Basic Helix-Loop-Helix Transcription Factors in Evolution: Roles in Development of Mesoderm and Neural Tissues

Fuki Gyoja

Addresses: 1, Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan

2, Department of Biology, Faculty of Science and Engineering, Konan University, 8-9-1 Okamoto, Higashinada, Kobe 658-8501, Japan

E-mail: 1, fgyoja@gmail.com

2, fgyoja@center.konan-u.ac.jp

Telephone: +81-78-435-2511

Fax: +81-78-435-2539

Short Title

bHLH Evo-Devo: Mesodermal/Neural Tissues

Keywords

Transcription Factors, Metaoza, Genome, Evo-Devo

Abstract

Basic helix-loop-helix (bHLH) transcription factors have attracted the attention of developmental and evolutionary biologists for decades because of their conserved functions in mesodermal and neural tissue formation in both vertebrates and fruit flies. Their evolutionary history is of special interest because it will likely provide insights into developmental processes and refinement of metazoan-specific traits. This review briefly considers advances in developmental biological studies on bHLHs/HLHs. I also discuss recent genome-wide surveys and molecular phylogenetic analyses of these factors in a wide range of metazoans. I hypothesize that interactions between metazoan-specific Group A, D, and E bHLH/HLH factors enabled a sophisticated transition system from cell proliferation to differentiation in multicellular development. This control mechanism probably emerged initially to organize a multicellular animal body and was subsequently recruited to form evolutionarily novel tissues, which differentiated during a later ontogenetic phase.

I. Molecular structure and classification of bHLHs/HLHs

Basic helix-loop-helix (bHLH) genes are evolutionarily conserved transcription factors that have been reported in wide range of eukaryotes, including metazoans, plants, fungi, and unicellular organisms, including the Choanoflagellida and Amoebozoa (Sebé-Pedrós et al., 2011; for reviews, see Massari and Murre, 2000; Jones, 2004; Fritzsch et al., 2015). A bHLH domain is composed of two regions, a basic region for recognition and binding to a target DNA sequence, and a helix-loop-helix (HLH) motif, a structure comprising two α-helices separated by a loop (Davis et al., 1990; Massari and Murre, 2000; Jones, 2004). HLH motifs mainly mediate dimerization with the same or another type of bHLH/HLH protein (Davis et al., 1990; Massari and Murre, 2000; Jones, 2004). Metazoan bHLHs/HLHs have been categorized into high-order Groups A-F, based on features such as molecular phylogenetic relationships, recognition of DNA sequences, and protein structures (Atchley and Fitch, 1997; Ledent and Vervoort, 2001). Group B includes genes such as bilaterian *myc*, *max*, and *SREBP*. All other groups seem to have originated from this most ancient group (a simplified diagram is shown in Fig. 1. For representatives of each group, see Table1). Group B members have been reported in metazoans and fungi as well as several unicellular organisms, such as those belonging to the Choanoflagellida and Holozoa (Sebé-Pedrós et al., 2011). They have biologically fundamental functions in control of metabolism and cell proliferation in both yeasts and metazoans (Massari and Murre, 2000; Jones, 2004). Groups A (*MyoD*, *Achaete-Scute* etc.), D (*ID* and *pearl*), and E (*HES*, *HEY* etc.) seem to be metazoan innovations (Sebé-Pedrós et al., 2011). This review will discuss mainly these three Groups, since many studies have revealed their roles in mesodermal and neural tissue differentiation, and these proteins can interact directly with one another (Massari and Murre, 2000; Jones, 2004). Since there have been many studies of these factors in vertebrates and fruit-flies, I will introduce examples from both, where possible. Group C members, such as *Clock* and *Bmal*, are involved in biological processes such as circadian clock and detoxification. Group C members have been reported in metazoans as well as a holozoan, *Capsaspora owczarzaki* (Sebé-Pedrós et al., 2011). Group F contains only one factor, *collier/olfactory/early B-cell factor*

(*COE*). This gene has been reported in metazoans, as well as in *C. owczarzaki* (Suga et al., 2013). It also has a conserved function in neural differentiation in metazoans (Dubois and Vincent, 2001; Demilly et al., 2011), which will be addressed later.

In molecular phylogenetic analyses, when two or more bHLH genes from phylogenetically distant organisms form a single clade with high confidence, those genes are regarded as members of an orthologous family (Ledent and Vervoort, 2001; also see Table1). This definition has been adopted in many studies (Ledent et al., 2002; Simionato et al., 2007; Gyoja et al., 2012; Gyoja, 2014; Bao et al., 2017). When I describe bHLH/HLH genes below, I will indicate to which orthologous family they belong.

II. Group A, D, and E bHLH/HLH interactions in development

II-1 Group A. Many Group A factors function as transcriptional activators that promote tissue differentiation. Although there are several exceptions, most Group A bHLHs function as heterodimers (For a review, see Massari and Murre, 2000). One Group A orthologous family, E12/E47, which is also called E-protein in mammals, is expressed almost ubiquitously, at least in rats and *Drosophila* (Cronmiller and Cummings, 1993; Roberts et al., 1993). On the other hand, most other Group A bHLH factors are expressed in a tissue-specific manner, and their transcription factor activities usually require heterodimerization with ubiquitous E-proteins (Murre et al., 1989). These Group A heterodimers specifically recognize DNA sequences called E-boxes (Fig. 2a).

Many Group A factors are involved in differentiation of mesodermal and neural tissues. I will begin with a well-known orthologous family involved in mesodermal tissue development. *MyoD* and its paralogs in vertebrates, which belong to orthologous family "MyoD," are potent myogenic factors (for review, see Olson, 1990). When introduced into cultured cells from rats, humans, or chickens, *MyoD* induces expression of skeletal muscle differentiation-markers (Weintraub et al., 1989). Mice have four *MyoD* paralogs, called *MyoD*, *Myf-5*, *myogenin*, and *MRF-4* (Olson, 1990;

Weintraub et al., 1991). Mice lacking functions of both *MyoD* and *Myf-5* are born without any detectable skeletal muscle and die soon after birth, probably because they cannot breathe (Rudnicki et al., 1993). This result shows that these genes are indispensable for skeletal muscle differentiation. MyoD-induced myogenesis is related to cell proliferation. Oncogenes *ras* and *fos* can prevent expression of *MyoD* and *myogenin*, as well as muscle differentiation markers in cultured mouse cells (Lassar et al., 1989). On the other hand, myogenesis is induced by inhibiting *ras* function (Lassar et al., 1989).

The *Drosophila melanogaster* genome has a single member of MyoD, called *nautilus*. This gene is involved in formation of subsets of muscle cells, although loss-of-function phenotypes are less severe than in mice (Balagopalan et al., 2001). Therefore, the function of this gene seems to be at least partially conserved in both deuterostomes and protostomes.

Group A bHLH factors also contribute greatly to neural fate specification and differentiation. Here I will describe an orthologous family, ACSa. Guillemot et al. (1993) reported that *Mash-1/Ascl1*, one of two paralogs in mice, is required for olfactory and autonomic neuron formation. This gene, together with another Group A factor, *Math3*/*NeuroD4*, promotes neuronal fate and represses glial fate in several cell populations during central nervous system (CNS) development in mice (Tomita et al., 2000). Genome-wide analysis for *Mash-1/Ascl1* recognition sites by ChIP-chip showed that this Group A bHLH promotes cell proliferation in early neurogenesis, and inhibits it in the late differentiation phase of neural differentiation (Castro et al., 2011).

The *D. melanogaster* genome has four tandemly aligned paralogs, *Achaete*, *Scute*, *Asense*, and *Lethal of Scute*. This cluster is called the Achaete-Scute complex. Its members have pivotal, partially redundant roles in neural differentiation, as in microchaete and macrochaete formation in *Drosophila* (For reviews, see Ghysen and Dambly-Chaudière, 1988; Campuzano and Modolell, 1992; Jan and Jan, 1993). They are called “proneural genes” because their expression endows ectodermal cells with neural cell-forming potential (Ghysen and Dambly-Chaudière, 1989; Jan and Jan, 1993; Bertrand et al., 2002). *Achaete* or *Scute* single-mutant flies lose subsets of sensory bristles in

wing discs, while double-mutant flies lack all sensory bristles (Ghysen and Dambly-Chaudière, 1988). Like *Mash-1*/*Ascl1*, *Asense* first promotes neuroblast division, and later suppresses its proliferation (Southall and Brand, 2009).

Some Group A factors such as murine *Twist* (Spicer et al., 1996) and murine *MyoR*/*Musculin* (Lu et al., 1999) act as transcriptional repressors that inhibit other Group A factors (Fig. 2b). Such inhibitory Group A factors form dimers with E-proteins, thereby antagonizing activator-type Group A factors. At least in some cases, these dimers bind to E-boxes, thereby inhibiting the binding of activator-type Group A factors/E-protein heterodimers (Lu et al., 1999).

II-2 Group D. Group D proteins have helix-loop-helix domains, but no basic region; therefore, they are HLH factors, but not bHLH factors. Here I describe one of two orthologous families of Group D, called EMC. *M. musculus* has four paralogs, *Inhibitor of DNA binding 1* (*Id1*), *Id2*, *Id3* and *Id4*, while *D. melanogaster* has one, which is called *extramacrochaete* (*EMC*). Mouse IDs dimerize with Group A factors such as MyoD, E12 and E47 and inhibit them from binding E-boxes, thereby repressing their transcriptional activity (Benezra et al., 1990; Fig. 2c). There are many reports on ID functions, such as promotion of cell growth, cell differentiation arrest, cell cycle progression, and oncogenesis (For a review, see Norton, 2000). For example, double knockout mice for *Id1* and *Id3* show aberrant neurogenesis and angiogenesis in brain, and cannot survive beyond E13.5 (Lyden et al., 1999). In those knockout mice, neuroblasts prematurely exit the cell cycle and start to express differentiation markers (Lyden et al., 1999). Therefore, those paralogs likely have an essential function in maintaining neuroblasts in an undifferentiated, proliferative state.

The gene name "*extramacrochaete*" of *D. melanogaster* came from its loss of function phenotype. Loss-of-function mutants of this gene have numerous extra-sensory organs, especially macrochaete (Botas et al., 1982). As described above, formation of these organs is largely dependent on *Achaete* and *Scute*. EMC protein inhibits Achaete and Scute proteins from binding to E-boxes *in vitro* (Van Doren et al., 1991). EMC has also been shown to inhibit those Group A factors

during the course of *Drosophila* development *in vivo* (Cubas and Modolell, 1992; For a review, see Jan and Jan, 1993).

II-3 Group E. An orthologous family, "Hairy/Enhancer of Split", sometimes abbreviated "HES," is the most vigorously studied orthologous family within Group E (for gene names, see Table1). Many studies on HES have revealed their functions as transcriptional repressors of Group A activators. Mice have six paralogs, *Hes1-3* and *5-7* (Kageyama et al., 2007). They are important in maintaining neural progenitors in undifferentiated, proliferating states (for review, see Kageyama et al., 2007; Imayoshi and Kageyama, 2014). *Hes1; Hes5* double knockout mice cannot maintain undifferentiated neural progenitor cells properly, formed neurons prematurely at the expense of other types of neural cells, such as glia. They form severely disorganized CNS structures (Hatakeyama et al., 2004). These phenotypes become even more severe in *Hes1; Hes3; Hes5* triple knockout mice (Hatakeyama et al., 2004). Repressional targets of *HES* genes include Group A neural differentiation factors such as *Mash-1/Ascl1*.

HES protein shows oscillatory expression, and this feature is proposed to be important for retaining cells in an undifferentiated, proliferative state (For reviews, see Kageyama et al., 2007; Imayoshi and Kageyama, 2014). A C-terminal WRPW motif, which will be discussed below, functions as a polyubiquitylation signal and induces rapid degradation of HES proteins (Hirata et al., 2002; Kang et al., 2005). In neural progenitor cells of mouse telencephalon, HES promotes cell-autonomous oscillatory transcription by the following mechanisms (Hirata et al., 2002). Once HES represses its own transcription, the concentration of HES protein within the cell decreases rapidly, because it is unstable. This weakens HES transcriptional repression, which restarts transcription and translation of HES, increasing the total amount of HES protein. For this reason, Ascl1, a Group A protein belonging to ASCa, also oscillates, because it is also unstable and its expression is suppressed by HES proteins. Oscillation of these Group A and E bHLH factors is thought to retain neural progenitor cells in a multi-potent, proliferating state (Imayoshi and Kageyama, 2014). On the

other hand, persistent, non-oscillatory expression of *Hes-1* keeps cells in an unproliferative, undifferentiated state, rather than in a proliferative state, in two boundary regions of the CNS, the zona limitans intrathalamica and the isthmus (Baek et al., 2006; Kageyama et al., 2007). When *Hes-1* is persistently activated in neural progenitor cells, cell proliferation rate is reduced (Baek et al., 2006).

HES proteins can repress Group A activators in several different ways (Fig. 2c and 2d; for review, see Kageyama et al., 2007; Imayoshi and Kageyama, 2014). First, homodimers of HES, or another Group E factor HEY, or heterodimers of HES and HEY, bind directly to DNA sequences, such as N-box or C-site, to repress transcription of Group A members in mammalian cultured cells (Fig. 2d; Sasai et al., 1992; Iso et al., 2001). An evolutionarily conserved tetra-peptide motif "WRPW" at HES C-termini then recruits a strong transcriptional repressor, Groucho/Transducin Like Enhancer of Split (TLE), suppressing transcriptional activity nearby (Fisher et al., 1996). A HEY protein molecule also has a similar, conserved YRPW motif at its C-terminus. At present, the role of the YRPW motif is not as clear as that of HES (Iso et al., 2001; Fischer and Gessler, 2007). Secondly, HES protein seems to form a dimer with Group A protein, preventing it from forming transcriptionally active heterodimers (Fig. 2e; Sasai et al., 1992; Fritzsch et al., 2015).

D. melanogaster has 11 paralogs of HES in the genome (Ledent and Vervoort, 2001; See Table 1). Among them, *Hairy* suppresses *Achaete* function to prevent ectopic sensory organ formation (Botas et al., 1982; Skeath and Carroll, 1991). *E(spl)* genes, other members of the *Drosophila* HES family, comprise a cluster in the genome, and also have a suppressive function in neural differentiation (for a review, see Campos-Ortega and Knust, 1990). Heterodimers of E(spl) m7/mγ and Group A bind to an E-box, representing the third mechanism for suppression of Group A by Group E (Fig. 2f; Giagtzoglou et al., 2003). By this interaction, heterodimerization of E-proteins with other Group A factors is antagonized, and transcriptional activation by Group A is repressed (Giagtzoglou et al., 2003). Combinations of these repression mechanisms apparently enable sophisticated and strictly controlled repression of Group A bHLHs by Group E factors (Giagtzoglou et al., 2003).

The sole Group F factor, *COE*, also has conserved functions in differentiation of mesodermal and neural tissues (Dubois and Vincent, 2001; Demilly et al., 2011). I will not discuss this group hereafter because direct molecular interactions between COE and bHLHs/HLHs of other groups have not been reported, so far as I am aware. Moreover, this factor seems to have evolved before metazoan emergence, because recently a candidate was reported in the genome of the non-metazoan holozoan, *C. owczarzaki* (Suga et al., 2013). But orthology between this gene and metazoan COE genes should be analyzed carefully in the future. There is a possibility that Group F HLH also has some indispensable roles in developmental processes, which are achieved by Group A, D, and E bHLHs/HLHs, and therefore we should pay attention to such studies of COE in future.

III Recent advances in evolutionary studies of Group A, D, and E bHLHs/HLHs

As described above, many developmental and cell biological studies of Group A, D, and E bHLHs/HLHs have been performed in mice and fruit-flies. Both of them belong to the Bilateria. In this section, I describe recent genome-wide bHLH/HLH surveys in basal metazoans, namely Porifera, Ctenophora, Placozoa, and Cnidaria, as well as Bilateria. Phylogeny of metazoans and a summary of this section are shown in Fig. 3.

(1) The origin of Group A and E bHLHs, inferred from the Phyla Porifera and Ctenophora

I begin with two metazoan phyla, the Porifera and Ctenophora, the most basal metazoans. There have been two contradictory theories regarding their phylogenetic positions. Some claim that the Porifera is the most basal, while others argue that the Ctenophora occupies that position (Ryan et al., 2013; Pisani et al., 2015). A recently published study using a large, high-quality dataset strongly supports the traditional view, which places the Porifera in the most basal position (Simion et al., 2017). Poriferans, also known as sponges, are sessile. They lack muscle, nerve, and gut tissue (Srivastava et al., 2010). Ctenophores, also known as comb jellies, are planktonic, and swim using eight rows of cilia, which resemble “combs.” Most species have smooth muscle, while one species has sarcomeric muscle (Ryan et al., 2013). They also have a nervous system (Ryan et al., 2013).

To date, no Group A or E bHLHs have been reported in non-metazoan genomes, including the Choanoflagellate, *Monosiga brevicollis* (Sebé-Pedrós et al., 2011) and the filasterean, *C. owczarzaki* (Suga et al., 2013). Simionato et al. (2007) comprehensively surveyed the genome of the sponge, *Amphimedon queenslandica* for bHLH genes. They reported one E12/E47 ortholog, and three Group A bHLHs (Simionato et al., 2007). One of the latter was later shown to be expressed in a cell population that may give rise to putative sensory cells in the sponge (Richards et al., 2008). Recently, Fortunato et al. (2016) reported a genome-wide survey in another poriferan, *Sycon ciliatum*. They reported 18 Group A bHLH genes, including two genes that belong to E12/E47 and some other putative orthologs for bilaterian bHLH genes, such as *Hand* (Fortunato et al., 2016). Therefore, the heterodimeric character of Group A bHLHs seems to be as old as the origin of Group A itself. Group A bHLH information reported in ctenophores is minimal at present.

A Group E factor, HEY, has been reported in the sponges, *A. queenslandica* (Simionato et al., 2007) and *S. ciliatum* (Fortunato et al., 2016). *A. queenslandica* HEY has an FRPW motif, which resembles the YRPW motif of mouse and *Drosophila* HEY, at its C-terminus (Gazave et al., 2014). Gazave et al. (2014) reported one HES ortholog with a C-terminal WRPW motif from a homoscleromorph sponge, *Oscarella chimeric*. They also suggest that this sponge has a putative ortholog of HELT, although supporting values in molecular phylogenetic analyses were low (Gazave et al., 2014). There are three HES candidates in the *M. leidyi* genome, all of which have a WRPW motif at their C-termini (Gazave et al., 2014; Copley, 2016). *Groucho* candidates are reported in the genomes of the phyla Porifera and Ctenophora; therefore, it is likely that WRPW motifs are recognized by Groucho proteins (Copley, 2016). Presently no Group D factors have been reported in genomes of any species belong to these two phyla (Simionato et al., 2007; Fortunato et al., 2016).

According to these surveys, great bHLH innovations seem to have accompanied the emergence of basal metazoans. At that time, Groups A and E emerged (Fig. 3). Heterodimerization of E12/E47 factors with other Group A bHLHs also seems to have occurred. HES proteins apparently

interact with Groucho via its C-terminal WRPW motif. Molecular biological studies will be needed to confirm these possible protein-protein interactions. At present there are only a few reports of bHLHs among these basal phyla. bHLH data from *A. queenslandica*, *O. chimeric*, S. *ciliatum*, and *M. leidyi* may not provide adequate insight into bHLH genes in those phyla. Genome-wide surveys of bHLHs should be continued in basal metazoan phyla to further clarify these matters.

(2) A simple appearance with delicate bHLH components: Placozoa

There are more than 20 members of Group A bHLHs in most bilaterians studied so far (Simionato et al., 2007). When and how did Group A factors increase their numbers? When and how did orthologous families such as MyoD and Neurogenin appear and become established? Since most bilaterian orthologous families have not been reported in either the Porifera or the Ctenophora, they probably appeared after separation of these clades from their common metazoan ancestor. Members of two orthologous families, Fer1 and ACSb, have not yet been reported in either of these phyla, but were reported in the *Trichoplax adhaerens* genome (Gyoja, 2014). Moreover, the *Trichoplax* genome contains at least 14 Group A bHLHs (Gyoja, 2014).

For Group E, the placozoan, *T. adhaerens*, has one member each for HES, HEY, and HELT in its genome (Gazave et al., 2014; Gyoja, 2014). There is a C-terminal WRPW motif in the HES predicted protein, although I could not find a YRPW motif in HEY (Gyoja, 2014). Presently no Group D factors have been reported in the placozoan genome either (Gyoja, 2014).

Since the body plan of *T. adhaerens*, the sole member of the Phylum Placozoa, seems to lack a nervous system and mesodermal-like tissues, its complex bHLH components are curious. At present, we do not know the reason, but this organism may have acquired its simple body plan secondarily, or it may have unknown tissue(s) in its life cycle (Srivastava et al., 2008).

(3) Cnidaria: Similar to, but also quite different from the Bilateria

The number of Group A and E bHLH components in *T. adhaerens* is still smaller than that in bilaterians. What about the Phylum Cnidaria, a sister clade of the Bilateria, which contains gelatinous animals such as jellyfish, corals, hydras, and sea anemones? Actually, they reportedly have as many Group A bHLH components as do bilaterians (Simionato et al., 2007; Gyoja et al., 2012). There are approximately 30 members of Group A bHLHs in genomes of the sea anemone, *Nematostella vectensis,* and the coral, *Acropora digitifera* (Simionato et al., 2007; Gyoja et al., 2012). They, along with bilaterians, have members of many orthologous families such as ASCa, Twist, and SCL (Simionato et al., 2007; Gyoja et al., 2012; Fig. 3). Spring et al. (2000) reported the expression pattern of *Twist* in a jellyfish, *Podocoryne carnea,* by whole mount *in situ* hybridization. This gene is expressed in myoepithelial cells of the polyp, and later in mesoderm-like entocodon cells, indicating active transcription in proliferating cells (Spring et al., 2000). This expression pattern probably resembles that of bilaterian *Twist* genes (Spring et al., 2000).

In both number and orthology of Group A bHLHs, cnidarian genomes are much more similar to bilaterian genomes than to other basal metazoans. However, there are many differences between them. Group A genes such as *MyoD*, *NeuroD*, *Neurogenin*, and *Atonal* have not been reported in cnidarian genomes (Simionato et al., 2007; Gyoja et al., 2012). Although Müller et al. (2003) reported a *MyoD*-like gene in *P. carnea*, similarity between its predicted amino acid sequence and those of bilaterian MyoDs is very low. In addition, recent genome-wide surveys in other cnidarian species failed to find *MyoD* candidates (Simionato et al., 2007; Gyoja et al., 2012). Therefore, at present it appears that cnidarians likely lack *MyoD*, although comprehensive surveys and molecular phylogenetic studies in cnidarians should be continued to obtain more information.

On the other hand, many cnidarian Group A members seem to lack orthologs in any known bilaterian genome (Simionato et al., 2007; Gyoja et al., 2012). In summary, cnidarian genomes contain as many Group A bHLHs as bilaterian genomes, but many cnidarian bHLH factors have no bilaterian orthologs. This implies that Group A bHLHs first increased their numbers, and conserved

orthologous families became established later (Gyoja, 2014; Fig. 3). Nevertheless, it is possible that part of the expansion may have occurred independently in those two clades.

There is another interesting trait in cnidarian bHLHs. Some cnidarians have many copies of *HES*. For example, *N. vectensis* has least 11 *HES* copies in its genome (Simionato et al., 2007). *A. digitifera* has at least eight *HES* copies (Gyoja et al., 2012). Six copies of HES have a WRPW motif at the C-termini of the predicted proteins, while no such motif could be detected in the remaining two (Gyoja et al., 2012). Both cnidarian genomes have one *HEY* gene along with a few unclassified Group E genes (Simionato et al., 2007; Gyoja et al., 2012). A similar situation can be found in bilaterians. Many bilaterians also have multiple *HES* paralogs. Sometimes the copy number exceeds ten, in cases such as the amphioxus, *Branchiostoma floridae* and *D. melanogaster* (Ledent and Vervoort, 2001; Simionato et al., 2007; Gazave et al., 2014). Therefore, *HES* has tended to increase its copy number as more Group A members emerged.

As mentioned above, no Group D factors have been reported yet in the phyla Porifera, Ctenophora, or Placozoa. The Cnidaria is the only non-bilaterian phylum in which Group D factors have been reported (Fig. 3). Both *N. vectensis* and *A. digitifera* have one member each of *EMC/ID* and another Group D member, *Pearl*, in their genomes (Simionato et al., 2007; Gyoja et al., 2012). Group D is therefore probably an innovation of the common ancestor of cnidarians and bilaterians. Evolution of Group A seems to have been tightly linked to evolution of its negative regulators.

III-4 Conservative Fellows: Bilateria

The Bilateria comprises more than 30 phyla, including the Arthropoda, Nematoda, Mollusca, Annelida, Echinodermata, and Vertebrata (For the Phylum Vertebrata, see Satoh et al., 2014; Satoh 2016). bHLH components have been surveyed genome-wide in many bilaterians. Many of them, including *H. sapiens*, *B. floridae*, *Capitella teleta*, and *Lottia gigantea* have a conserved repertoire of bHLH factors (Ledent et al., 2002; Simionato et al., 2007). Most bilaterians seem to retain ancestral bHLH gene families that the Urbilateria, a hypothetical common ancestor of all bilaterians, had

in its genome (Simionato et al., 2007). Furthermore, there seem to be only small scale innovations of Group A bHLH components after the emergence of bilaterian phyla (Simionato et al., 2007). Notably, there are several exceptions, such as *C. elegans* and an ascidian, *Ciona intestinalis* (Ledent and Vervoort, 2001; Satou et al., 2003). These have lost several bHLH orthologous families and developed several unassigned bHLHs, unlike other bilaterians (Ledent and Vervoort, 2001; Satou et al., 2003).

Group E bHLHs, especially *HES*, seem to be less evolutionarily conservative. As described above, there are variable numbers of *HES* genes in bilaterian genomes (Ledent and Vervoort, 2001; Ledent et al., 2002; Simionato et al., 2007; Gazave et al., 2014). Some bilaterians, such as *B. floridae* and *D. melanogaster*, have more than ten *HES* genes in their genomes, while there is only a single copy in the *C. elegans* genome, and three in that of *C. intestinalis* (Ledent and Vervoort, 2001; Satou et al., 2003). Ambiguous molecular phylogenetic relationships within the *HES* clade makes this evolutionary variability more complicated (Satou et al., 2003; Gyoja, 2014; Gazave et al., 2014). HES proteins from various metazoan taxa tend to cluster in molecular phylogenetic analyses, suggesting that they arose from a single ancestral gene, although statistical support is often low (Simionato et al., 2007; Gyoja, 2014; Gazave et al., 2014). However, resolution tends to be so low within the HES clade that in most cases it is difficult to tell whether *HES* members in one species are descendants from an ancient *HES* population of almost the same size, or whether they originated from one ancestral *HES* gene. Molecular phylogenetic analyses by Gazave et al. (2014) seem to support the latter scenario. What this evolutionary character of *HES* means is an important question for bilaterian bHLH research.

IV Hypotheses about roles of Group A, D, and E bHLHs in Metazoan Evolution

There are pivotal questions yet to be answered in metazoan bHLH research. Group A and E bHLHs and multicellular metazoan organisms emerged at around the same time. Was there a functional connection between those events? What were the initial function(s) of those novel groups? Did they

contribute to mesodermal and/or neural tissue invention in the course of evolution? If so, how? Taking recent studies into account, in subsequent sections I will present hypotheses relative to these questions.

(1) One of the initial functions of Group A and E bHLHs may have been cell proliferation control

As mentioned above, Group A and E bHLHs have very ancient origins in metazoan evolution. Molecular interactions characterizing them seem equally ancient. It is likely that they had some important function(s) in multicellularization.

If so, what were the initial function(s) of Group A and E bHLHs? As described above, both of these groups are thought to have originated from Group B (Fig. 1; Sebé-Pedrós et al., 2011). Group B members are involved in fundamental biological processes such as metabolism and cell proliferation in both metazoans and yeasts (Massari and Murre, 2000 and references therein; Jones, 2004). For example, a Group B factor, *myc*, often together with its dimerization partner *max*, strongly promotes cell proliferation and represses differentiation in mice and *Drosophila* (For a review, see Eilers and Eisenman 2008). Since development of an ordered multicellular animal body requires strict cell proliferation control, one of the initial functions of Group A and E bHLHs may have been cell proliferation control. Group E bHLHs may have promoted cell proliferation and/or may have repressed Group A bHLH transcriptional activity (Fig. 4a). As development proceeded, cells likely exited from this proliferation phase and shifted to a differentiation phase, probably initiated by Group A activators (Fig. 4b, c). Another, not mutually exclusive, possibility is that ancient Group E bHLHs inhibited metabolism. A recent study by Hashimshony et al. (2015) showed that endoderm is the most ancient germ layer, and that its characters, like metabolic control, originated from similar functions already present in ancestral unicellular organisms. Once they became multicellular, organisms probably acquired additional cell population(s) like ectoderm, and this novel differentiation process may have been achieved by repression of “endodermal” genes by Group E

bHLHs. Recently, Yasuoka et al. (2016) showed that a T-box gene, *Brachyury*, activates ectodermal genes and represses endodermal genes to maintain the ectoderm-endoderm border in the coral, *Acropora digitifera*. They suggest that this "demarcation" activity may have been the ancestral function of this transcription factor (Yasuoka et al., 2016). Functionally, this suggestion is somewhat similar to my hypothesis above.

(2) Controlled transition from an undifferentiated, proliferating state, to differentiation may have enabled mesodermal and/or neural tissue formation

Group A, D, and E bHLHs/HLHs are deeply involved in mesodermal and neural tissue formation in extant bilaterians. One of the most important questions regarding bHLHs/HLHs in metazoan evo-devo is why did Group A, D, and E factors become especially involved in differentiation of mesodermal and neural tissues? Are there differences between tissues that are deeply regulated by these subsets of bHLH/HLHs and other tissues? Recently Hashimshony et al. (2015) suggested that genes with evolutionarily conserved endodermal expression, tend to be expressed early in development, followed by such genes for ectoderm. Such genes for mesoderm tend to be expressed comparatively later (Hashimshony et al., 2015). This trait seems to be conserved among several distant metazoan clades (Hashimshony et al., 2015). This probably means that tissues that evolved later tend to require a longer, undifferentiated, proliferating state than their progenitor cells in development, while differentiation and/or fate restriction of cell populations such as endoderm and/or ectoderm, with more ancient evolutionary origins, starts earlier. This "retarded differentiation" of novel tissues probably requires strict maintenance of an undifferentiated proliferating state, partly by prevention of endodermal and epidermal differentiation, as well as a regulated conversion system from proliferation to differentiation. "Retarded differentiation" may be required because morphology of cells that constitute "novel" tissues, such as skeletal muscle or nervous system, are often so specialized that it is difficult for those cells to actively divide once they have become differentiated. As de-

scribed before, Group E factors can repress Group A activators both by binding directly to their enhancer DNAs and through physical interactions with them (Fig. 2d, 2e and 2f). The latter mechanism may have been important for cell fate specification before the "retarded differentiation" took place, because it allows expression of differentiation factor(s) in an inactive state in cells. "More ancient" endodermal and epidermal tissues possibly also required this conversion system. But these tissues likely started to differentiate earlier (Hashimshony et al., 2015) and may not have required strong suppression systems for differentiation into other types of tissue. Therefore, systems for maintenance of an undifferentiated state and for the transition to differentiation have probably remained simple. Cells comprising them perhaps have been able to divide after initiation of differentiation, when their cell populations are still small and cell morphology is simple. My hypothesis is that Group A and E bHLHs, which initially controlled cell proliferation during development in ancient multicellular animals (Fig. 4a), were recruited for formation of novel tissues later in evolution because their functions were suited for a sophisticated proliferation/differentiation control system (Fig. 4b, c). Group D HLHs likely attended this system later. Such a view also offers a possible explanation for mesodermal-like and/or neural tissues in the Ctenophora and/or the Cnidaria. Those tissues may be not homologous to bilaterian mesodermal and/or neural tissues in a strict sense. Nevertheless, they may share some features with bilaterian mesodermal and/or neural tissues, concerning this switching system, in which Group D HLHs and/or E bHLHs suppress functions of Group A differentiation factors.

Duplication and diversification of regulatory genes have been proposed as important in innovation of novel cell types in multicellular animals (For reviews, see Pan et al., 2012; Arendt et al., 2016). It is noteworthy that, among six High Order Groups, Group A and E members expanded largely during early metazoan evolution, while other Groups expanded only slightly (Ledent and Vervoort, 2001; Simionato et al., 2007; Sebé-Pedrós et al., 2011; Gyoja 2014). As I suggested, members of Group A and E may have been recruited frequently for novel tissue formation in early metazoans. If a member of Group A or E duplicated to yield two copies, one of those copies may

have acquired a new function in formation of novel tissues, allowing both copies to survive. The gradual increase of Group A and E members in metazoan evolution and the emergence of Group D may have finally enabled an explosion of bilaterian phyla with sophisticated mesoderm and nervous systems that could be invested with capacities for locomotion, feeding, escaping, and so on. Experimental approaches, especially in basal metazoans, will likely provide some important keys to test and refine these hypothetical roles for Group A, D, and E bHLHs/HLHs in metazoan evolution.

V Perspectives in "orthologous family" evolution

Many important questions in the bHLH evo-devo field have yet to be resolved. For example, what factors established and maintained orthologous families in the course of evolution? As described above, many extant Group A orthologous families seem to have been established after the expansion of Group A members. Most extant bilaterians still have an ancient set of orthologous protein families that urbilaterians may have had in their genomes, with slight modifications (Simionato et al., 2007). Once an orthologous family became established, it should have been tightly retained in many cases. Expansion of Group A may have enabled complex, refined animal development, involving differentiation of novel tissues, as described above. Probably, amino acid sequences of proteins became tightly conserved, allowing only neutral molecular evolution, as specialized function(s) of the protein became requisite. Nevertheless, what the nature of those "specialized functions" was and how they became requisite, remains to be solved. Why several Group A factors became conservatively shared by bilaterians, but not by poriferans and diploblasts, is also an intriguing question. For example, bHLH factors, the expression of which promotes ectodermal cells to initiate the development of a neural lineage, are called proneural genes (For a review, see Bertrand et al., 2002). Three proneural orthologous families are known either in mammals and/or *Drosophila*: ACSa, Neurogenin, and Atonal (Bertrand et al., 2002). While ACSa is reported in both bilaterians and cnidarians, the latter two have been reported only in bilaterians (Ledent and Vervoort,

2001; Simionato et al., 2007). They may have enabled emergence of more complicated and sophisticated bilaterian neural tissues. By considering these questions, we may gain insight into why many extant bilaterian phyla became established during a relatively short period (the Cambrian Explosion) (Peterson et al., 2004; Simakov et al., 2015).

There is another important question: Why are there exceptions among bilaterians, such as *C. elegans* and *C. intestinalis*, described above? Since development of *C. elegans* proceeds rapidly and an adult worm is composed of a surprisingly small number of cells, the transition system from an undifferentiated proliferating state to a differentiated state may have been less important than for most other bilaterians. From this point of view, it is interesting that this animal seems to lack *EMC* homologs in its genome (Massari and Murre, 2000; Simionato et al., 2007). Larvae of *C. intestinalis* are also comprised of a small number of cells, although this species forms an adult body through active cell proliferation after metamorphosis (Satoh, 2014). I suggest that seeking to answer these questions, together with testing my hypothesis, will provide more profound insights into metazoan development and evolution.

Concluding Remarks

Recent genome-wide studies have greatly advanced our knowledge of bHLH components in a wide range of metazoans. I hypothesize that interactions between Group A and E, which maintain cells in an undifferentiated, proliferating state, was essential in development of ancient multi-cellular animals. When development of an animal reaches an appropriate stage, those cells began to differentiate, probably by recruiting Group A factor(s). I propose that recruitment of this conversion system may have enabled emergence of more complex and refined animal body plans comprising novel tissues. Testing this hypothesis and solving other questions concerning bHLH evolution will support fascinating studies in metazoan evo-devo.

Acknowledgements

I am deeply grateful to Dr. Nori Satoh for encouraging me to write this review and for critical reading of the manuscript. I also thank Dr. Hiroki Nishida at Osaka University and his laboratory members for their kind hospitality, since most of this manuscript was prepared while I was there. I thank Dr. Steven D. Aird for editing my English, and two anonymous reviewers for helpful comments on the manuscript.

Conflict of Interests

The author declares no conflicts of interest.

References

Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, Pavlicev M, Schlosser G, Widder S, Laubichler MD, Wagner GP. 2016. The origin and evolution of cell types. Nat Rev Genet 17: 744-757.

Atchley WR, Fitch WM. 1997. A natural classification of the basic helix-loop-helix class of transcription factors. Proc Natl Acad Sci U S A 94: 5172-5176.

Baek JH, Hatakeyama J, Sakamoto S, Ohtsuka T, Kageyama R. 2006. Persistent and high levels of Hes1 expression regulate boundary formation in the developing central nervous system. Development 133: 2467-2476.

Balagopalan L, Keller CA, Abmayr SM. 2001. Loss-of-function mutations reveal that the *Drosophila nautilus* gene is not essential for embryonic myogenesis or viability. Dev Biol 231: 374-382.

Bao Y, Xu F, Shimeld SM. 2017. Phylogenetics of lophotrochozoan bHLH genes and the evolution of lineage-specific gene duplicates. Genome Biol Evol 9: 869-886.

Benezra R, Davis RL, Lockshon D, Turner DL, Weintraub H. 1990. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell 61: 49-59.

Bertrand N, Castro DS, Guillemot F. 2002. Proneural genes and the specification of neural cell types. Nat Rev Neurosci 3: 517-530.

Botas J, Moscoso del Prado J, García-Bellido A. 1982. Gene-dose titration analysis in the search of trans-regulatory genes in *Drosophila*. EMBO J 1: 307-310.

Campos-Ortega JA, Knust E. 1990. Molecular analysis of a cellular decision during embryonic development of *Drosophila melanogaster*: epidermogenesis or neurogenesis. Eur J Biochem 190: 1-10.

Campuzano S, Modolell J. 1992. Patterning of the *Drosophila* nervous system: the *achaete-scute* gene complex. Trends Genet 8: 202-208.

Castro DS, Martynoga B, Parras C, Ramesh V, Pacary E, Johnston C, Drechsel D, Lebel-Potter M, Garcia LG, Hunt C, Dolle D, Bithell A, Ettwiller L, Buckley N, Guillemot F. 2011. A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. Genes Dev 25: 930-945.

Copley RR. 2016. The Unicellular Ancestry of Groucho-Mediated Repression and the Origins of Metazoan Transcription Factors. Genome Biol Evol 8: 1859-1867.

Cubas P, Modolell J. 1992. The *extramacrochaetae* gene provides information for sensory organ patterning. EMBO J 11: 3385-3393.

Cronmiller C, Cummings CA. 1993. The *daughterless* gene product in *Drosophila* is a nuclear protein that is broadly expressed throughout the organism during development. Mech Dev 42: 159-169.

Davis RL, Cheng PF, Lassar AB, Weintraub H. 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. Cell 60: 733-746.

Demilly A, Simionato E, Ohayon D, Kerner P, Garcès A, Vervoort M. 2011. *Coe* genes are expressed in differentiating neurons in the central nervous system of protostomes. PLoS One 6: e21213.

Dubois L, Vincent A. 2001. The COE--Collier/Olf1/EBF--transcription factors: structural conservation and diversity of developmental functions. Mech Dev 108: 3-12.

Eilers M, Eisenman RN. 2008. Myc's broad reach. Genes Dev. 22: 2755-2766.

Fischer A, Gessler M. 2007. Delta-Notch--and then? Protein interactions and proposed modes of repression by Hes and Hey bHLH factors. Nucleic Acids Res 35: 4583-4596.

Fisher AL, Ohsako S, Caudy M. 1996. The WRPW motif of the hairy-related basic helix-loop-helix repressor proteins acts as a 4-amino-acid transcription repression and protein-protein interaction domain. Mol Cell Biol 16: 2670-2677.

Fortunato SA, Vervoort M, Adamski M, Adamska M. 2016. Conservation and divergence of bHLH genes in the calcisponge *Sycon ciliatum*. Evodevo 7: 23.

Fritzsch B, Jahan I, Pan N, Elliott KL. 2015. Evolving gene regulatory networks into cellular networks guiding adaptive behavior: an outline how single cells could have evolved into a centralized neurosensory system. Cell Tissue Res 359: 295-313.

Gazave E, Guillou A, Balavoine G. 2014. History of a prolific family: the *Hes*/*Hey*-related genes of the annelid *Platynereis*. Evodevo 5: 29.

Ghysen A, Dambly-Chaudière C. 1988. From DNA to form: the *achaete-scute* complex. Genes Dev 2: 495-501.

Ghysen A, Dambly-Chaudière C. 1989. Genesis of the *Drosophila* peripheral nervous system. Trends Genet 5: 251-255.

Giagtzoglou N, Alifragis P, Koumbanakis KA, Delidakis C. 2003. Two modes of recruitment of E(spl) repressors onto target genes. Development 130: 259-270.

Guillemot F, Lo LC, Johnson JE, Auerbach A, Anderson DJ, Joyner AL. 1993. Mammalian *achaete-scute* homolog 1 is required for the early development of olfactory and autonomic neurons. Cell 75: 463-476.

Gyoja F. 2014. A genome-wide survey of bHLH transcription factors in the Placozoan *Trichoplax adhaerens* reveals the ancient repertoire of this gene family in metazoan. Gene 542: 29-37.

Gyoja F, Kawashima K, Satoh N. 2012. A genomewide survey of bHLH transcription factors in the coral *Acropora digitifera* identifies three novel orthologous families, pearl, amber, and peridot. Dev Genes Evol 222: 63-76.

Gyoja F, Satoh N. 2013. Evolutionary aspects of variability in bHLH orthologous families: insights from the pearl oyster, *Pinctada fucata*. Zoolog Sci 30: 868-876.

Hatakeyama J, Bessho Y, Katoh K, Ookawara S, Fujioka M, Guillemot F, Kageyama R. 2004. *Hes* genes regulate size, shape and histogenesis of the nervous system by control of the timing of neural stem cell differentiation. Development 131: 5539-5550.

Hashimshony T, Feder M, Levin M, Hall BK, Yanai I. 2015. Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. Nature 519: 219-222.

Hirata H, Yoshiura S, Ohtsuka T, Bessho Y, Harada T, Yoshikawa K, Kageyama R. 2002. Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop. Science 298: 840-843.

Imayoshi I, Kageyama R. 2014. bHLH factors in self-renewal, multipotency, and fate choice of neural progenitor cells. Neuron 82: 9-23.

Iso T, Sartorelli V, Poizat C, Iezzi S, Wu HY, Chung G, Kedes L, Hamamori Y. 2001. HERP, a novel heterodimer partner of HES/E(spl) in Notch signaling. Mol Cell Biol 21: 6080-6089.

Jan YN, Jan LY. 1993. HLH proteins, fly neurogenesis, and vertebrate myogenesis. Cell 75: 827-830.

Jones S. 2004. An overview of the basic helix-loop-helix proteins. Genome Biol 5: 226.

Kageyama R, Ohtsuka T, Kobayashi T. 2007. The Hes gene family: repressors and oscillators that orchestrate embryogenesis. Development 134: 1243-1251.

Kang SA, Seol JH, Kim J. 2005. The conserved WRPW motif of Hes6 mediates proteasomal degradation. Biochem Biophys Res Commun 332: 33-36.

Lassar AB, Thayer MJ, Overell RW, Weintraub H. 1989. Transformation by activated *ras* or *fos* prevents myogenesis by inhibiting expression of MyoD1. Cell 58: 659-667.

Ledent V, Paquet O, Vervoort M. 2002. Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biol 3: 0030.1-0030.18.

Ledent V, Vervoort M. 2001. The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. Genome Res 11: 754-770.

Lu J, Webb R, Richardson JA, Olson EN. 1999. MyoR: a muscle-restricted basic helix-loop-helix transcription factor that antagonizes the actions of MyoD. Proc Natl Acad Sci U S A 96: 552-557.

Lyden D, Young AZ, Zagzag D, Yan W, Gerald W, O'Reilly R, Bader BL, Hynes RO, Zhuang Y, Manova K, Benezra R. 1999. Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. Nature 401: 670-677.

Massari ME, Murre C. 2000. Helix-Loop-Helix proteins: regulators of transcription in eukaryotic organisms. Mol Cell Biol 20: 429-440.

Müller P, Seipel K, Yanze N, Reber-Müller S, Streitwolf-Engel R, Stierwald M, Spring J, Schmid V. 2003. Evolutionary aspects of developmentally regulated helix-loop-helix transcription factors in striated muscle of jellyfish. Dev Biol 255: 216-229.

Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabrera CV, Buskin JN, Hauschka SD, Lassar AB, Weintraub H, Baltimore D. 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell 58: 537-544.

Norton JD. 1990. ID helix-loop-helix proteins in cell growth, differentiation and tumorigenesis. J Cell Sci 113: 3897-3905.

Olson EN. 1990. MyoD family: a paradigm for development? Genes Dev 4: 1454-1461.

Pan N, Kopecky B, Jahan I, Fritzsch B. 2012. Understanding the evolution and development of neurosensory transcription factors of the ear to enhance therapeutic translation. Cell Tissue Res 349: 415-432.

Peterson KJ, Lyons JB, Nowak KS, Takacs CM, Wargo MJ, McPeek MA. 2004. Estimating metazoan divergence times with a molecular clock. Proc Natl Acad Sci U S A 101: 6536-6541.

Pisani D, Pett W, Dohrmann M, Feuda R, Rota-Stabelli O, Philippe H, Lartillot N, Wörheide G. 2015. Genomic data do not support comb jellies as the sister group to all other animals. Proc Natl Acad Sci U S A 112: 15402-15407.

Richards GS, Simionato E, Perron M, Adamska M, Vervoort M, Degnan BM. 2008. Sponge genes provide new insight into the evolutionary origin of the neurogenic circuit. Curr Biol 18: 1156–1161.

Roberts VJ, Steenbergen R, Murre C. 1993. Localization of E2A mRNA expression in developing and adult rat tissues. Proc Natl Acad Sci U S A 90: 7583-7587.

Rudnicki MA, Schnegelsberg PN, Stead RH, Braun T, Arnold HH, Jaenisch R. 1993. MyoD or Myf-5 is required for the formation of skeletal muscle. Cell 75: 1351-1359.

Ryan JF, Pang K, Schnitzler CE, Nguyen AD, Moreland RT, Simmons DK, Koch BJ, Francis WR, Havlak P, NISC Comparative Sequencing Program, Smith SA, Putnam NH, Haddock SH, Dunn CW, Wolfsberg TG, Mullikin JC, Martindale MQ, Baxevanis AD. 2013. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. Science 342: 1242592.

Sasai Y, Kageyama R, Tagawa Y, Shigemoto R, Nakanishi S. 1992. Two mammalian helix-loop-helix factors structurally related to *Drosophila hairy* and *Enhancer of split*. Genes Dev 6: 2620-2634.

Satoh N. 2014. Developmental Genomics of Ascidians. New York: Wiley Blackwell. p145-158.

Satoh N. 2016. Chordate Origins and Evolution. Elsevier/Academic Press.

Satoh N, Rokhsar D, Nishikawa T. 2014. Chordate evolution and the three-phylum system. Proc Biol Sci 281: 20141729.

Satou Y, Imai KS, Levine M, Kohara Y, Rokhsar D, Satoh N. 2003. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. I. Genes for bHLH transcription factors. Dev Genes Evol 213: 213-221.

Sebé-Pedrós A, Mendoza A, Lang BF, Degnan BM, Ruiz-Trillo I. 2011. Unexpected repertoiere of metazoan transcription factors in the unicellular holozoan *Capsaspora owczarzaki*. Mol Biol Evol 28: 1241-1254. Nat. Genet. 36: 1301-1305.

Simakov O, Kawashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, Hisata K, Bredeson J, Shoguchi E, Gyoja F, Yue JX, Chen YC, Freeman RM Jr, Sasaki A, Hikosaka-Katayama T, Sato A, Fujie M, Baughman KW, Levine J, Gonzalez P, Cameron C, Fritzenwanker JH, Pani AM, Goto H, Kanda M, Arakaki N, Yamasaki S, Qu J, Cree A, Ding Y, Dinh HH, Dugan S, Holder M, Jhangiani SN, Kovar CL, Lee SL, Lewis LR, Morton D, Nazareth LV, Okwuonu G, Santibanez J, Chen R, Richards S, Muzny DM, Gillis A, Peshkin L, Wu M,

Humphreys T, Su YH, Putnam NH, Schmutz J, Fujiyama A, Yu JK, Tagawa K, Worley KC, Gibbs RA, Kirschner MW, Lowe CJ, Satoh N, Rokhsar DS, Gerhart J. 2015. Hemichordate genomes and deuterostome origins. Nature 527: 459-465.

Simionato E, Ledent V, Richards G, Thomas-Chollier M, Kerner P, Coornaert D, Degnan BM, Vervoort M. 2007. Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. BMC Evol Biol 7: 33.

Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Di Franco A, Roure B, Satoh N, Quéinnec É, Ereskovsky A, Lapébie P, Corre E, Delsuc F, King N, Wörheide G, Manuel M. 2017. A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. Curr Biol 27: 958-967.

Skeath JB, Carroll SB. 1991. Regulation of *achaete-scute* gene expression and sensory organ pattern formation in the *Drosophila* wing. Genes Dev 5: 984-995.

Southall TD, Brand AH. 2009. Neural stem cell transcriptional networks highlight genes essential for nervous system development. EMBO J 28: 3799-3807.

Spicer DB, Rhee J, Cheung WL, Lassar AB. 1996. Inhibition of myogenic bHLH and MEF2 transcription factors by the bHLH protein Twist. Science 272: 1476-1480.

Spring J, Yanze N, Middel AM, Stierwald M, Gröger H, Schmid V. 2000. The mesoderm specification factor twist in the life cycle of jellyfish. Dev Biol 228: 363-375.

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 ME, 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.

Suga H, Chen Z, de Mendoza A, Sebé-Pedrós A, Brown MW, Kramer E, Carr M, Kerner P, Vervoort M, Sánchez-Pons N, Torruella G, Derelle R, Manning G, Lang BF, Russ C, Haas BJ, Roger AJ, Nusbaum C, Ruiz-Trillo I. 2013. The *Capsaspora* genome reveals a complex unicellular prehistory of animals. Nat Commun 4: 2325.

Tomita K, Moriyoshi K, Nakanishi S, Guillemot F, Kageyama R. 2000. Mammalian *achaete-scute* and *atonal* homologs regulate neuronal versus glial fate determination in the central nervous system. EMBO J 19: 5460-5472.

Van Doren M, Ellis HM, Posakony JW. 1991. The *Drosophila extramacrochaetae* protein antagonizes sequence-specific DNA binding by *daughterless*/*achaete-scute* protein complexes. Development 113: 245-255.

Weintraub H, Tapscott SJ, Davis RL, Thayer MJ, Adam MA, Lassar AB, Miller AD. 1989. Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. Proc Natl Acad Sci U S A 86: 5434-5438.

Weintraub H, Dwarki VJ, Verma I, Davis R, Hollenberg S, Snider L, Lassar A, Tapscott SJ. 1991. Muscle-specific transcriptional activation by MyoD. Genes Dev 5: 1377-1386.

Yasuoka Y, Shinzato C, Satoh N. 2016. The Mesoderm-Forming Gene *brachyury* Regulates Ectoderm-Endoderm Demarcation in the Coral *Acropora digitifera*. Curr Biol 26: 2885-2892.

Figure Legends

Figure 1

A simplified diagram, indicating that Group A, C, and E bHLHs originated from ancient members of Group B. Phylogenetic relationships shown here are only schematic and do not necessarily reflect actual phylogeny. Groups D and F are not shown because they are HLH genes lacking basic regions. They have often been excluded from molecular phylogenetic studies. Groups A, C, and E are nearly monophyletic, although two members of Group B, *AP4* and *Fig-α*, tend to cluster with Group A (Simionato et al., 2007; Sebé-Pedrós et al., 2011). Group B members are paraphyletic (Simionato et al., 2007; Sebé-Pedrós et al., 2011). For more detail, see Sebé-Pedrós et al. (2011).

Figure 2

Several ways to repress Group A activators. Most diagrams are based on Massari and Murre (2000), Kageyama et al. (2007) and Imayoshi and Kageyama (2014). (a) A heterodimer comprising a Group A activator (magenta) and an E-protein can bind to an E-box to activate transcription, thereby promoting tissue differentiation and suppressing cell proliferation. (b) A repressive Group A factor (dark purple) forms a dimer with an E-protein. In some cases, this dimer can bind an E-box, without activating transcription. By this reaction, binding of a Group A activator to an E-protein, an E-box is antagonized. (c) A Group D protein (pale blue) forms a dimer with an E-protein. By this means, dimerization of Group A activators and E-proteins can be inhibited. (d) Group E factors HES and HEY (dark blue) can form both homo- and heterodimers. These dimers recognize specific target DNA sequences, such as an N-box. A WRPW motif at the HES C-terminus recruits a strong transcription repressor, Groucho. This complex represses transcription of target genes, including Group A activators. (e) A Group E factor can make a dimer with a Group A protein such as E-protein, inhibiting formation of Group A-Group A heterodimer. (f) An E protein molecule forms a complex with a Group A protein molecule. This complex binds to an E-box sequence and represses transcription, which also requires Groucho.

Figure 3

A simple diagram showing phylogeny of metazoan and its sister group Choanoflagellida, based on Simion et al. (2017). Presumed positions of emergence of metazoan specific High Order Groups (magenta), emergence of some orthologous families (blue) and expansion of High Order Groups members (green) are also shown.

Figure 4

A hypothesis showing roles of Group A and E bHLHs in metazoan evolution. (a) A virtual unicellular ancestor state. The dark brown oval represents a single cell with endodermal-like character. Arrows represent cell divisions. This virtual ancestor likely proliferated continuously under suitable conditions for survival. (b) Development of a virtual multicellular ancestor with endoderm-like and epidermis-like tissues. The white oval represents an undifferentiated cell and a light brown oval represents a differentiated, epidermis-like cell. A Group E bHLH, represented by a magenta square, may have acted to maintain a relatively undifferentiated state by repressing a Group A factor, represented by a light brown "A." When Group E functions were eliminated, the Group A factor activated cell differentiation, which is represented by a light brown oval. For simplification, numbers of cell divisions shown here are very small. Conversion from an undifferentiated, proliferating state to differentiation likely occurred more gradually. (c) Development of a virtual, multicellular ancestor with multiple novel tissues. Orange ovals represent differentiated cells, while red ovals represent cells of another kind. Because these novel tissues likely required elongated, undifferentiated states and a more strictly controlled switching system from undifferentiated state to differentiation, Group A and E factors may have been recruited for these purposes.