Transcription Factors and the Origin of Animal Multicellularity

Arnau Sebé-Pedrós and Alex de Mendoza

Abstract The transition from a unicellular life-style, with temporal cell differentiation to a multicellular life-style, with both temporal and spatial cell differentiation, required the expansion of the regulatory capabilities of ancient animals. In this chapter, we describe how this change occurred from the perspective of transcription factor (TF) evolution. First, we revise TF diversification throughout eukaryotes. We trace the evolutionary origins of major TF classes and describe general patterns of TF content and how they correlate with multicellular life-styles in eukaryotes. We then focus on the animal TF developmental toolkit at the gene family level. Many of the metazoan developmental TFs originated in a unicellular context; yet there are also many TFs that evolved at the onset of Metazoa. Finally, we describe the changes that led to the establishment of gene regulatory networks that control animal multicellularity and review different case-examples that have provided illuminating insights into this question.

Keywords *Cis* evolution · *Trans* evolution · Co-option · Gene regulatory networks · Developmental toolkit · Choanoflagellates · *Capsaspora owczarzaki* · Porifera

Introduction

Animal development involves the orchestrated deployment of gene batteries to control spatial and temporal cell differentiation. Conversely, unicellular life cycles mainly require temporal changes, in order to regulate the transitions from one life stage to another or to regulate metabolic activities and responses to environmental cues. Moreover, control over cell proliferation (in order to minimize the emergence of non-cooperating cells) is a critical requirement in multicellular lineages (Grosberg and Strathmann 2007). Transcription factors (TFs) are key players in these processes, as they bind DNA in a sequence-specific manner and enhance or repress gene expression. Indeed, it has long been hypothesized that the complexity of the transcription

A. de Mendoza (✉) · A. S.-Pedrós
Institut de Biologia Evolutiva, CSIC-Universitat Pompeu Fabra, Passeig Marítim de la Barceloneta 37–49, 08003 Barcelona, Spain
e-mail: alexmendozasoler@gmail.com

I. Ruiz-Trillo, A. M. Nedelcu (eds.), *Evolutionary Transitions to Multicellular Life,* Advances in Marine Genomics 2, DOI 10.1007/978-94-017-9642-2_18

regulation system is correlated with organismal complexity (Levine and Tjian 2003). TFs can be viewed as regulatory hubs, where the information processed by the cell must be translated into biogenesis. For this reason, TFs are often key downstream targets of the major metazoan developmental signaling pathways (Pires-daSilva and Sommer 2003). Taking all this into account, it has been hypothesized that the onset of animal multicellularity was accompanied by an increase in transcription factor diversity (Rokas 2008).

Despite the striking diversity of body plans found in metazoans, there is a transcription factor toolkit largely shared by all metazoans (Vaquerizas et al. 2009; Degnan et al. 2009). This toolkit is known to be involved in embryonic development in both bilaterian and non-bilaterian animal lineages (Technau and Steele 2011; Adamska et al. 2011, Chapter "The Evolution of Developmental Signalling in Dictyostelia from an Amoebozoan Stress Response"). Therefore, animal TFs represent a shared patterning language that facilitate multicellularity and govern development. This fact brings up several questions. Since multicellularity evolved more than once in the history of eukaryotes, can we identify similarities or differences between animal multicellularity and other multicellularities in terms of transcription factor toolkits? When did the metazoan TF types emerge and how did they evolve? Finally, if metazoan TFs were present in the unicellular ancestors of Metazoa, how were they adapted for use in the context of a multicellular developmental program?

A Natural History of Eukaryotic Transcription Factors

The acquisition of multicellularity by eukaryotes is a story of evolutionary convergence, since at least 26 independent transitions to multicellularity have been reported (Grosberg and Strathmann 2007; Parfrey and Lahr 2013, Chapter "Timing the Origins of Multicellular Eukaryotes Through Phylogenomics and Relaxed Molecular Clock Analyses"). If we restrict our definition of multicellularity to complex multicellularity, which involves extensive spatial cellular differentiation (see Chapter "Independent Emergence of Complex Multicellularity in the Brown and Red Algae" for a discussion of complex multicellularity), we can identify at least seven independent groups that have acquired complex multicellularity, including animals, plants, fungi, and brown and red algae (Knoll 2011; Niklas and Newman 2013). Thus, it is possible to analyze the sequenced genomes of species belonging to complex multicellular lineages and compare their transcription factor repertoires with those of the closely related unicellular species. Indeed, eukaryotes are known to exhibit great diversity in terms of TF abundance and TF type composition (Weirauch and Hughes 2011), which we define as the TFome, and by looking at patterns of abundance and diversity we can infer trends associated with multicellularity.

We recently performed such analyses and identified some general trends in the total abundance of TFs (Fig. 1; de Mendoza et al. 2013). For example, taxa with complex multicellular development taxa, i.e. Embryophyta and Metazoa, present a dramatic increase in TF numbers compared to other eukaryotes. Moreover, the morphologically simpler forms within these groups (sponges, mosses, etc.) have fewer

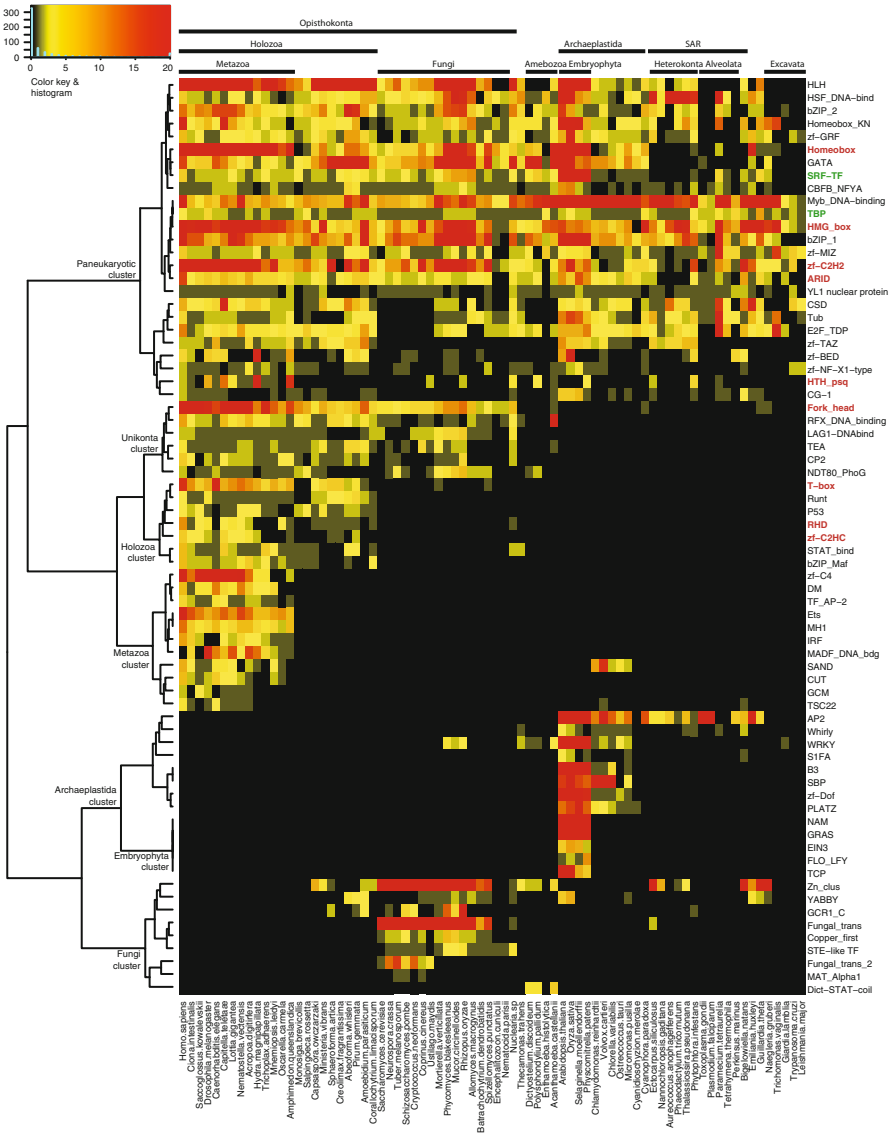


Fig. 1 Presence and abundance of transcription factors (TFs) in eukaryotes. The heat map depicts absolute TF counts according to the colour scale. TFs were identified using the HMM profiles of their respective DNA-binding domains. TF types (*rows*) are clustered according to abundance and distribution, and species (*columns*) are grouped according to phylogenetic affinity. Major eukaryotic lineages are indicated (*Top*). (Adapted from de Mendoza et al. 2013)

TFs (de Mendoza et al. 2013). Irrespective of their phylogenetic relationships, the lowest numbers of TFs are observed in parasitic eukaryotes, an example of convergent simplification (Iyer et al. 2008; de Mendoza et al. 2013). Nonetheless, some species described as parasitic or symbiotic, such as most ichthyosporeans and the filasterean *Capsaspora owczarzaki* (both close relatives of metazoans; see Chapter "Filastereans and Ichthyosporeans: Models to Understand the Origin of Metazoan Multicellularity") have a relatively rich TF repertoire, suggesting a more complex life cycle or an unknown free-living stage. Paradoxically, choanoflagellates, which are free-living and in some cases colonial and prey-catching organisms (Chapter "Choanoflagellates: Perspective on the Origin of Animal Multicellularity"), have a lower total number of TFs compared to related lineages (such as filastereans and ichthyosporeans). Another factor that can explain particularly rich TF repertoires is the occurrence in some groups of Whole Genome Duplications (WGD) (such as those found in vertebrates, zygomycetes, the ciliate *Paramecium* and embryophytes; also see Chapter "The Evolution of Transcriptional Regulation in the Viridiplantae and its Correlation with Morphological Complexity") (Maere et al. 2005; Van De Peer et al. 2009). WGD tends to lead to the deletion of excess copies of duplicated genes; however, the TFs are one of the most resilient genes to this type of loss (Van De Peer et al. 2009; De Smet et al. 2013). Therefore life-style and genome dynamics influence the total number of TFs in eukaryotic genomes.

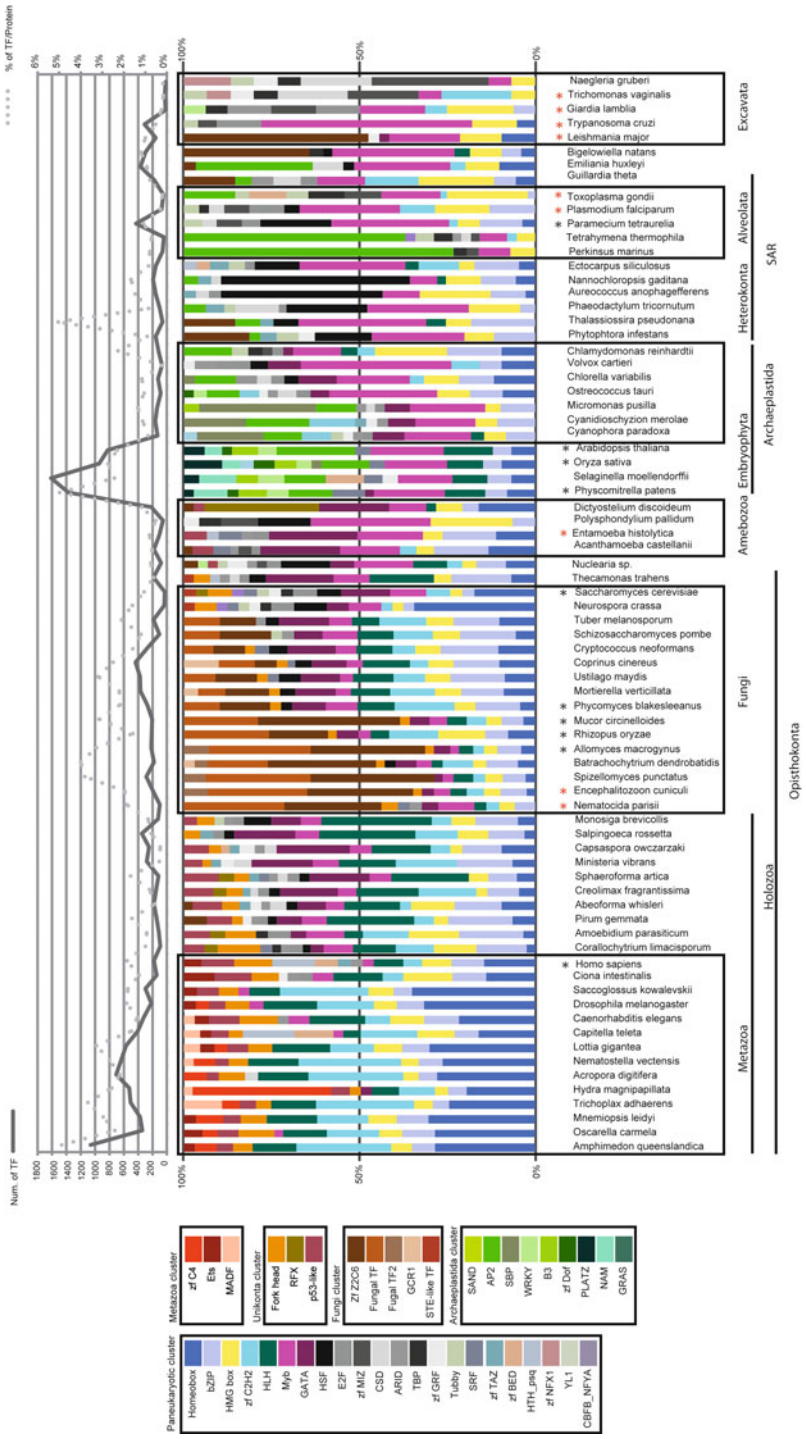
Based on their phylogenetic distribution, TFs can be divided into two groups (Fig. 1): those that are paneukaryotic and those that are lineage-specific. The first group includes TFs that have undergone independent expansions in several lineages, such as Myb_DNA-binding, HLH, GATA, SRF-TF, bZIP, Homeobox, HMGbox and zf-C2H2 (Fig. 1). The expansion of ancient TF families was an evolutionary source of innovation that allowed diversification of the TFome, in some cases related to multicellularity (de Mendoza et al. 2013). The second group includes plant-, fungal- and animal-specific clusters of TFs, indicating that some lineages have evolved their own repertoires of TFs by diversifying their ancestral TFome. This correlation between clusters of lineage-specific TFs and eukaryotic groups with multicellular lineages might also be due to sampling bias as those lineages represent the most-sampled taxonomic groups. In these groups, more genomes are available and, more importantly, many TF families have been investigated experimentally. Thus, new specific TF clusters are likely to be found in other eukaryotic groups, such as stramenopiles or rhizarians, once their TFome are surveyed. The absence of experimental studies may also explain some exceptionally TF-poor taxa in which unknown TF families may have an important role. The case of the IBD family unique to *Trichomonas vaginalis* provides a good example of this (Iyer et al. 2008). It seems clear that ancestral TF types evolved dynamically through gene family expansions in some lineages, while new TFs were added to the ancestral repertoire in the lineages that led to fungi, plants and animals.

As we discussed above, there are strong phylogenetic patterns of TF diversity, in terms of both lineage-specificity and abundance. In addition, there are patterns in the relative contribution of each TF type in the TFome of each species, measured as the number of genes of each TF class as a percentage of the total number of TFs in

that genome (Fig. 2). For example, animal genomes are dominated by Homeobox, zf-C2H2 and bHLH TFs (accounting for more than 50 % of the TFome), whereas unicellular holozoans have a distinct TFome profile, indicating that the transition to multicellularity involved a system-level change in TF type proportions (de Mendoza et al. 2013). The TF types that became predominant in animal genomes are mainly those involved in patterning and differentiation in animals, a structural requirement for an organism with diverse cell types (Degnan et al. 2009; Sebé-Pedrós et al. 2011). Conversely, p53-like TFs (e.g. T-box, Runx, p53 and others) or bZIP seem to be proportionally more important in non-bilaterian metazoans, with less cell types and simpler patterning. In contrast, they represent a small percentage in bilaterian lineages. The higher proportion of p53-like or bZip TFs in non-bilaterians and unicellular holozoan TFomes could reflect their potential role in regulating house-keeping functions, such as metabolic processes, proliferation or immunity response, although T-box is an interesting exception (Hammonds et al. 2013; Sebé-Pedrós et al. 2013).

Finally, the simplicity of the TFome of other complex multicellular eukaryote lineages, mainly red algae and brown algae (phaeophytes) (discussed in Chapter "Independent Emergence of Complex Multicellularity in the Brown and Red Algae"), is an intriguing question. Indeed, there is no expansion of total number of TFs when comparing the multicellular brown algae *Ectocarpus silicosus* and its unicellular relative *Nanochloropsis gaditana* (Cock et al. 2010; Radakovits et al. 2012). Judging from the large number of unique TFs in other complex multicellular lineages (plants and animals), there are likely to be undiscovered stramenopile- or brown-algae-specific TFs, but there is a dearth of functional studies in this group (Peters et al. 2008; Coelho et al. 2011). It is worth mentioning that *E. siliculosus* does not have a typical embryonic development, but rather a modular growth strategy, although this is not the case for other multicellular brown algae, such as *Fucus spiralis* (Bouget et al. 1998). We hypothesize that a richer TF repertoire will be found in brown algae with embryonic development. The case of red algae is similar, since the TFomes of both *Pyropia yezoensis* and *Chondrus crispus* are surprisingly simple compared to the unicellular red algae *Cyanidioschyzon merolae* (Nakamura et al. 2013; Collen et al. 2013).

The simplicity of the TF repertoires of some multicellular eukaryotes may be explained by their modes of development, although the identification of new lineage-specific TF types remains a critical issue to be resolved through future research in these groups. We can conclude that complex multicellularity is associated with enrichment of the TF toolkit (both in terms of abundance and innovation) in lineages with complex embryonic development: plants and animals. However, this toolkit is greatly influenced by the TF repertoires of their respective unicellular ancestors (de Mendoza et al. 2013). In the next section, we will focus on the early evolution of animal TF families in order to gain further insights into which genes evolved in a unicellular context.


% of TF/Protein
Num. of TF
Naegleria gruberi
Trichomonas vaginalis
Giardia lamblia
Trypanosoma cruzi
Leishmania major
Bigelowiella natans
Emiliania huxleyi
Guillardia theta
Toxoplasma gondii
Plasmodium falciparum
Paramecium tetraurelia
Tetrahymena thermophila
Perkinsus marinus
Ectocarpus siliculosus
Nannochloropsis gaditana
Aureococcus anophagefferens
Phaeodactylum tricornutum
Thalassiosira pseudonana
Phytophtora infestans
Chlamydomonas reinhardtii
Volvox cartieri
Chlorella variabilis
Ostreococcus tauri
Micromonas pusilla
Cyanidioschyzion merolae
Cyanophora paradoxa
Arabidopsis thaliana
Oryza sativa
Selaginella moellendorffii
Physcomitrella patens
Dictyostelium discoideum
Polysphondylium pallidum
Entamoeba histolytica
Acanthamoeba castellanii
Nucleariа sp.
Thecamonas trahens
Saccharomyces cerevisiae
Neurospora crassa
Tuber melanosporum
Schizosaccharomyces pombe
Cryptococcus neoformans
Coprinus cinereus
Ustilago maydis
Mortierella verticillata
Phycomyces blakesleeanus
Mucor circinelloides
Rhizopus oryzae
Allomyces macrogynus
Batrachochytrium dendrobatidis
Spizellomyces punctatus
Encephalitozoon cuniculi
Nematocida parisii
Monosiga brevicollis
Salpingoeca rossetta
Capsaspora owczarzaki
Ministeria vibrans
Sphaeroforma artica
Creolimax fragrantissima
Abeoforma whisleri
Pirum gemmata
Amoebidium parasiticum
Corallochytrium limacisporum
Homo sapiens
Ciona intestinalis
Saccoglossus kowalevskii
Drosophila melanogaster
Caenorhabditis elegans
Capitella teleta
Lottia gigantea
Nematostella vectensis
Acropora digitifera
Hydra magnipapillata
Trichoplax adhaerens
Mnemiopsis leidyi
Oscarella carmela
Amphimedon queenslandica
Excavata
Alveolata
Heterokonta
SAR
Archaeplastida
Embryophyta
Amebozoa
Fungi
Metazoa
Holozoa
Opisthokonta
Metazoa cluster
zf C4
Ets
MADF
Unikonta cluster
Fork head
RFX
p53-like
Fungi cluster
Zf Zn2Cys6
Fungal TF
Fugal TF2
GCR1
STE-like TF
Archaeplastida cluster
SAND
AP2
SBP
WRKY
B3
zf Dof
PLATZ
NAM
GRAS
Paneukaryotic cluster
Homeobox
bZIP
HMG box
zf C2H2
HLH
Myb
GATA
HSF
E2F
zf MIZ
CSD
ARID
TBP
zf GRF
Tubby
SRF
zf TAZ
zf BED
HTH_psq
zf NFX1
YL1
CBFB_NFYA

Early Origins of Metazoan Developmental TFs

TFs have very disparate evolutionary origins and patterns. While some TF families seem to have remained quite static throughout evolution, preserving a high degree of similarity between orthologous gene families (Sebé-Pedrós et al. 2011), others such as zf-C2H2 evolved at a faster rate, making it extremely difficult to assign them to specific orthologous gene families over large evolutionary distances (Vaquerizas et al. 2009). Although the field of evolutionary developmental biology has identified many well-known animal developmental TF families in different phyla, these are only a subset of all TFs in metazoan genomes. Even among the members of a particular TF type, usually only some of them are developmental TFs. Therefore the identification of such TFs requires precise class and family phylogeny-based classifications. Here we will review the evolutionary histories of animal multi-gene TF families.

Homeobox is one of the most abundant and diverse developmental TF types in animals, and a milestone in the evolution of animal development (Bürglin 2011). During metazoan evolution, Homeobox genes increased from a rather simple gene complement to a highly diversified toolkit (Larroux et al. 2008). This expansion coincided with an increase in domain combinations (Bürglin 2011), including Pou, Lim, Cut or Paired domains that are found adjacent to the homeodomain. Some of these co-occurring domains are metazoan innovations (Pou, Cut, Paired, Six, Prospero or Iro-box), while others are ancient eukaryotic domains (e.g. LIM or zf-C2H2) fused to Homeobox proteins by domain shuffling. Unicellular relatives of metazoans do not possess a rich Homeobox complement; for example, choanoflagellates have just two Homeobox genes, both belonging to the TALE superclass (King et al. 2008). In fact, TALE Homeoboxes represent one of the two ancient lineages of Homeobox that, together with non-TALE Homeoboxes, have been present since the origin of eukaryotes (Derelle et al. 2007). In animals, the TALE superclass diversified into few families, namely Iroquois, PBX, Meis/PREP and Tgif (Larroux et al. 2008; Bürglin 2011; Fig. 3). In contrast, non-TALE Homeoboxes are extremely diverse. Sponges and ctenophores, which are potentially the two earliest branching metazoans, already possess members of the 5 classes of non-TALE Homeobox (ANTP, PRD-like, POU, LIM and SINE) (Larroux et al. 2008, Ryan et al. 2010). Later in

←

Fig. 2 Phylogenetic patterns of TFome composition across eukaryotes. For clarity, TF types representing $< 2\,\%$ of the corresponding TFome are not considered, and some are summarized in higher-level categories according to structural similarities. This is the case of (i) the Homeobox supergroup, which comprises Homeobox and Homeobox_KN/TALE; (ii) the bZIP supergroup, which comprise bZIP_1, bZIP_2, and bZIP_Maf; and (iii) the p53-like supergroup, which comprise p53, STAT, Runx, NDT80, LAG1, and RHD. To the *Left*, the total number of TF types present in each taxon and the relative abundance of each DNA Binding Domain (DBD) type in the TFome of every species are depicted using the colour code in the legend of DBDs. In the line graph (*Top*), the *solid line* indicates the total number of TFs in each species, and the *dotted line* indicates this number as a percentage of the total number of proteins. The *black asterisks* indicate species with WGDs. The *red asterisks* indicate strict parasites. (Adapted from de Mendoza et al. 2013)

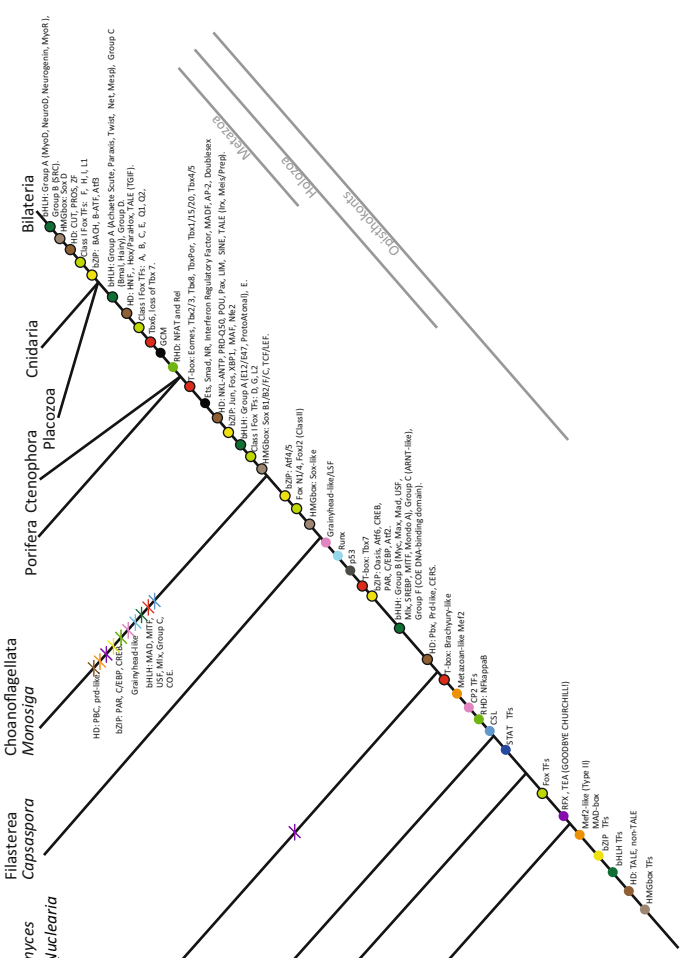


Fig. 3 Cladogram representing metazoan TF evolution. Colours are unique to each domain class. A *coloured dot* indicates the hypothetical origin of the domain. A *black-circled dot* indicates where a specific protein family appears in our taxon sampling. A *cross* means the loss of the domain or specific protein family in a lineage. (Adapted from Sebé-Pedrós et al. 2011)

animal evolution, other classes emerged, including HNF in the last common ancestor of Placozoa, Cnidaria, and Bilateria, and CUT, PROS and ZF in the bilaterian stem (Putnam et al. 2007; Srivastava et al. 2010). This rapid process of duplication and subfunctionalization is characteristic of metazoan Homeobox gene families, which, once they evolve, tend to maintain a conserved domain architecture and clear aminoacidic motifs typical of each gene (Bürglin 2011). In contrast, clear orthologs outside animals are scarce and generally do not bear the key amino acids or concurrent protein domains that characterize animal Homeobox families (Sebé-Pedrós et al. 2011).

T-box genes are also key developmental TFs in animals. They originated in the common ancestor of Opisthokonta and were secondarily lost in Dikarya fungi and in choanoflagellates. Brachyury was the ancestral T-box class; a second class, Tbx7, originated in the common ancestor of Holozoa (Sebé-Pedrós et al. 2013). All other T-box classes (including Tbx1/15/20, Tbx2/3, Tbx4/5, Eomes, TbxPor and Tbx8) originated at the stem of Metazoa, except Tbx6, which appeared in the cnidarian-bilaterian ancestor (Fig. 3; Sebé-Pedrós et al. 2013). Therefore, like Homeobox TFs, T-box TFs underwent rapid radiation and subfunctionalization at the origin of Metazoa, and some classes, like Tbx8 or TbxPor, were later lost in many metazoan lineages. In contrast to Homeoboxes, little innovation occurred in the T-box TF family after the early stages of animal evolution, except for the diversification of Tbx1/15/20 at the stem of the cnidarian-bilaterian ancestor.

Forkhead domain containing genes (Fox) have an ancient eukaryotic origin, and although they are scarce in bikonts, they are quite abundant in opisthokont genomes (Figs. 1, 2). All Fox genes found in non-metazoans are of Class II, including clear homologs of Fox N1/4 and Fox J2 in choanoflagellates (Larroux et al. 2008; Sebé-Pedrós et al. 2011). Class I Fox genes originated at the onset of Metazoa, with Fox D, G and L2 present in non-bilaterian metazoans; Fox A, B, C, E, Q1 and Q2 are found in Placozoa + Cnidaria + Bilateria; and Fox F, H, L1 and I originated in Bilateria (Fig. 3; Larroux et al. 2008; Sebé-Pedrós et al. 2011).

Basic Helix-Loop-Helix (bHLH) transcription factors are one of the main eukaryotic TF types (Figs. 1 and 3) and metazoans have six groups (A, B, C, D, E and F) (Simionato et al. 2007). Most bHLH TFs conserved between metazoans and their unicellular relatives belong to Group B. For example, Myc/Max/Mad, SREBP, USF, Mlx, MondoA, and MITF can be found in both lineages, while Src, AP4, FigAlpha and MNT are metazoan innovations (Fig. 3; Sebé-Pedrós et al. 2011). We find Group C genes in *C. owczarzaki*, which are co-orthologs of metazoan Group C bHLH TFs. Group F, also known as the COE family, includes TFs with a COE DNA-binding domain that also has an HLH domain at the C-terminal part of the protein. This COE domain can be found in the *C. owczarzaki* genome, but it does not have any trace of the HLH domain (also degenerated in some metazoan orthologs) (Suga et al. 2013). Finally, of the six metazoan groups, bHLH A, D and E are unique to metazoans and have diversified into many subfamilies (Group A genes), most of which are heterodimerizing classes (Simionato et al. 2007).

The HMGbox domain encompasses a large family of DNA-binding proteins, some of which have sequence-specific TF activity. Metazoans have a unique set of subtypes of these TFs, with the Sox family being the most diverse. Of the five known

groups of human Sox genes, B1, B2, C, E and F are present in the non-bilaterian metazoans, while Group D appeared at the stem of Bilateria (Fig. 3; Fortunato et al. 2012). Tcf/Lef genes are also animal-specific HMGbox TFs, and act as the regulators of the Wnt pathway (Larroux et al. 2008). No Sox or Tcf/Lef genes are found outside Metazoa.

bZIP is another paneukaryotic TF family (Fig. 1), in this case quite homogenously distributed throughout eukaryotes in terms of relative abundance (Fig. 2). Some metazoan bZIP classes, including Oasis, Atf6, CREB, Atf2, C/EBP, PAR and Atf4/5, originated in unicellular holozoans (Fig. 3), constituting homodimerizing classes in most cases. In contrast, the heterodimerizing bZIP classes Jun, Fos, XBP1, MAF and Nfe2 are metazoan innovations, with bZIP BACH, B-ATF and Atf3 evolving later in the metazoan lineage (Sebé-Pedrós et al. 2011).

NFkappaB genes have also been found to have a more ancient origin than previously thought (Sebé-Pedrós et al. 2011; de Mendoza et al. 2013). As for T-box genes, they originated in the last common ancestor of opisthokonts and were lost in both choanoflagellates and fungi (Fig. 3). This ancestral gene had the typical NFkappaB domain structure, with a RHD domain followed by Ankyrin repeats. Later on in metazoans, two other RHD domain-containing TFs without Ankyrin repeats evolved: NFAT in Placozoa + Cnidaria + Bilateria and Rel in Bilateria (Gauthier and Degnan 2008; Sebé-Pedrós et al. 2011).

STAT TFs were present in the last common ancestor of opisthokonts and apusozoans and were lost in fungi. STAT TFs are found in organisms with tyrosine kinase genes, such as choanoflagellates and filastereans. In this context, STAT TFs interact with Y-phosphorylated proteins through their C-terminal SH2 domain, constituting the transcriptional outputs of tyrosine kinase signaling.

Although MADS-box (SRF domain) TFs are paneukaryotic, TFs with specific Mef2 signatures appeared at the stem of the Opisthokonta, are were secondarily lost in Dikarya fungi and choanoflagellates. Finally, p53 and Runx TFs are holozoan innovations.

In summary, metazoan developmental TFs have three main sources: first, *de novo* types that emerged at the onset of Metazoa; second, co-opted genes that originated in their unicellular relatives; and third, duplication of pre-existing TFs, sometimes accompanied by diversification of co-occurring domains. Among the truly metazoan innovations we find Smad (MH1 domain), Ets, DoubleSex, AP-2 and the Nuclear Receptors (zf-C4 domain) (Fig. 3). Key TF types that originated in unicellular Holozoa include NFkappaB, T-box, p53, the Myc/Max network, Grainyhead and LSF, and Runx. Finally, significantly enriched developmental TF types in metazoans include Homeobox, T-box, Fox, Sox and NFkappaB (Fig. 1). All of these TF types were present before the emergence of animal multicellularity, but new specific families involved in development emerged during metazoan evolution.

Establishment of the Metazoan TF Regulatory System

Transcriptional regulation by sequence-specific TFs goes beyond the TFome content itself. TFs act in a cellular context and their activity requires interaction with other proteins and, of course, with specific DNA sites. TFs bind to enhancers and promoters of genes and also interact with cofactors that modulate their activity in various ways (for example, controlling translocation to the nucleus or restricting DNA binding specificity). Moreover, TF binding is influenced by the epigenetic context, such as nucleosome positioning, histone modifications and DNA methylation patterns (Spitz and Furlong 2012). The cross-talk between TFs and all of these elements is what ultimately defines the TF regulatory system.

Currently, little is known about the functions of TFs in the unicellular relatives of animals, so it is not possible at this time to have a clear picture on how TF activities changed during the transition to animal multicellularity. Nevertheless, experimental approaches have provided some insights into this issue. There are two main evolutionary changes that may have accompanied an increase in regulatory interactions in animals: *trans* and *cis* regulatory changes. *Trans* regulatory changes involve the appearance of new physical interactions between TFs and cofactors (which may be either new or ancient) and/or new DNA-binding motif specificity. *Cis* regulatory changes involve the evolution of new DNA binding sites for a particular TF, which results in new downstream targets controlled by a TF and, therefore, the re-wiring of the TF network. The following cases provide examples of such processes.

The evolution of Brachyury, a well-known T-box TF with essential roles in gastrulation and mesoderm specification in animals (Technau 2001), is an example of a *trans* regulatory change. The presence of a clear Brachyury ortholog in *C. owczarzaki* was one of the most striking findings in the first unicellular holozoan TF survey (Sebé-Pedrós et al. 2011), and the exploration of its functional conservation using heterologous expression in *Xenopus laevis* and protein binding microarray (PBM) experiments revealed a suggestive pattern. *C. owczarzaki* Brachyury can rescue artificially inhibited embryo gastrulation in *Xenopus laevis*, although it does so in a rather unspecific way, activating diverse downstream genes that are commonly activated by different T-box gene families in metazoans. In contrast, sponge and ctenophore Brachyury homologs behave very much like endogenous *Xenopus laevis* Brachyury (Yamada et al. 2010; Sebé-Pedrós et al. 2013). Moreover, the results of the PBM experiments showed that the differences between *C. owczarzaki* Brachyury and metazoan Brachyury were not due to changes in the DNA-binding motif specificity of the different Brachyury genes studied (Sebé-Pedrós et al. 2013). Taken together, these two results strongly suggest that Brachyruy target specificity in metazoans arose through interaction with cofactors (Smad and probably others) that were probably established at the onset of Metazoa (Fig. 4a).

Generally, transcription binding site recognition motifs seem to be highly conserved in some TF families, such as bHLH or T-box (Jolma et al. 2013; Sebé-Pedrós et al. 2013). In contrast, other families are more labile in their sequence binding preferences (Nakagawa et al. 2013), such as in Homeoboxes, zinc fingers (C2H2)

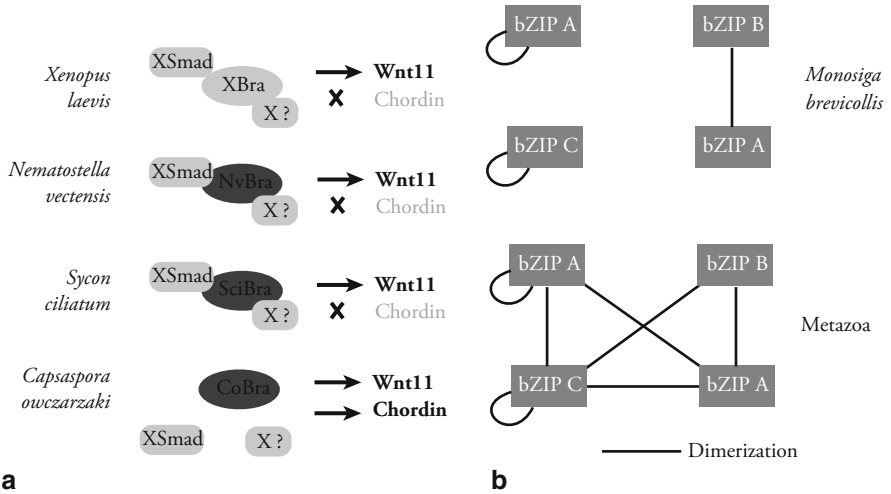


Fig. 4 *Trans* changes during the evolution of metazoan TF networks. **a** Schematic representation of the results from Sebé-Pedrós et al. 2013. Brachyury, an essential TF for animal development, is found in the unicellular *Capsaspora owczarzaki*. In a heterologous expression assay, only the early-metazoan Brachyury orthologs (of the sponge *Sycon ciliatum* and the cnidarian *Nematostella vectensis*) show the same molecular phenotype as the endogenous *Xenopus laevis* Brachyury ortholog (Activation of Wnt11 and non-activation of Chordin). This suggests that the *trans* regulatory interactions between Brachyury and cofactors like Smad (and probably other unknown cofactors, shown as "X?") were established at the onset of Metazoa. **b** Schematic representation of the results of Reinke et al. 2013. Analysis of the *in vitro* interactions of all bZIPs of several metazoans and of the choanoflagellate *Monosiga brevicollis* showed that the proportion of heterodimeric interactions (versus homodimeric interactions) was much lower in a unicellular context. This result suggests an increase in the complexity of the bZIP heterodimerizing network during the transition to animal multicellularity, which probably allowed new regulatory outputs by combining old (and also new) bZIP TFs

and Nuclear receptors (Jolma et al. 2013), all of which are major components of animal TF toolkits (de Mendoza et al. 2013). The plasticity of DNA-binding motif recognition is crucial to avoid cross-activation between TF paralogs. Thus, the transition to animal multicellularity involved the expansion of some of these labile families, which may have facilitated the acquisition of developmental complexity by allowing denser, non-overlapping readers of *cis*-regulatory information.

An open question is how these gene regulatory networks changed, not only quantitatively (more interactions), but also qualitatively. For example, it is conceivable to find that *C. owczarzaki* Brachyury regulates fewer downstream genes, and is more directly connected to batteries of effector genes than to other regulatory TFs. The degree of hierarchy of these interactions (where each level is defined as a master regulatory gene regulating another regulatory gene) would probably be lower in the case of unicellular gene regulatory networks. As a working hypothesis, shallower regulatory networks could have been intercalated into more complex wiring when developmental multicellularity evolved (Davidson and Erwin 2006).

Another example of *trans* evolution is the increase in interactivity found in bZIP TFs when comparing metazoans with unicellular eukaryotes (Reinke et al. 2013). bZIP TFs bind to DNA as dimers, usually homodimers (two bZIPs of the same class) but also as heterodimers, although not all possible heterodimeric interactions occur. Phylogenetically, some heterodimeric interactions can be predicted, but most remain elusive in non-model organisms. Reinke and co-workers examined the *in vitro* dimerization affinities of all bZIPs in several species, including a unicellular relative of metazoans, the choanoflagellate *Monosiga brevicollis*. They found that metazoan bZIPs have a higher proportion of heterodimeric interactions than *M. brevicollis* Therefore, there was an increase in complexity of the bZIP interaction network, which generated new combinatorial binding specificities (Fig. 4b).

bHLHs are another type of heterodimerizing TF. Myc is an animal bHLH TF responsible for cell cycle control, growth and apoptosis, and is a crucial oncogene in many types of cancer. It binds to DNA in association with Max, another bHLH TF. Max has the ability to interact with other bHLH TFs, mainly Mxd/Mad and MNT. Mxd/Mad and MNT proteins antagonize Myc, controlling cell cycle arrest and gene repression. This network was already established in a pre-metazoan context, where we find members of the Myc, Max and Mlx/Mad families, while MNT appeared later on, in the Eumetazoan split (Sebé-Pedrós et al. 2011; Young et al. 2011). It has been shown that *M. brevicollis* Myc and Max orthologs heterodimerize, revealing experimentally that the heterodimerization network was present and functionally conserved (Young et al. 2011). Moreover, a core set of ancient eukaryotic genes involved in ribosome biogenesis seems to be regulated by Myc in Holozoan genomes. The E-box motif (the typical DNA-binding motif of Myc/Max dimers) is found to be enriched in the promoters of *M. brevicollis* and animals that have Myc and Max orthologs, whereas fungi and *C. elegans*, which both lack Myc genes, are depleted of E-boxes in the promoters of those ribosomal genes (Brown et al. 2008). Myc network data from unicellular holozoans tell us that not only were the physical interactions between different TFs already in place, but also that there is some cis-regulatory conservation in the downstream genes involved in basic cell processes.

Despite this rather simple cis-regulatory conservation between unicellular taxa and metazoans, there is evidence of an increase in cis-regulatory interactions in metazoans. The 5′ intergenic regions of TFs and other regulatory genes (where many regulatory proteins bind) are expanded in metazoan genomes (compared to the mean intergenic distance of the genome), which allows for a more complex regulatory landscape (both in terms of the number and combination of regulatory modules, such as enhancers or repressors) and, ultimately, for complex spatiotemporal regulatory states (Nelson et al. 2004; Suga et al. 2013). This complex regulatory landscape explains why some TFs, such as NK or Hox, have evolved in syntenic blocks. Other TFs have retained by-stander genes, forming micro-syntenic blocks. The by-stander gene contains cis-regulatory elements embedded in its gene body that regulate the expression of the TF, preventing genomic recombination that would separate the two neighbours (Irimia et al. 2012). Overall, we can observe how the unique metazoan

genomic architecture is greatly influenced by regulatory interactions between enhancers and downstream genes. These interactions allow complex developmental gene regulation by TFs.

In summary, current evidence suggests that not only did the TF repertoire itself changed during the origin of animal multicellularity, but a fundamental change also occurred in the gene regulatory networks in which these genes were embedded. These created more complex patterns of spatiotemporal regulation of gene expression, an essential feature of a complex multicellular entity. Some of these major regulatory changes occurred early in metazoan history, and became virtually frozen, producing what is known as kernel gene regulatory networks; while others were incorporated later in a phylum-, class- or species-specific manner, forming the basis of the morphological and functional diversity of extant metazoans (Davidson and Erwin 2006).

Summary

1. From a broad eukaryotic perspective, it is clear that phylogenetic inertia is an important factor that conditions the TF toolkit of different origins of multicellularity. Therefore, studying the TF repertoire of metazoans' unicellular relatives is essential for understanding the foundation of the metazoan TFome.
2. Various evolutionary forces have shaped the metazoan TFome, including *de novo* gene origin, gene family expansion and gene co-option.
3. The establishment of complex gene regulatory networks accompanied the origin of Metazoa. In the context of these networks many TFs were locked into specific developmental processes.
4. A global rearrangement of both TFome content and *cis* and *trans* interactions facilitated an explosion in the regulatory capabilities of TFs in Metazoa.

References

Adamska M, Degnan BM, Green K, Zwafink C (2011) What sponges can tell us about the evolution of developmental processes. Zoology (Jena) 114:1–10. doi:10.1016/j.zool.2010.10.003

Bouget FY, Berger F, Brownlee C (1998) Position dependent control of cell fate in the Fucus embryo: role of intercellular communication. Development 125:1999–2008

Brown SJ, Cole MD, Erives AJ (2008) Evolution of the holozoan ribosome biogenesis regulon. BMC Genomics 9:442. doi:10.1186/1471-2164-9-442

Bürglin T (2011) Homeodomain subtypes and functional diversity. In: Hughes TR (ed) A handbook of transcription factors, SE–5. Springer, The Netherlands, pp 95–122

Cock JM, Sterck L, Rouzé P et al (2010) The Ectocarpus genome and the independent evolution of multicellularity in brown algae. Nature 465:617–621. doi:10.1038/nature09016

Coelho SM, Godfroy O, Arun A et al (2011) OUROBOROS is a master regulator of the gametophyte to sporophyte life cycle transition in the brown alga Ectocarpus. Proc Natl Acad Sci U S A 108:11518–11523. doi:10.1073/pnas.1102274108/

Collen J, Porcel B, Carre W et al (2013) Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. Proc Natl Acad Sci U S A 110:5247–5252. doi:10.1073/pnas.1221259110

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

De Mendoza A, Sebé-Pedrós A, Sestak MS et al (2013) Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proc Natl Acad Sci U S A 110:E4858–4866. doi:10.1073/pnas.1311818110

De Smet R, Adams KL, Vandepoele K et al (2013) Convergent gene loss following gene and genome duplications creates single-copy families in fl owering plants. Proc Natl Acad Sci U S A 110:2898–2903. doi:10.1073/pnas.1300127110/

Degnan BM, Vervoort M, Larroux C, Richards GS (2009) Early evolution of metazoan transcription factors. Curr Opin Genet Dev 19:591–599. doi:10.1016/j.gde.2009.09.008

Derelle R, Lopez P, Le Guyader H et al (2007) Homeodomain proteins belong to the ancestral molecular toolkit of eukaryotes. Evol Dev 9:212–219. doi:10.1111/j.1525-142X.2007.00153.x

Fortunato S, Adamski M, Bergum B et al (2012) Genome-wide analysis of the sox family in the calcareous sponge Sycon ciliatum: multiple genes with unique expression patterns. Evodevo 3:14. doi: 10.1186/2041-9139-3-14

Gauthier M, Degnan BM (2008) The transcription factor NF-kappaB in the demosponge Amphimedon queenslandica: insights on the evolutionary origin of the Rel homology domain. Dev Genes Evol 218:23–32. doi:10.1007/s00427-007-0197-5

Grosberg RK, Strathmann RR (2007) The evolution of multicellularity: a minor major transition? Annu Rev Ecol Evol Syst 38:621–654. doi:10.1146/annurev.ecolsys.36.102403.114735

Hammonds AS, Bristow CA, Fisher WW et al (2013) Spatial expression of transcription factors in Drosophila embryonic organ development. Genome Biol 342:R140. doi:10.1186/gb-2013-14-12-r140

Irimia M, Tena JJ, Alexis MS et al (2012) Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. Genome Res 22:2356–2367. doi:10.1101/gr.139725.112

Iyer LM, Anantharaman V, Wolf MY, Aravind L (2008) Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. Int J Parasitol 38:1–31. doi:http://dx.doi.org/10.1016/j.ijpara.2007.07.018

Jolma A, Yan J, Whitington T et al (2013) DNA-binding specificities of human transcription factors. Cell 152:327–339. doi:10.1016/j.cell.2012.12.009

King N, Westbrook M, Young S, Kuo A et al (2008) The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. Nature 21:4300–4305. doi:10.1091/mbc.E10-03-0251

Knoll AH (2011) The multiple origins of complex multicellularity. Annu Rev Earth Planet Sci 39:217–239. doi:10.1146/annurev.earth.031208.100209

Larroux C, Luke GN, Koopman P et al (2008) Genesis and expansion of metazoan transcription factor gene classes. Mol Biol Evol 25:980–996. doi:10.1093/molbev/msn047

Levine M, Tjian R (2003) Transcription regulation and animal diversity. Nature 424:147–151. doi:10.1038/nature01763

Maere S, De Bodt S, Raes J et al (2005) Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci U S A 102:5454–5459. doi:10.1073/pnas.0501102102

Nakagawa S, Gisselbrecht SS, Rogers JM et al (2013) DNA-binding specificity changes in the evolution of forkhead transcription factors. Proc Natl Acad Sci U S A 110:12349–12354. doi:10.1073/pnas.1310430110

Nakamura Y, Sasaki N, Kobayashi M et al (2013) The first symbiont-free genome sequence of marine red alga, Susabi-nori (Pyropia yezoensis). PLoS One 8:e57122. doi:10.1371/journal.pone.0057122

Nelson CE, Hersh BM, Carroll SB (2004) The regulatory content of intergenic DNA shapes genome architecture. Genome Biol 5:R25.1–R25.15

Niklas KJ, Newman SA (2013) The origins of multicellular organisms. Evol Dev 15:41–52. doi:10.1111/ede.12013

Parfrey LW, Lahr DJG (2013) Multicellularity arose several times in the evolution of eukaryotes. Bioessays 35:339–347. doi:10.1002/bies.201200143

Peters AF, Scornet D, Ratin M et al (2008) Life-cycle-generation-specific developmental processes are modified in the immediate upright mutant of the brown alga Ectocarpus siliculosus. Development 135:1503–1512. doi:10.1242/dev.016303

Pires-daSilva A, Sommer RJ (2003) The evolution of signalling pathways in animal development. Nat Rev Genet 4:39–49. doi:10.1038/nrg977

Putnam NH, Srivastava M, Hellsten U et al (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317:86–94. doi:10.1126/science.1139158

Radakovits R, Jinkerson RE, Fuerstenberg SI et al (2012) Draft genome sequence and genetic transformation of the oleaginous alga Nannochloropis gaditana. Nat Commun 3:686. doi:10.1038/ncomms1688

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

Rokas A (2008) The origins of multicellularity and the early history of the genetic toolkit for animal development. Annu Rev Genet 42:235–251. doi:10.1146/annurev.genet.42.110807.091513

Ryan JF, Pang K, Mullikin JC et al (2010) The homeodomain complement of the ctenophore Mnemiopsis leidyi suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa. Evodevo 1:9. doi:10.1186/2041-9139-1-9

Sebé-Pedrós A, de Mendoza A, Lang BF et al (2011) Unexpected repertoire of metazoan transcription factors in the unicellular holozoan Capsaspora owczarzaki. Mol Biol Evol 28:1241–1254. doi:10.1093/molbev/msq309

Sebé-Pedrós A, Ariza-Cosano A, Weirauch MT et al (2013) Early evolution of the T-box transcription factor family. Proc Natl Acad Sci U S A 110:16050–16055. doi:10.1073/pnas.1309748110

Simionato E, Ledent V, Richards G et al (2007) Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. BMC Evol Biol 7:1–18. doi:10.1186/1471-2148-7-33

Spitz F, Furlong EEM (2012) Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 13:613–626. doi:10.1038/nrg3207

Srivastava M, Simakov O, Chapman J et al (2010) The Amphimedon queenslandica genome and the evolution of animal complexity. Nature 466:720–726. doi:10.1038/nature09201

Suga H, Chen Z, de Mendoza A et al (2013) The Capsaspora genome reveals a complex unicellular prehistory of animals. Nat Commun 4:1–9. doi:10.1038/ncomms3325

Technau U (2001) Brachyury, the blastopore and the evolution of the mesoderm. Bioessays 23:788–794. doi:10.1002/bies.1114

Technau U, Steele RE (2011) Evolutionary crossroads in developmental biology: Cnidaria. Development 138:1447–1458. doi:10.1242/dev.090472

Van De Peer Y Maere S Meyer A (2009) The evolutionary significance of ancient genome duplications. Nat Rev Genet 10:725–732. doi:10.1038/nrg2600

Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM (2009) A census of human transcription factors: function, expression and evolution. Nat Rev Genet 10:252–263. doi:10.1038/nrg2538

Weirauch MT, Hughes TR (2011) A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. In: A handbook of transcription factors. Springer, The Netherlands. Subcell Biochem 52:25–73. doi:10.1007/978-90-481-9069-0

Yamada A, Martindale MQ, Fukui A, Tochinai S (2010) Highly conserved functions of the Brachyury gene on morphogenetic movements: Insight from the early-diverging phylum Ctenophora. Dev Biol 339:212–222. doi:10.1016/j.ydbio.2009.12.019

Young SL, Diolaiti D, Conacci-Sorrell M et al (2011) Premetazoan ancestry of the Myc–Max network. Mol Biol Evol 28:2961–2971. doi:10.1093/molbev/msr132