved. Certain bZIP heterodimeric interactions appear to be constrained for over a billion years. For example, the JUN-FOS, ATF4-CEBP, PAR-L-CEBP, and ATF2-JUN dimers originated before the metazoan divergence and persist throughout metazoan evolution (Figure 3). Many of these conserved interactions involve hub bZIPs, promiscuous proteins that form a variety of dimers. Well-annotated duplications in the promiscuous PAR-L, CEBP, ATF2, JUN and FOS bZIP subfamilies (Figure 1, Figure 2) are associated with increases in bZIP network connectedness (Figure 2, Figure 3) simply because the associated interactomes

have more copies of promiscuous proteins with similar interactive topology. For example, the duplication of JUN, FOS, PAR-L and CEBP subfamilies alone drives a majority of interactome connectedness in cnidarians (Figure 4) and vertebrates (Figure 2, Figure 3). This is potentially unexpected because proteins that form many different interactions typically evolve at a slower rate than less connected proteins (Alvarez-Ponce and Fares 2012; Eanes 2011). Stringent bZIPs also retain similar topologies. The NFE2 and MAF subfamilies emerged at the base of the animal tree, and form dimers in every animal possessing both subfamilies. Similarly, the XBP1-ATF6 interaction predates the metazoan lineage (Figure 3, Supplemental Figure 3).

While duplication events are primary driving forces of changing interactome topologies, the behavior of genes independent from count also shapes interactome evolution. For example, while *Nematostella* possesses more JUN and FOS duplicates the cnidarian JUN and FOS sequences are also more promiscuous themselves than orthologs in bilaterians like *Capitella* and *Caenorhabditis elegans* (Figure 3, Supplemental Figure 4). Multiple duplications of the promiscuous CEBP family expanded the total dimers formed within vertebrate interactomes, and only one of the five CEBP paralogs in humans lost its ancestral promiscuity (Figure 3, Supplementary Figure 4. Supplementary Table 4). However, in bilaterian invertebrates, the promiscuity of CEBP duplicates fluctuates within species, and as a consequence the CEBP subfamily is less connected in *Capitella* and *Drosophila* than in vertebrates and *Nematostella* (Figure 3, Supplementary Figure 4). This dynamic rewiring happens in non-hub bZIPs, leading to species-specific topologies formed by otherwise stringent dimerizing proteins. Species or lineage-specific changes in dimerization potential include: increases in promiscuity of NFE2 and MAF in

the vertebrate lineage (Figure 3), the unique XBP1 circuit in *Ciona* that expanded on the XBP1-ATF6 interaction by adding interactions with OASIS and the PARs (Supplementary Figure 3), and the uncharacteristic promiscuity of ATF2 in the highly connected *Amphimedon* interactome (Figure 3).

Conservation of promiscuity reaches outside of individual interactomes. Promiscuous or stringent bZIPs in a single interactome will likely retain similar degrees of heterodimerization activity in every animal interactome. This suggests that the dimerization potential of bZIPs is generally constrained. If the bZIP interactosphere were unconstrained, one would expect more bZIPs to form completely unique heterodimeric nodes as dimerization topology diverges. Our results support previous evidence that naturally occurring leucine zippers- as opposed to those designed in a lab- explore a constrained interactosphere (Grigoryan, et al. 2009; Newman and Keating 2003). Otherwise, we would observe more instances of dramatic species-specific changes in bZIP dimerization formation.

Constraint on dimerization dynamics would still allow the emergence of new dimerization patterns, but only within and between closely related bZIP subfamilies. This would avoid pleiotropic consequences of evolving completely new interaction dynamics. Our results, however, present possible challenges to this notion. The bZIP subfamilies that have undergone rapid radiations (PAR-L, CEBP and FOS) are among the most connected genes within the bZIP interactomes (Figure 7). However, this interpretation only takes dimerization events into consideration, and does not address connections to other proteins, subcellular compartmentalization, or the expression pattern of the protein. The molecular underpinnings of the changes in developmental processes and

environmental response pathways are dependent on protein-protein interactions (Rudra, et al. 2012; Tepass 2012). Shifts in dimerization activity and spatiotemporal expression of transcription factors can have concerted effects on the gene regulatory elements that govern development and tissue integrity (Davis, et al. 2016; Pajares, et al. 2016). Developmental functions of bZIPs often shift in different organisms (Bogeska and Pahl 2013; Sharma, et al. 2014), suggesting dynamic rewiring of bZIP dimers in GRN circuitry.

It is important to note that annotation of protein-protein interactions, whether from a computational model or *in vitro*, cannot tell us how these proteins will actually interact in animals. Practically, defining protein-protein interactions empirically can lead to inaccuracies: false positives (Banks, et al. 2015) and unreported negative results (Blohm, et al. 2014). Computational models are only as good as the empirical data on which the program is trained. Tissue specific expression (Reitzel, et al. 2016), stage specific expression (Cui, et al. 2004; Ikuzawa, et al. 2006), and general expression parameters (Booth and King 2016) may mean that many possible interactions never happen within an organism. However, differential regulation of transcription may express potential dimerization partners that have never interacted before in the same tissue. These new interactions could lead to a mechanism for organismal novelty. Understanding general annotation of interactions may help guide researchers to discover novel interactions within already well-studied systems and in non-model organisms.

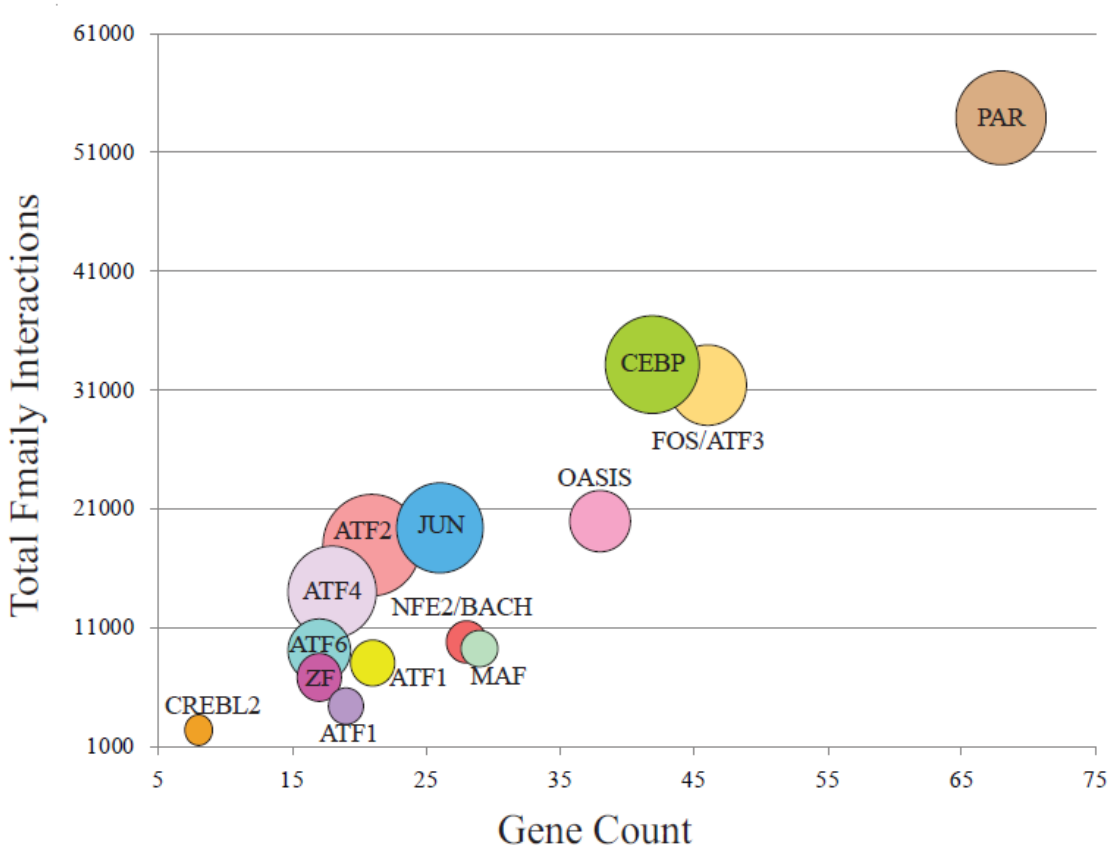


Figure 7: Total Dimers Formed by bZIPs is Gene Count Dependent. Each circle is labeled with corresponding subfamily. Area of circles represents average percent of the dataset that members of the subfamilies interact with. There is a linear relationship between total dimers formed and bZIP gene count. The PAR/L family is an outlier that drives this trend. Subfamily-specific promiscuity causes multiple orthologs (JUN, ATF2, ATF4) to form more dimers than predicted by this linear trend. These are among the most frequently rewired families.

REFERENCES

Acharya A, Rishi V, Vinson C 2006. 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 45: 11324-11332. doi: 10.1021/bi060822u

Alvarez-Ponce D, Fares MA 2012. Evolutionary rate and duplicability in the *Arabidopsis thaliana* protein–protein interaction network. Genome Biology and Evolution 4: 1263-1274. doi: 10.1093/gbe/evs101

Amoutzias GD, Robertson DL, Van de Peer Y, Oliver SG 2008. Choose your partners: dimerization in eukaryotic transcription factors. Trends in Biochemical Sciences 33: 220-229. doi: http://doi.org/10.1016/j.tibs.2008.02.002

Amoutzias GD, Veron AS, Weiner IJ, Robinson-Rechavi M, Bornberg-Bauer E, Oliver SG, Robertson DL 2007. One billion years of bZIP transcription factor evolution: conservation and change in dimerization and DNA-binding site specificity. Molecular Biology and Evolution 24: 827-835. doi: 10.1093/molbev/msl211

An JH, Blackwell TK 2003. SKN-1 links *C. elegans* mesendodermal specification to a conserved oxidative stress response. Genes & Development 17: 1882-1893. doi: 10.1101/gad.1107803

Andreani J, Guerois R 2014. Evolution of protein interactions: From interactomes to interfaces. Archives of Biochemistry and Biophysics 554: 65-75. doi: http://doi.org/10.1016/j.abb.2014.05.010

Banks CAS, Boanca G, Lee ZT, Florens L, Washburn MP 2015. Proteins interacting with cloning scars: a source of false positive protein-protein interactions. Scientific reports 5: 8530. doi: 10.1038/srep08530

Blohm P, Frishman G, Smialowski P, Goebels F, Wachinger B, Ruepp A, Frishman D 2014. Negatome 2.0: a database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis. Nucleic Acids Research 42: D396-D400. doi: 10.1093/nar/gkt1079

Bogeska R, Pahl HL 2013. Elevated Nuclear factor erythroid-2 levels promote epo-independent erythroid maturation and recapitulate the hematopoietic stem cell and common myeloid progenitor expansion observed in polycythemia vera patients. Stem cells translational medicine 2: 112-117. doi: 10.5966/sctm.2012-0046

Booth DS, King N 2016. Evolution: gene regulation in transition. Nature 534: 482-483. doi: 10.1038/nature18447

Chan JY, Kwong M, Lu R, Chang J, Wang B, Yen TSB, Kan YW 1998. Targeted disruption of the ubiquitous CNC-bZIP transcription factor, Nrf-1, results in anemia and embryonic lethality in mice. The EMBO Journal 17: 1779-1787. doi: 10.1093/emboj/17.6.1779

Cheatle Jarvela AM, Brubaker L, Vedenko A, Gupta A, Armitage BA, Bulyk ML, Hinman VF 2014. Modular evolution of DNA-binding preference of a tbrain transcription factor provides a mechanism for modifying gene regulatory networks. Molecular Biology and Evolution 31: 2672-2688. doi: 10.1093/molbev/msu213

Cheatle Jarvela AM, Hinman VF 2015. Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks. EvoDevo 6: 3. doi: 10.1186/2041-9139-6-3

Claverie J-M 2001. What if there are only 30,000 human genes? Science 291: 1255-1257. doi: 10.1126/science.1058969

Cui W, Tomarev SI, Piatigorsky J, Chepelinsky AB, Duncan MK 2004. Mafs, Prox1, and Pax6 can regulate chicken βB1-crystallin gene expression. Journal of Biological Chemistry 279: 11088-11095. doi: 10.1074/jbc.M312414200

Darriba D, Taboada GL, Doallo R, Posada D 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27: 1164-1165. doi: 10.1093/bioinformatics/btr088

Davidson EH, Erwin DH 2006. Gene regulatory networks and the evolution of animal body plans. Science 311: 796-800. doi: 10.1126/science.1113832

Davis KR, Giesy SL, Long Q, Krumm CS, Harvatine KJ, Boisclair YR 2016. XBP1 regulates the biosynthetic capacity of the mammary gland during lactation by controlling cpithelial expansion and endoplasmic reticulum formation. Endocrinology 157: 417-428. doi: 10.1210/en.2015-1676

Deppmann CD, Alvania RS, Taparowsky EJ 2006. Cross-species annotation of basic leucine zipper factor interactions: insight into the evolution of closed interaction networks. Molecular Biology and Evolution 23: 1480-1492. doi: 10.1093/molbev/msl022

Dibner C, Schibler U 2015. Circadian timing of metabolism in animal models and humans. Journal of Internal Medicine 277: 513-527. doi: 10.1111/joim.12347

Dunn CW, Ryan JF 2015. The evolution of animal genomes. Current Opinion in Genetics & Development 35: 25-32. doi: http://doi.org/10.1016/j.gde.2015.08.006

Eanes WF 2011. Molecular population genetics and selection in the glycolytic pathway. The Journal of Experimental Biology 214: 165-171. doi: 10.1242/jeb.046458

Echave J, Spielman SJ, Wilke CO 2016. Causes of evolutionary rate variation among protein sites. Nat Rev Genet 17: 109-121. doi: 10.1038/nrg.2015.18

Edgar RC 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32.

Elcock AH, McCammon JA 2001. Identification of protein oligomerization states by analysis of interface conservation. Proceedings of the National Academy of Sciences 98: 2990-2994. doi: 10.1073/pnas.061411798

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M 2014. Pfam: the protein families database. Nucleic Acids Research 42: D222-D230. doi: 10.1093/nar/gkt1223

Fong JH, Keating AE, Singh M 2004. Predicting specificity in bZIP coiled-coil protein interactions. Genome Biology 5: R11. doi: 10.1186/gb-2004-5-2-r11

Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW 2002. Evolutionary rate in the protein interaction network. Science 296: 750-752. doi: 10.1126/science.1068696

Gachon F, Olela FF, Schaad O, Descombes P, Schibler U 2006. The circadian PAR-domain basic leucine zipper transcription factors DBP, TEF, and HLF modulate basal and inducible xenobiotic detoxification. Cell Metabolism 4: 25-36. doi: http://doi.org/10.1016/j.cmet.2006.04.015

Gasiorek JJ, Blank V 2015. Regulation and function of the NFE2 transcription factor in hematopoietic and non-hematopoietic cells. Cellular and Molecular Life Sciences 72: 2323-2335. doi: 10.1007/s00018-015-1866-6

George H, Terracol R 1997. The vrille Gene of *Drosophila* is a maternal enhancer of decapentaplegic and encodes a new member of the bZIP family of transcription factors. Genetics 146: 1345-1363.

Grigoryan G, Keating AE 2006. Structure-based prediction of bzip partnering specificity. Journal of Molecular Biology 355: 1125-1142. doi: http://doi.org/10.1016/j.jmb.2005.11.036

Grigoryan G, Reinke AW, Keating AE 2009. Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature 458: 859-864. doi: http://www.nature.com/nature/journal/v458/n7240/suppinfo/nature07885_S1.html

Guo M, Chen Y, Du Y, Dong Y, Guo W, Zhai S, Zhang H, Dong S, Zhang Z, Wang Y, Wang P, Zheng X 2011. The bZIP Transcription Factor MoAP1 Mediates the Oxidative Stress Response and Is Critical for Pathogenicity of the rice blast fungus magnaporthe oryzae. PLOS Pathogens 7: e1001302. doi: 10.1371/journal.ppat.1001302

Hamer R, Luo Q, Armitage JP, Reinert G, Deane CM 2010. i-Patch: Interprotein contact prediction using local network information. Proteins: Structure, Function, and Bioinformatics 78: 2781-2797. doi: 10.1002/prot.22792

Hart-Smith G, Yagoub D, Tay AP, Pickford R, Wilkins MR 2016. Large scale mass spectrometry-based identifications of enzyme-mediated protein methylation are subject to high false discovery rates. Molecular & Cellular Proteomics 15: 989-1006. doi: 10.1074/mcp.M115.055384

Hayashi A, Kasahara T, Iwamoto K, Ishiwata M, Kametani M, Kakiuchi C, Furuichi T, Kato T 2007. The role of brain-derived neurotrophic factor (BDNF)-induced XBP1 splicing during brain development. Journal of Biological Chemistry 282: 34525-34534. doi: 10.1074/jbc.M704300200

Howell SH 2013. Endoplasmic reticulum stress responses in plants. Annual review of plant biology 64: 477-499.

Ikuzawa M, Shimizu K, Yasumasu S, Iuchi I, Shi Y-B, Ishizuya-Oka A 2006. Thyroid hormone-induced expression of a bZip-containing transcription factor activates epithelial cell proliferation during *Xenopus* larval-to-adult intestinal remodeling. Development Genes and Evolution 216: 109-118. doi: 10.1007/s00427-005-0037-4

Jindrich K, Degnan BM 2016. The diversification of the basic leucine zipper family in eukaryotes correlates with the evolution of multicellularity. BMC Evolutionary Biology 16: 28. doi: 10.1186/s12862-016-0598-z

King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, Marr M, Pincus D, Putnam N, Rokas A, Wright KJ, Zuzow R, Dirks W, Good M, Goodstein D, Lemons D, Li W, Lyons JB, Morris A, Nichols S, Richter DJ, Salamov A, Sequencing JGI, Bork P, Lim WA, Manning G, Miller WT, McGinnis W, Shapiro H, Tjian R, Grigoriev IV, Rokhsar D 2008. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. Nature 451: 783-788.

Kusserow A, Pang K, Sturm C, Hrouda M, Lentfer J, Schmidt HA, Technau U, von Haeseler A, Hobmayer B, Martindale MQ 2005. Unexpected complexity of the Wnt gene family in a sea anemone. Nature 433: 156-160.

Letunic I, Bork P 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23: 127-128. doi: 10.1093/bioinformatics/btl529

Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, Senderovich N, Kovalev E, Silver DH, Feder M, Fernandez-Valverde SL, Nakanishi N, Simmons D, Simakov O, Larsson T, Liu S-Y, Jerafi-Vider A, Yaniv K, Ryan JF, Martindale MQ, Rink JC, Arendt D, Degnan SM, Degnan BM, Hashimshony T, Yanai I 2016. The mid-developmental transition and the evolution of animal body plans. Nature 531: 637-641. doi: 10.1038/nature16994

Luisi P, Alvarez-Ponce D, Pybus M, Fares MA, Bertranpetit J, Laayouni H 2015. Recent positive selection has acted on genes encoding proteins with more interactions within the whole human interactome. Genome Biology and Evolution 7: 1141-1154. doi: 10.1093/gbe/evv055

Lynch VJ, Brayer K, Gellersen B, Wagner GP 2009. HoxA-11 and FOXO1A cooperate to regulate decidual prolactin expression: towards inferring the core transcriptional regulators of decidual genes. PLoS ONE 4.

Lynch VJ, May G, Wagner GP 2011. Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. Nature 480: 383-386.

Ma X, Zhang H, Yuan L, Jing H, Thacker P, Li D 2011. CREBL2, interacting with CREB, induces adipogenesis in 3T3-L1 adipocytes. Biochemical Journal 439: 27-38. doi: 10.1042/bj20101475

Miller DJ, Ball EE, Technau U 2005. Cnidarians and ancestral genetic complexity in the animal kingdom. Trends in Genetics 21: 536-539. doi: http://doi.org/10.1016/j.tig.2005.08.002

Miller M, Shuman JD, Sebastian T, Dauter Z, Johnson PF 2003. Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/Enhancer-binding protein α. Journal of Biological Chemistry 278: 15178-15184. doi: 10.1074/jbc.M300417200

Miller MA, Pfeiffer W, Schwartz T editors. 2010 Gateway computing environments workshop (GCE). 2010 14-14 Nov. 2010.

Mosca R, Pache RA, Aloy P 2012. The role of structural disorder in the rewiring of protein interactions through evolution. Molecular & Cellular Proteomics 11. doi: 10.1074/mcp.M111.014969

Mozgova I, Hennig L 2015. The polycomb group protein regulatory network. Annual review of plant biology 66: 269-296.

Nakagawa S, Gisselbrecht SS, Rogers JM, Hartl DL, Bulyk ML 2013. DNA-binding specificity changes in the evolution of forkhead transcription factors. Proceedings of the National Academy of Sciences 110: 12349-12354. doi: 10.1073/pnas.1310430110

Newman JRS, Keating AE 2003. Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science 300: 2097-2101. doi: 10.1126/science.1084648

Nicholas KB, Nicholas HB, Deerfield D 1997. GeneDoc: analysis and visualization of genetic variation. Embnew. news 4.

Pajares M, Jiménez-Moreno N, García-Yagüe ÁJ, Escoll M, de Ceballos ML, Van Leuven F, Rábano A, Yamamoto M, Rojo AI, Cuadrado A 2016. Transcription factor NFE2L2/NRF2 is a regulator of macroautophagy genes. Autophagy 12: 1902-1916. doi: 10.1080/15548627.2016.1208889

Potapov V, Kaplan JB, Keating AE 2015. Data-driven prediction and design of bzip coiled-coil interactions. PLOS Computational Biology 11: e1004046. doi: 10.1371/journal.pcbi.1004046

Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317: 86-94. doi: 10.1126/science.1139158

Rebeiz M, Patel NH, Hinman VF 2015. Unraveling the tangled skein: the evolution of transcriptional regulatory networks in development. Annual review of genomics and human genetics 16: 103-131.

Reinke AW, Baek J, Ashenberg O, Keating AE 2013. Networks of bZIP protein-protein interactions diversified over a billion years of evolution. Science 340: 730-734. doi: 10.1126/science.1233465

Reinke AW, Grant RA, Keating AE 2010. A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. Journal of the American Chemical Society 132: 6025-6031. doi: 10.1021/ja907617a

Reitzel AM, Pang K, Martindale MQ 2016. Developmental expression of "germline"- and "sex determination"-related genes in the ctenophore *Mnemiopsis leidyi*. EvoDevo 7: 17. doi: 10.1186/s13227-016-0051-9

Reitzel AM, Tarrant AM, Levy O 2013. Circadian clocks in the cnidaria: environmental entrainment, molecular regulation, and organismal outputs. Integr Comp Biol 53. doi: 10.1093/icb/ict024

Rivera AS, Pankey MS, Plachetzki DC, Villacorta C, Syme AE, Serb JM, Omilian AR, Oakley TH 2010. Gene duplication and the origins of morphological complexity in pancrustacean eyes, a genomic approach. BMC Evolutionary Biology 10: 123. doi: 10.1186/1471-2148-10-123

Rudra D, deRoos P, Chaudhry A, Niec RE, Arvey A, Samstein RM, Leslie C, Shaffer SA, Goodlett DR, Rudensky AY 2012. Transcription factor Foxp3 and its protein partners form a complex regulatory network. Nat Immunol 13: 1010-1019.

Sayou C, Monniaux M, Nanao MH, Moyroud E, Brockington SF, Thévenon E, Chahtane H, Warthmann N, Melkonian M, Zhang Y, Wong GK-S, Weigel D, Parcy F, Dumas R 2014. A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. Science 343: 645-648. doi: 10.1126/science.1248229

Sebé-Pedrós A, de Mendoza A. 2015. Transcription factors and the origin of animal multicellularity. In: Ruiz-Trillo I, Nedelcu AM, editors. Evolutionary Transitions to Multicellular Life: Principles and mechanisms. Dordrecht: Springer Netherlands. p. 379-394.

Sebé-Pedrós A, de Mendoza A, Lang BF, Degnan BM, Ruiz-Trillo I 2011. Unexpected repertoire of metazoan transcription factors in the unicellular holozoan *Capsaspora owczarzaki*. Molecular Biology and Evolution 28: 1241-1254. doi: 10.1093/molbev/msq309

Sharma PP, Gupta T, Schwager EE, Wheeler WC, Extavour CG 2014. Subdivision of arthropod cap-n-collar expression domains is restricted to Mandibulata. EvoDevo 5: 3. doi: 10.1186/2041-9139-5-3

Sone M, Zeng X, Larese J, Ryoo HD 2013. A modified UPR stress sensing system reveals a novel tissue distribution of IRE1/XBP1 activity during normal *Drosophila* development. Cell Stress and Chaperones 18: 307-319. doi: 10.1007/s12192-012-0383-x

Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML, Signorovitch AY, Moreno MA, Kamm K, Grimwood J, Schmutz J, Shapiro H, Grigoriev IV, Buss LW, Schierwater B, Dellaporta SL, Rokhsar DS 2008. The *Trichoplax* genome and the nature of placozoans. Nature 454: 955-960.

Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U, Larroux C, Putnam NH, Stanke M, Adamska M, Darling A, Degnan SM, Oakley TH, Plachetzki DC, Zhai Y, Adamski M, Calcino A, Cummins SF, Goodstein DM, Harris C, Jackson DJ, Leys SP, Shu S, Woodcroft BJ, Vervoort M, Kosik KS, Manning G, Degnan BM, Rokhsar DS

2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature 466: 720-726.

Stamatakis A 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22. doi: 10.1093/bioinformatics/btl446

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30: 2725-2729. doi: 10.1093/molbev/mst197

Taylor JS, Raes J 2004. Duplication and divergence: the evolution of new genes and old ideas. Annu. Rev. Genet. 38: 615-643.

Tepass U 2012. The apical polarity protein network in *Drosophila* epithelial cells: regulation of polarity, junctions, morphogenesis, cell growth, and survival. Annual review of cell and developmental biology 28: 655-685.

Thompson KS, Vinson CR, Freire E 1993. Thermodynamic characterization of the structural stability of the coiled-coil region of the bZIP transcription factor GCN4. Biochemistry 32: 5491-5496. doi: 10.1021/bi00072a001

Veraksa A, McGinnis N, Li X, Mohler J, McGinnis W 2000. Cap 'n' collar B cooperates with a small Maf subunit to specify pharyngeal development and suppress deformed homeotic function in the Drosophila head. Development 127.

Vo Tommy V, Das J, Meyer Michael J, Cordero Nicolas A, Akturk N, Wei X, Fair Benjamin J, Degatano Andrew G, Fragoza R, Liu Lisa G, Matsuyama A, Trickey M, Horibata S, Grimson A, Yamano H, Yoshida M, Roth Frederick P, Pleiss Jeffrey A, Xia Y, Yu H 2016. A proteome-wide fission yeast interactome reveals network evolution principles from yeasts to human. Cell 164: 310-323. doi: http://doi.org/10.1016/j.cell.2015.11.037

Voordeckers K, Pougach K, Verstrepen KJ 2015. How do regulatory networks evolve and expand throughout evolution? Current Opinion in Biotechnology 34: 180-188. doi: http://doi.org/10.1016/j.copbio.2015.02.001

Yu X, Ivanic J, Memišević V, Wallqvist A, Reifman J 2011. Categorizing biases in high-confidence high-throughput protein-protein interaction data sets. Molecular & Cellular Proteomics 10. doi: 10.1074/mcp.M111.012500

Zhang K, Kaufman RJ 2008. From endoplasmic-reticulum stress to the inflammatory response. Nature 454: 455-462.

Zhang R, Rapin N, Ying Z, Shklanka E, Bodnarchuk TW, Verge VMK, Misra V 2013. Zhangfei/CREB-ZF – A potential regulator of the unfolded protein response.

PLoS ONE 8: e77256. doi: 10.1371/journal.pone.0077256

APPENDIX A: SCANNED DATABASES

Species queried for bZIP repertoire and corresponding databases used to collect bZIP candidates.

Saccharomyces cerevisiae	UCSC Genome Browser	https://genome.ucsc.edu/cgi-bin/hgGateway
Capsaspora owczarzaki	Sanger Institute	http://pfam.xfam.org/
Monosiga brevicollis	Broad Insititute, JGI	http://genome.jgi.doe.gov/Monbr1/Monbr1.home.html
Mnemopsis leydii	Mnmiopsis Genome Portal	https://research.nhgri.nih.gov/mnemiopsis/
Amphimedon queenslandica	Sanger Institute, Ensembl	http://pfam.xfam.org/, http://metazoa.ensembl.org/Amphimedon_queenslandica/Info/Index
Hydra magnipapillata	Metazome	https://metazome.jgi.doe.gov/mzmine/results.do?trail=%7Cquery
Acropora digitifera	OIST Marine Genomics	http://marinegenomics.oist.jp/coral/viewer/info?project_id=3
Nematostella vectensis	Sanger Institute, JGI, NCBI	http://pfam.xfam.org/, http://genome.jgi.doe.gov/Nemve1/Nemve1.home.htmlhttps://blast.ncbi.nlm.nih.gov/Blast.cgi
Capitella teleta	NCBI	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Lottia gigantea	NCBI	https://blast.ncbi.nlm.nih.gov/Blast.cgi

Caenorhabditis elegans	Sanger Institute, NCBI	http://pfam.xfam.org/, https://blast.ncbi.nlm.nih.gov/Blast.cgi
Drosophila melanogaster	Sanger Institute, NCBI	http://pfam.xfam.org/, https://blast.ncbi.nlm.nih.gov/Blast.cgi
Daphnia pulex	Sanger Institute, JGI, NCBI	http://pfam.xfam.org/, http://genome.jgi.doe.gov/Dappu1/Dappu1.home.htmlhttps://blast.ncbi.nlm.nih.gov/Blast.cgi
Strongylocentrotus purpuratus	NCBI	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Ciona intestinalis	Sanger Institute, NCBI	http://pfam.xfam.org/, https://blast.ncbi.nlm.nih.gov/Blast.cgi
Branchiostoma floridae	NCBI	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Mus musculus	UCSC Genome Browser	https://genome.ucsc.edu/cgi-bin/hgGateway
Homo sapiens	UCSC Genome Browser	https://genome.ucsc.edu/cgi-bin/hgGateway

APPENDIX B: ALL BZIP SEQUENCES QUERIED

Please reference SupplementaryTable2.xlsx, Excel Spreadsheet, 11 KB

APPENDIX C: OMITTED BZIP SEQUENCES

24 bZIP candidates removed from phylogenetic analysis after manual inspection. All sequences collected from pfam database unless otherwise indicated

Please reference SupplementaryTable3.xslx, Excel spreadsheet, 11 KB

APPENDIX D: TOPOLOGY COMPARISON WITH LITERATURE

Comparison of subfamily presence and count between this paper and the Jindrich and Degnan bZIP repertoire

	CEBP	PAR	JUN	FOS/ATF3	ATF2	ATF4	OASIS	ATF6	NFE2/BACH	MAF	ATF1/CREB	XBP1	Zhangfei	CREBL2
	(KMR/JD)	**(KMR/JD)**	**(KMR/JD)**	**(KMR/JD)**	**(KMR/JD)**	**(KMR/JD)**	**(KMR/JD)**	**(KMR/JD)**	**(KMR/JD)**	**(KMR/JD)**	**(KMR/JD)**	**(KMR/JD)**	**(KMR/JD)**	**(KMR/JD)**
ad	(3/2)	(2/3)	(3/1)	(1/1)	(0/0)	(1/1)	(2/3)	(0/0)	(1/1)	(4/4)	(1/1)	(1/1)	(0/NA)	(0/NA)
aq	(2/2)	(2/2)	(1/1)	(1/1)	(1/1)	(1/1)	(2/2)	(1/1)	(2/2)	(1/1)	(1/1)	(1/1)	(2/NA)	(0/NA)
bf	(1/1)	(5/5)	(1/1)	(1/1?)	(0/0)	(1/1)	(2/2)	(1/1)	(1/1)	(2/3)	(0/0)	(1/1)	(0/NA)	(1/NA)
ce	(2/NA)	(5/NA)	(1/NA)	(1/NA)	(1/NA)	(2/NA)	(1/NA)	(1/NA)	(1/NA)	(0/NA)	(1/NA)	(1/NA)	(0/NA)	(0/NA)
ci	(0/NA)	(0/NA)	(1/NA)	(2/NA)	(1/NA)	(1/NA)	(2/NA)	(0/NA)	(2/NA)	(2/NA)	(2/NA)	(4/NA)	(1/NA)	(0/NA)
co	(1/1?)	(1/1?)	(1/1?)	(1/1?)	(1/1)	(2/0)	(0/1)	(0/1?)	(0/0)	(0/0)	(1/1)	(0/0)	(1/NA)	(1/NA)
ct	(1/2)	(7/7)	(1/1)	(2/1)	(1/1)	(0/1)	(2/2)	(1/1)	(1/1)	(2/2)	(1/1)	(1/1)	(1/NA)	(1/NA)
dm	(1/1)	(6/6)	(1/1)	(2/2)	(0/0)	(2/1)	(1/1)	(1/1)	(1/1)	(0/1)	(1/1)	(1/1)	(0/NA)	(1/NA)
dp	(2/3)	(4/5)	(1/1)	(2/2)	(1/1?)	(1/1)	(2/2)	(1/1)	(2/1)	(0/2)	(1/1)	(1/1)	(2/NA)	(0/NA)
hm	(2/2)	(3/3)	(2/1)	(2/3)	(1/1)	(1/1)	(2/2)	(1/1)	(1/1)	(2/2)	(1/1)	(1/2)	(0/NA)	(0/NA)
hs	(5/6)	(4/4)	(3/3)	(8/8)	(3/3)	(2/2)	(5/5)	(2/2)	(6/6)	(6/6)	(3/3)	(1/1)	(1/NA)	(1/NA)
lg	(3/3)	(5/8)	(1/1)	(3/2)	(1/1)	(1/1)	(3/3)	(1/1)	(1/1)	(1/2)	(1/1)	(1/1)	(2/NA)	(0/NA)
mb	(2/2)	(1/1)	(0/0)	(0/0)	(1/1)	(0/0)	(2/2?)	(1/1)	(0/0)	(0/0)	(0/0)	(0/0)	(1/NA)	(0/NA)
ml	(1/1)	(0/0)	(1/1)	(2/2)	(1/1)	(0/0)	(2/2)	(2/1)	(0/1?)	(0/1)	(1/1)	(1/1)	(0/NA)	(0/NA)
mm	(4/NA)	(4/NA)	(3/NA)	(8/NA)	(4/NA)	(2/NA)	(4/NA)	(1/NA)	(6/NA)	(4/NA)	(3/NA)	(1/NA)	(2/NA)	(1/NA)
nv	(2/2)	(8/3)	(3/2)	(4/3)	(2/1)	(1/0)	(2/2)	(1/1)	(1/1)	(1/3)	(1/1)	(1/1)	(2/NA)	(1/NA)
sp	(2/2)	(5/5)	(1/1)	(2/3)	(1/1)	(0/0)	(2/2)	(1/1)	(1/1)	(2/2)	(1/1)	(1/1)	(2/NA)	(1/NA)
ta	(1/1)	(3/2)	(1/1)	(0/0)	(1/1)	(0/1)	(1/1)	(1/1)	(1/1)	(1/2)	(1/1)	(1/1)	(0/NA)	(0/NA)

APPENDIX E: HOMODIMER PREDICTIONS

Count of homodimers formed by bZIP subfamilies in each queried species. The number on the right of each column corresponds to number of orthologues and the number on the left corresponds to the number of homodimers.

	CEBP	PAR	JUN	FOS	ATF2	ATF4	OASIS	ATF6	NFE2	MAF	ATF1	XBP1	Zhangfei	CREBL2
ad	(3/3)	(2/2)	(2/3)	(1/1)	(0/0)	(1/1)	(2/2)	(0/0)	(0/1)	(4/4)	(0/1)	(1/1)	(0/0)	(0/0)
aq	(2/2)	(1/2)	(1/1)	(0/1)	(1/1)	(0/1)	(2/2)	(1/1)	(1/2)	(1/1)	(1/1)	(1/1)	(2/2)	(0/0)
bf	(1/1)	(4/5)	(0/1)	(0/1)	(0/0)	(1/1)	(2/2)	(1/1)	(0/1)	(2/3)	(0/0)	(1/1)	(0/0)	(1/1)
ce	(2/2)	(5/5)	(1/1)	(0/1)	(1/1)	(2/2)	(1/1)	(1/1)	(0/1)	(0/0)	(1/1)	(1/1)	(0/0)	(0/0)
ci	(0/0)	(0/0)	(0/1)	(1/2)	(1/1)	(0/1)	(2/2)	(0/0)	(0/2)	(2/2)	(2/2)	(3/4)	(0/1)	(0/0)
co	(1/1)	(1/1)	(1/1)	(1/1)	(1/1)	(3/3)	(0/0)	(0/0)	(0/0)	(0/0)	(1/1)	(0/0)	(1/1)	(1/1)
ct	(1/2)	(6/7)	(0/1)	(1/2)	(1/1)	(0/1)	(2/2)	(1/1)	(1/1)	(2/2)	(1/1)	(1/1)	(1/1)	(0/1)
dm	(1/1)	(5/6)	(1/1)	(0/2)	(0/0)	(2/2)	(0/1)	(1/1)	(0/1)	(0/1)	(1/1)	(1/1)	(0/0)	(0/1)
dp	(2/3)	(3/4)	(1/1)	(0/2)	(1/1)	(1/1)	(2/2)	(1/1)	(0/2)	(0/2)	(1/1)	(0/1)	(2/2)	(0/0)
hm	(2/2)	(3/3)	(1/2)	(0/1)	(1/1)	(1/1)	(0/2)	(1/1)	(0/1)	(2/2)	(1/1)	(1/1)	(0/0)	(0/0)
hs	(5/5)	(4/4)	(1/3)	(3/8)	(3/3)	(1/2)	(5/5)	(2/2)	(2/6)	(6/6)	(3/3)	(1/1)	(1/1)	(1/1)
lg	(2/3)	(3/4)	(0/1)	(2/3)	(1/1)	(1/1)	(2/3)	(1/1)	(0/1)	(1/2)	(1/1)	(1/1)	(2/2)	(0/0)
mb	(2/2)	(1/1)	(0/0)	(0/0)	(1/1)	(0/0)	(2/2)	(1/1)	(0/0)	(0/0)	(0/0)	(0/0)	(1/1)	(0/0)
ml	(1/1)	(0/0)	(1/1)	(2/2)	(1/1)	(0/0)	(2/2)	(2/2)	(0/0)	(0/1)	(1/1)	(0/1)	(0/0)	(0/0)
mm	(4/4)	(4/4)	(1/3)	(2/8)	(3/4)	(1/2)	(4/4)	(1/1)	(2/6)	(3/4)	(3/3)	(1/1)	(1/1)	(0/1)
nv	(2/2)	(7/8)	(2/3)	(0/4)	(2/2)	(1/1)	(2/2)	(1/1)	(1/1)	(1/3)	(1/1)	(1/1)	(1/2)	(0/1)
sp	(2/2)	(4/5)	(0/1)	(1/2)	(1/1)	(0/0)	(2/2)	(1/1)	(0/1)	(2/2)	(1/1)	(1/1)	(2/2)	(0/1)
ta	(1/1)	(2/3)	(0/1)	(0/0)	(1/1)	(0/1)	(1/1)	(1/1)	(0/1)	(1/2)	(1/1)	(1/1)	(0/0)	(0/0)

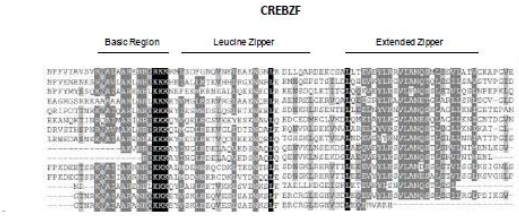
APPENDIX F: FONG ALGORITHM PREDICTIONS

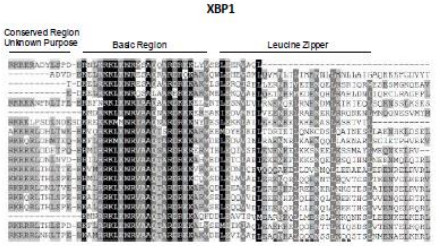
Interactome calculations based on the stringent Fong algorithm. Yellow bars represent heterodimers while green bars represent homodimers.

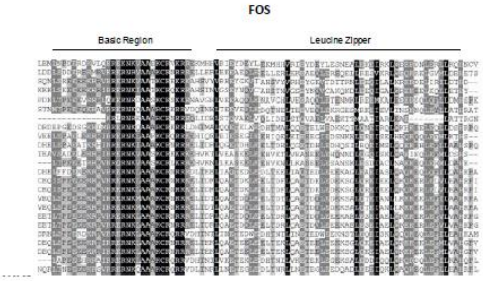
Please reference SupplementaryTable6.xslx, Excel spreadsheet, 11 KB

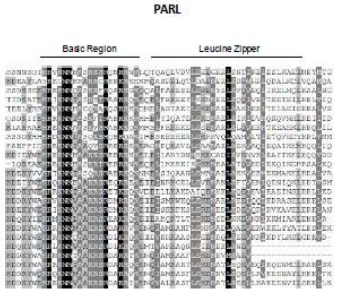
APPENDIX G: BZIP SIBFAMILY-SPECIFIC ALIGNMENTS

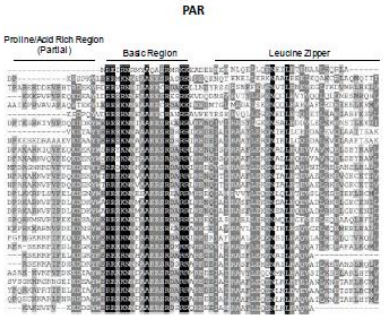
Alignments of the bZIP and other specific domains of each bZIP subfamily used for manual inspection.

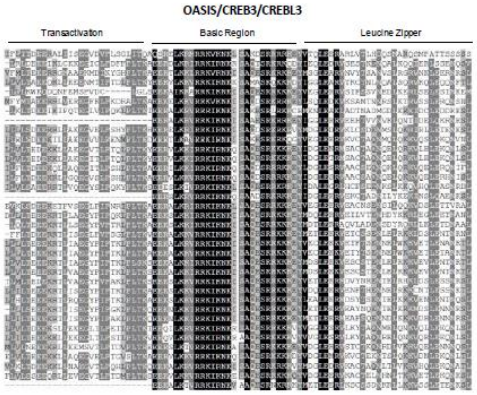






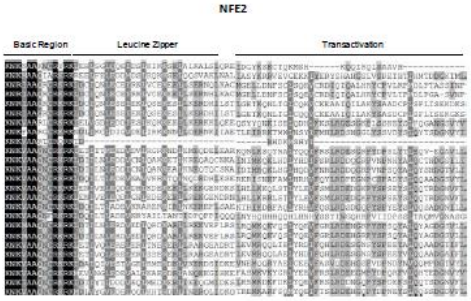




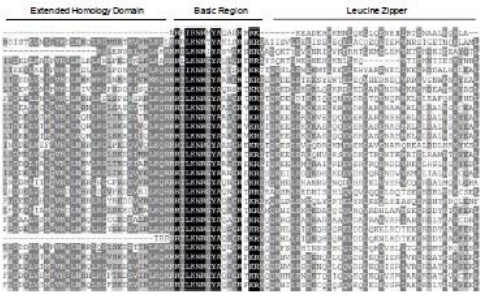


NFIL3

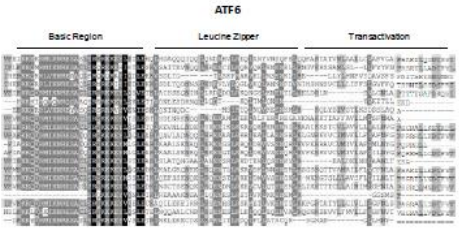
NFE2



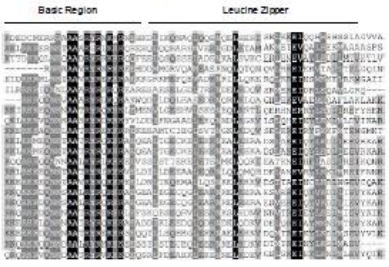
MAF



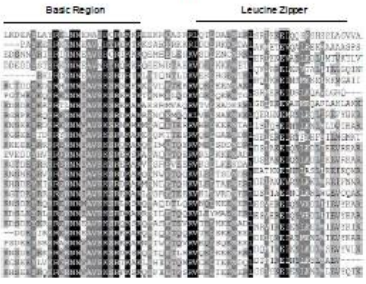
ATF6



ATF4

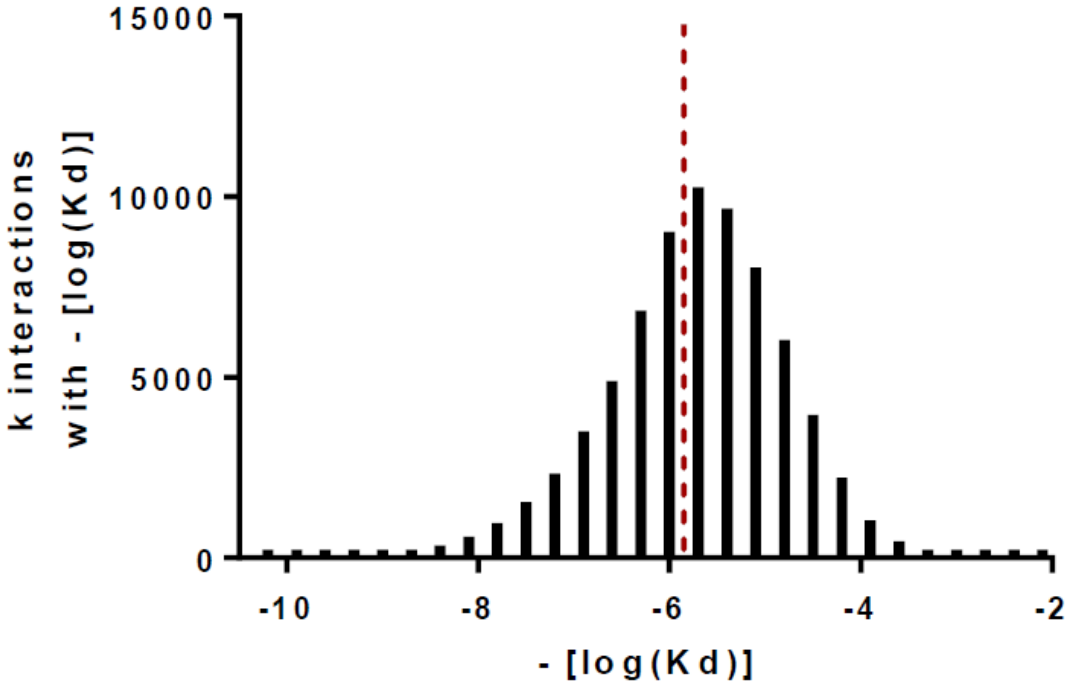


C/EBP



APPENDIX H: DISTRIBUTION OF DIMER PREDICTION SCORES

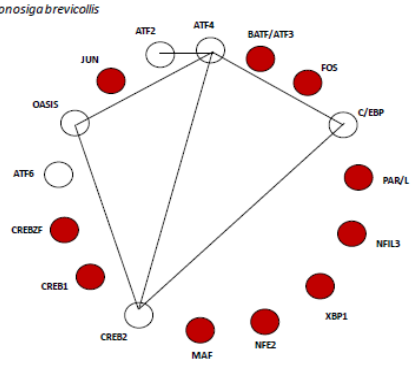
Distribution of the interaction scores called by the Potapov algorithm. The red dotted line indicates the interaction cut off value.



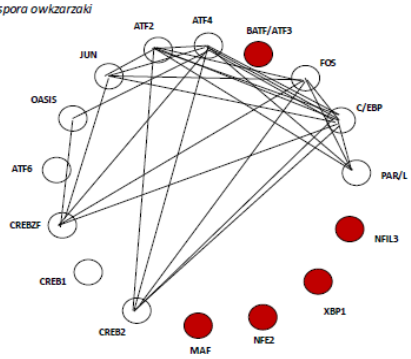
APPENDIX I: ALL INTERACTOME MAPS

Interactome maps of each species, tracking heterodimerization dynamics. Red circles denote the absence of a bZIP subfamily.

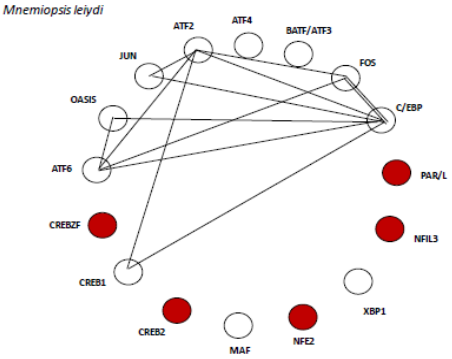
Monosiga brevicollis



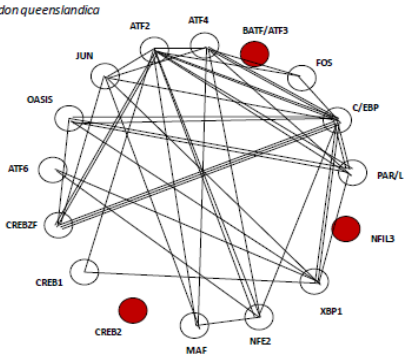
Capsaspora owkzarzaki



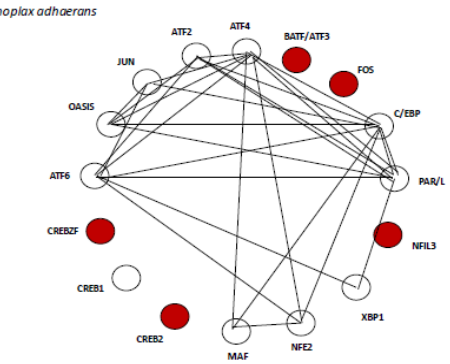
Mnemiopsis leiydi



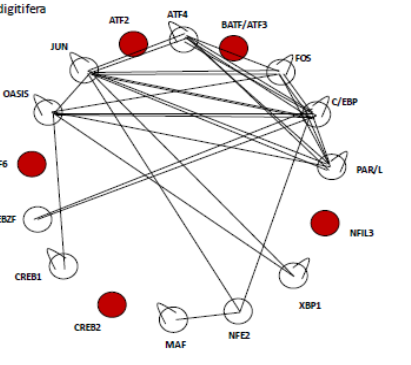
Amphimedon queenslandica



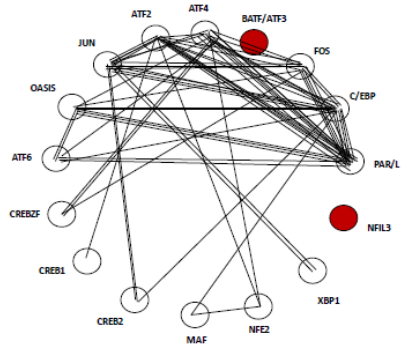
Trichoplax adhaerans



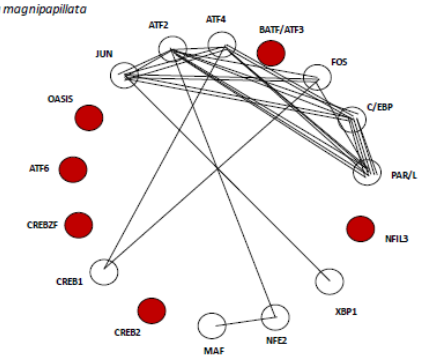
Acropora digitifera



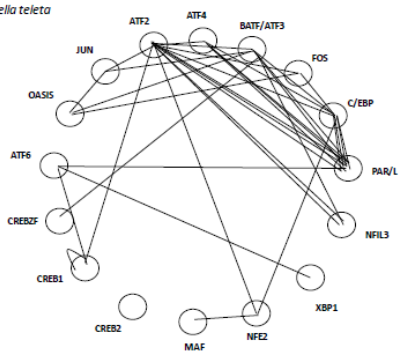
Nematostella vectensis



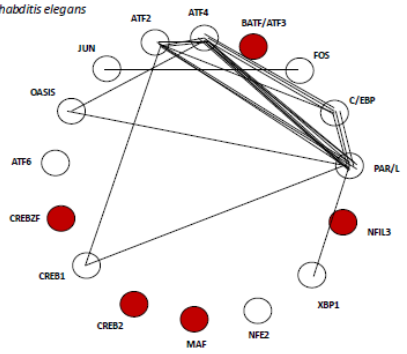
Hydra magnipapillata



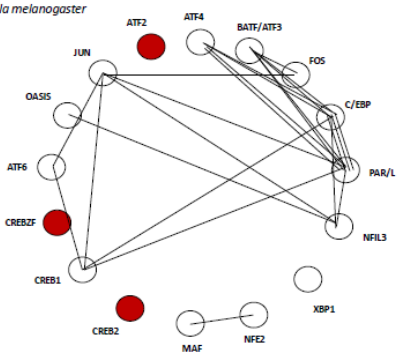
Capitella teleta



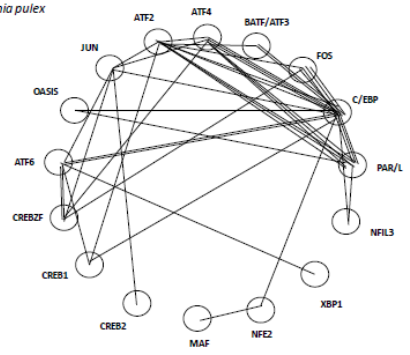
Caenorhabditis elegans



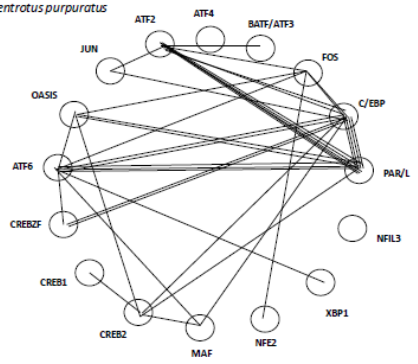
Drosophila melanogaster



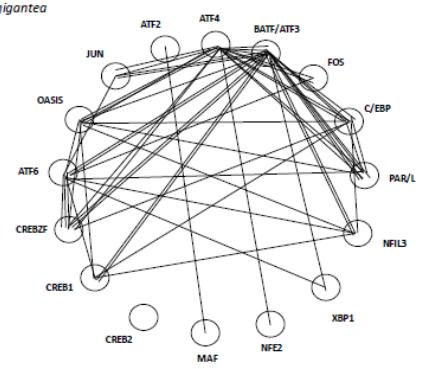
Daphnia pulex



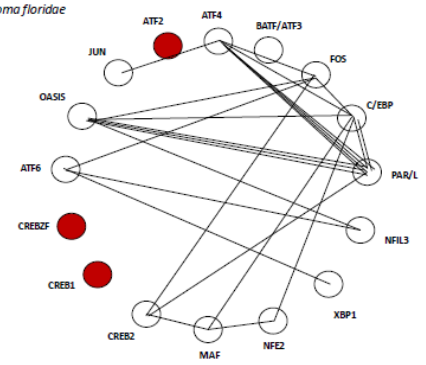
Strongylocentrotus purpuratus



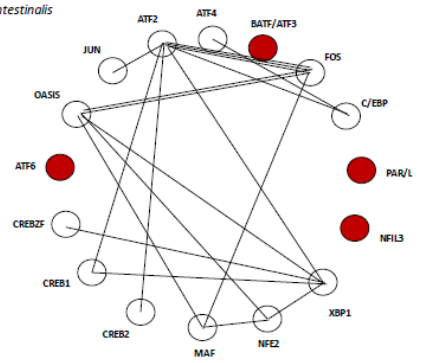
Lottia gigantea



Branchiostoma floridae



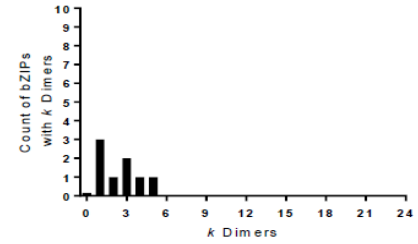
Ciona intestinalis



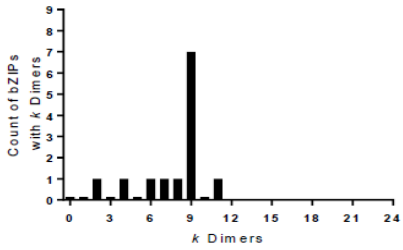
APPENDIX J: SPECIES-SPECIFIC DISTRIBTION OF BZIP PROMISCUITY

Count of genomic bZIPs that form a given amount of dimers in each species.

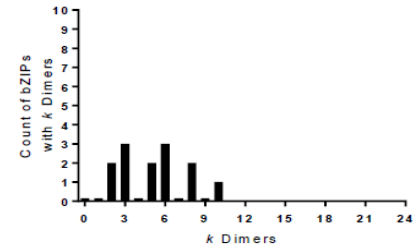
Monosiga brevicollis



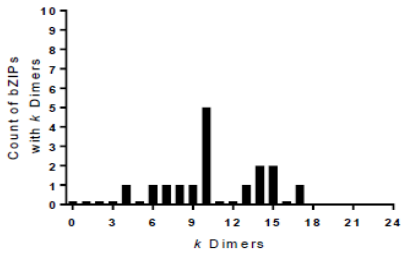
Capsaspora owczarzaki



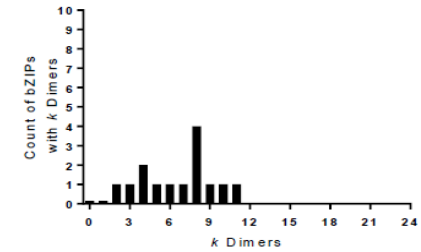
Mnemiopsis leiydi



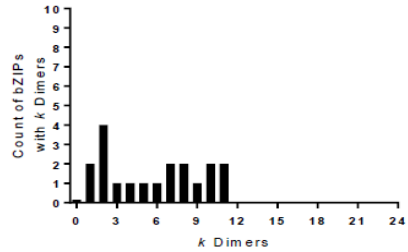
Amphimedon queenslandica



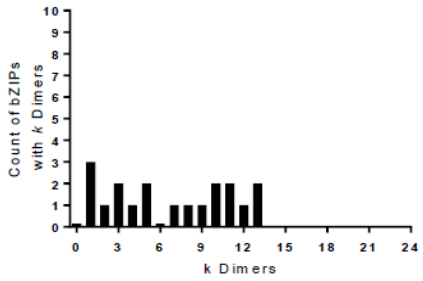
Trichoplax adhaerens



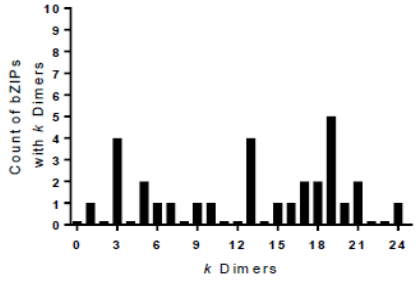
Hydra magnipapillata



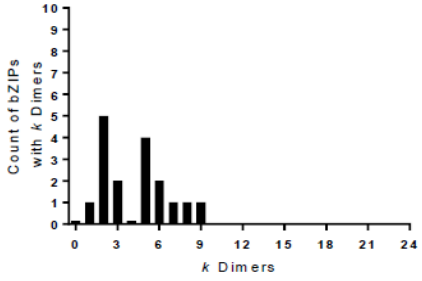
Acropora digitifera



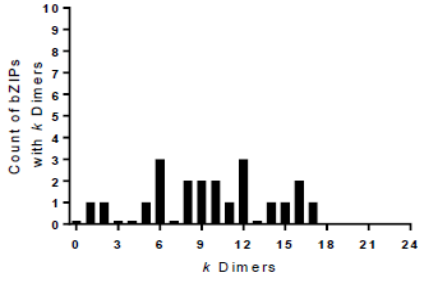
Nematostella vectensis



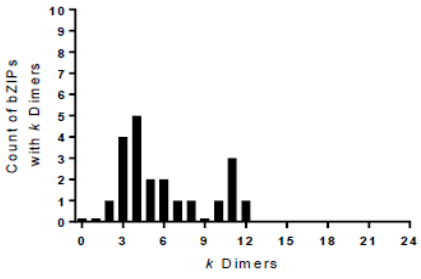
Capitella teleta



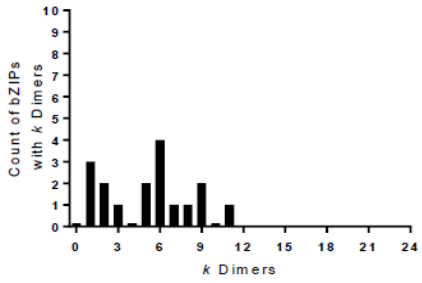
Lottia gigantea



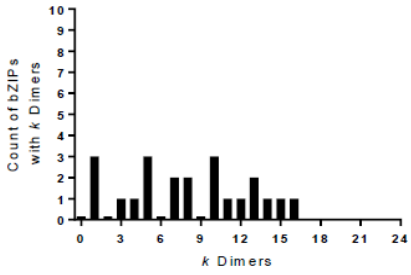
Caenorhabditis elegans



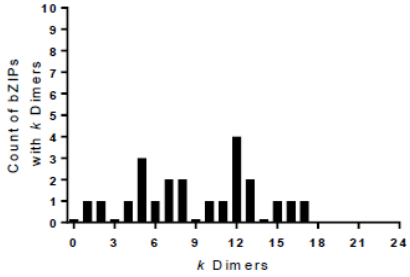
Drosophila melanogaster



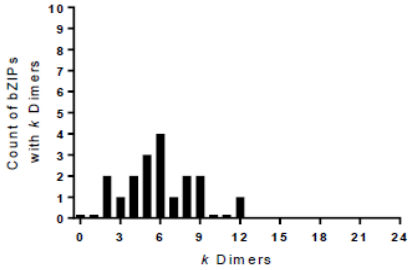
Daphnia pulex



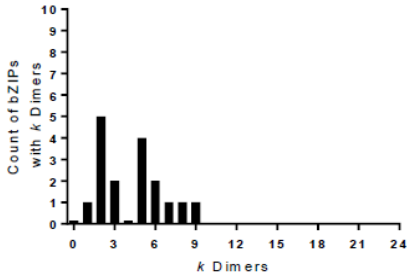
Strongylocentrotus pupuratus



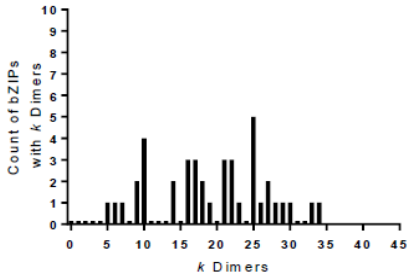
Branchiostoma floridae



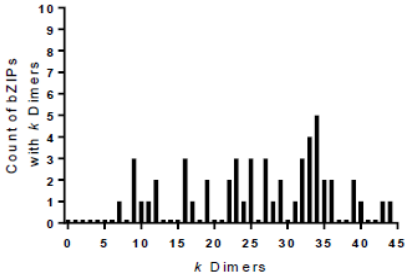
Ciona intestinalis



Mus musculus

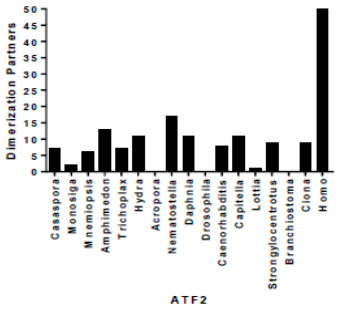


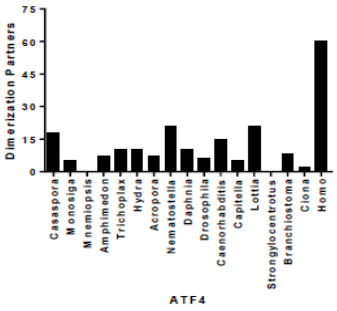
Homo sapiens

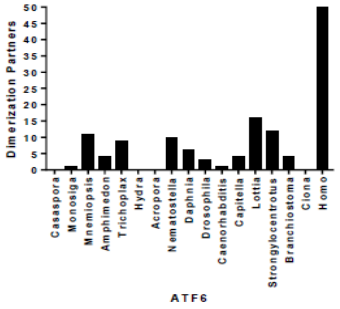


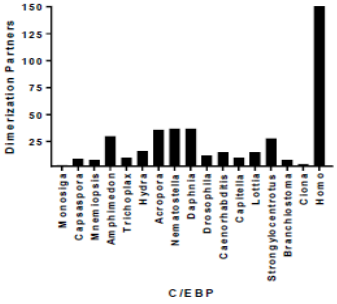
APPENDIX K: SUBFAMILY-SPECIFIC DIMER ACTIVITY ACROSS SPECIES

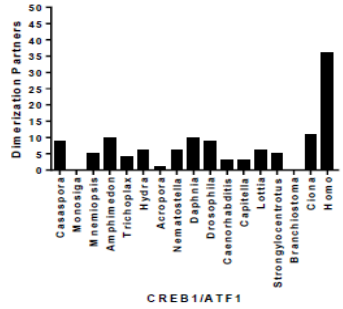
Distribution of subfamily-specific dimerization capacity in each species.

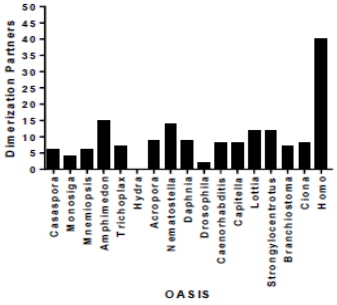


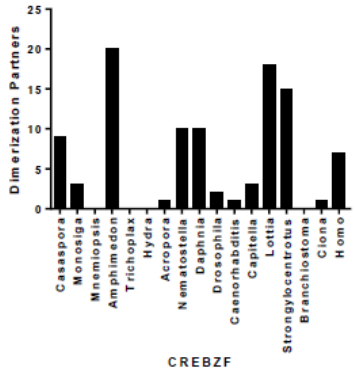


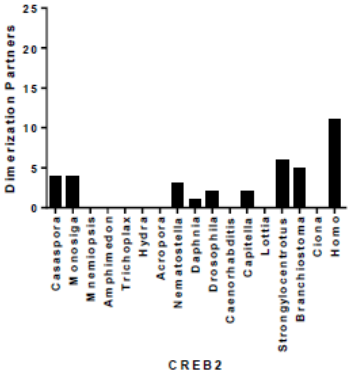


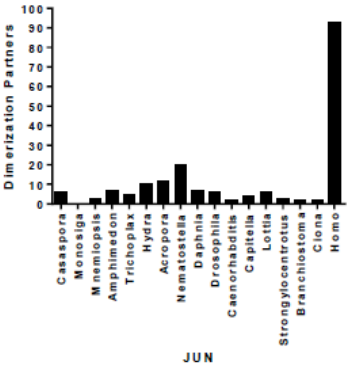


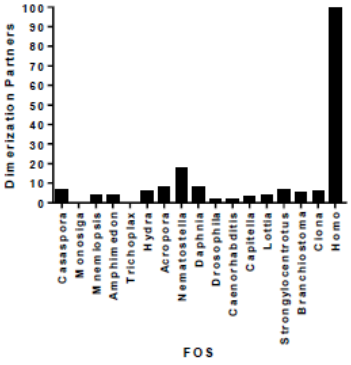


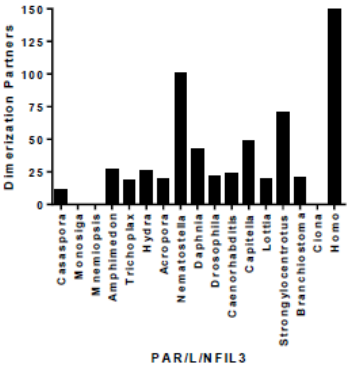


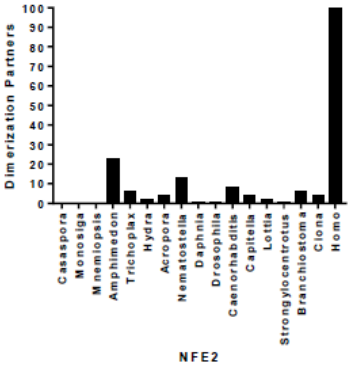

Dimerization Partners
0
5
10
15
20
25
Casaspora
Monosiga
Mnemiopsis
Amphimedon
Trichoplax
Hydra
Acropora
Nematostella
Daphnia
Drosophila
Caenorhabditis
Capitella
Lottia
Strongylocentrotus
Branchiostoma
Ciona
Homo
CREBZF

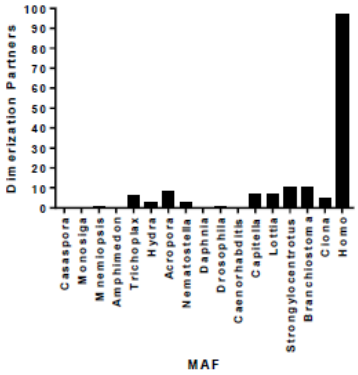

Dimerization Partners
0
5
10
15
20
25
Casaspora
Monosiga
Mnemiopsis
Amphimedon
Trichoplax
Hydra
Acropora
Nematostella
Daphnia
Drosophila
Caenorhabditis
Capitella
Lottia
Strongylocentrotus
Branchiostoma
Ciona
Homo
CREB2

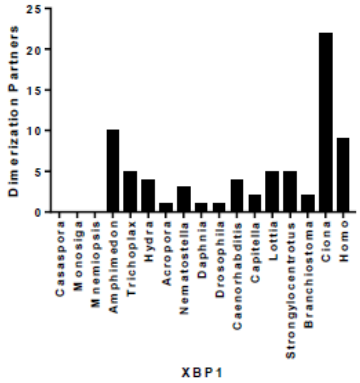

Dimerization Partners
0
10
20
30
40
50
60
70
80
90
100
Casaspora
Monosiga
Mnemiopsis
Amphimedon
Trichoplax
Hydra
Acropora
Nematostella
Daphnia
Drosophila
Caenorhabditis
Capitella
Lottia
Strongylocentrotus
Branchiostoma
Ciona
Homo
JUN


Dimerization Partners
0
10
20
30
40
50
60
70
80
90
100
Casaspora
Monosiga
Mnemiopsis
Amphimedon
Trichoplax
Hydra
Acropora
Nematostella
Daphnia
Drosophila
Caenorhabditis
Capitella
Lottia
Strongylocentrotus
Branchiostoma
Ciona
Homo
FOS


Dimerization Partners
0
25
50
75
100
125
150
Casaspora
Monosiga
Mnemiopsis
Amphimedon
Trichoplax
Hydra
Acropora
Nematostella
Daphnia
Drosophila
Caenorhabditis
Capitella
Lottia
Strongylocentrotus
Branchiostoma
Ciona
Homo
PAR/L/NFIL3


Dimerization Partners
0
10
20
30
40
50
60
70
80
90
100
Casaspora
Monosiga
Mnemiopsis
Amphimedon
Trichoplax
Hydra
Acropora
Nematostella
Daphnia
Drosophila
Caenorhabditis
Capitella
Lottia
Strongylocentrotus
Branchiostoma
Ciona
Homo
NFE2


Dimerization Partners
100
90
80
70
60
50
40
30
20
10
0
Casaspora
Monosiga
Mnemiopsis
Amphimedon
Trichoplax
Hydra
Acropora
Nematostella
Daphnia
Drosophila
Caenorhabditis
Capitella
Lottia
Strongylocentrotus
Branchiostoma
Ciona
Homo
MAF


Dimerization Partners
25
20
15
10
5
0
Casaspora
Monosiga
Mnemiopsis
Amphimedon
Trichoplax
Hydra
Acropora
Nematostella
Daphnia
Drosophila
Caenorhabditis
Capitella
Lottia
Strongylocentrotus
Branchiostoma
Ciona
Homo
XBP1