www.elsevier.com/locate/developmentalbiology

Origin and evolution of the metazoan non-coding regulatory genome

Federico Gaiti, Andrew D. Calcino, Miloš Tanurdžić, Bernard M. Degnan

PII: S0012-1606(16)30573-5
DOI: http://dx.doi.org/10.1016/j.ydbio.2016.11.013
Reference: YDBIO7298

To appear in: *Developmental Biology*

Received date: 11 September 2016
Revised date: 14 November 2016
Accepted date: 18 November 2016

Cite this article as: Federico Gaiti, Andrew D. Calcino, Miloš Tanurdžić and Bernard M. Degnan, Origin and evolution of the metazoan non-coding regulatory genome, *Developmental Biology,* http://dx.doi.org/10.1016/j.ydbio.2016.11.013

Origin and evolution of the metazoan non-coding regulatory genome

Federico Gaiti[11], Andrew D. Calcino[22], Miloš Tanurdžić[13], and Bernard M. Degnan[1*4]
[1]School of Biological Sciences, University of Queensland, Brisbane, Australia
[2] Department of Molecular Evolution and Development, University of Vienna, Vienna, Austria
FG: f.gaiti@uq.edu.au
ADC: andrew.calcino@univie.ac.at
MT: m.tanurdzic@uq.edu.au
BMD: b.degnan@uq.edu.au
[*]Corresponding author.
Bernard M. Degnan, School of Biological Sciences, University of Queensland, Brisbane QLD 4072 Australia. b.degnan@uq.edu.au

Author ORCIDs:

Abstract

Animals rely on genomic regulatory systems to direct the dynamic spatiotemporal and cell-type specific gene expression that is essential for the development and maintenance of a multicellular lifestyle. Although it is widely appreciated that these systems ultimately evolved from genomic regulatory mechanisms present in single-celled stem metazoans, it remains unclear how this occurred. Here, we focus on the contribution of the non-coding portion of the genome to the evolution of animal gene regulation, specifically on recent insights from non-bilaterian metazoan lineages, and unicellular and colonial holozoan sister taxa. High-throughput next-generation sequencing, largely in bilaterian model species, has led to the discovery of tens of thousands of non-coding RNA genes (ncRNAs), including short, long and circular forms, and uncovered the central roles they play in development. Based on the analysis of non-bilaterian metazoan, unicellular holozoan, and fungal

[1] FG: 0000-0001-5111-8816
[2] ADC: 0000-0002-3956-1273
[3] MT: 0000-0002-7564-0868
[4] BMD: 0000-0001-7573-8518

genomes, the evolution of some ncRNAs, such as Piwi-interacting RNAs, correlates with the emergence of metazoan multicellularity, while others, including microRNAs, long non-coding RNAs, and circular RNAs, appear to be more ancient. Analysis of non-coding regulatory DNA and histone post-translational modifications have revealed that some *cis*-regulatory mechanisms, such as those associated with proximal promoters, are present in non-animal holozoans while others appear to be metazoan innovations, most notably distal enhancers. In contrast, the cohesin-CTCF system for regulating higher-order chromatin structure and enhancer-promoter long-range interactions appears to be restricted to bilaterians. Taken together, most bilaterian non-coding regulatory mechanisms appear to have originated before the divergence of crown metazoans. However, differential expansion of non-coding RNA and *cis*-regulatory DNA repertoires in bilaterians may account for their increased regulatory and morphological complexity relative to non-bilaterians.

This article is part of a Special Issue entitled: Changes in developmental controls underlying the evolution of animal body plans, edited by Dr. Jr-Kai Yu and Dr. Yi-Hsien Su.

Keywords: evolution, gene regulation, metazoans, complexity, *cis*-regulation, non-coding RNAs, enhancers, multicellularity, holozoans

The evolution of metazoan multicellularity must have involved the emergence of genomic regulatory capacities to allow for dynamic spatiotemporal and cell-type specific gene expression. To a great extent, these capacities rely on the interplay between available transcription and other regulatory factors with each other, and with non-coding regulatory DNA and RNA sequences (Davidson and Peter, 2015).

The evolution and diversification of protein-coding genes that regulate the formation and maintenance of animals, namely transcription factors (TFs) and other signaling pathway

components, have been a point of focus since the discovery over 30 years ago that deeply conserved developmental genes populate the genomes of most, if not all, bilaterians (vertebrates, insects, worms and their numerous allies) (Finnerty and Martindale, 1999; Finnerty et al., 2003; Levine and Tjian, 2003; McGinnis et al., 1984; Quiring et al., 1994; Rokas, 2008; Slack et al., 1993). Although many of these TF and signaling gene families evolved after the divergence of metazoan and choanoflagellate lineages and thus are correlated with the evolution of animal multicellularity [*e.g.*, (Degnan et al., 2009; Larroux et al., 2008; Richards and Degnan, 2009)], others have an older origin and are found in unicellular holozoans (de Mendoza et al., 2013; King, 2004; King et al., 2003; Richter and King, 2013; Sebé-Pedrós and de Mendoza, 2015; Sebé-Pedrós et al., 2011; Sebé-Pedrós et al., 2010; Sebé-Pedrós et al., 2012).

The striking conservation of the TF family repertoire in non-bilaterian metazoan lineages (cnidarians, placozoans, ctenophores and sponges) – and to a lesser extent diverse unicellular holozoan lineages (choanoflagellates, filastereans and ichthyosporeans) – suggests that other regulatory features, including *cis*-regulatory DNA, non-coding RNAs (ncRNAs) and histone post-translational modifications (PTMs) (Davidson and Peter, 2015), were instrumental in the evolution of animal multicellularity and complexity (Fig. 1). Here, we review recent insights into the evolution of the non-coding portion of the regulatory genome, specifically seeking to find evidence for its potential contribution to the evolution of the metazoan multicellular condition and body plan complexity.

Evolution of the metazoan non-coding RNA landscape

RNA performs a wide range of developmental and physiological roles beyond its cardinal function in the flow of genetic information (Crick, 1970). The discovery of heterogeneous nuclear RNA, retrotransposons and introns in the 1960s and 1970s provided the first hint that animal cells might contain extensive RNA-based regulatory networks (Berget et al., 1977; Britten and Davidson, 1969, 1971; Chow et al., 1977; Davidson et al., 1977; Warner et al., 1966). The next milestone in metazoan RNA biology was the identification of small regulatory RNAs (*lin-4* and *let-7*) (Lee et al., 1993; Reinhart et al., 2000) as harbingers of the RNA interference pathway (Fire et al., 1998). Since then, the classes of small RNAs have expanded to include small interfering RNAs (siRNAs) that deplete RNA transcript levels, Piwi-interacting RNAs (piRNAs) involved in primarily germ cell differentiation and microRNAs (miRNAs) that regulate multiple steps in gene expression (Morris and Mattick, 2014).

In the last decade, the application of next-generation sequencing (NGS) technologies, largely in bilaterian model species, has revealed that animal genomes encode thousands of long non-coding RNAs (lncRNAs) (Bertone et al., 2004; Cabili et al., 2011; Carninci et al., 2005; Derrien et al., 2012; Djebali et al., 2012; Kapranov et al., 2007; Khalil et al., 2009; Okazaki and al., 2002; Ponting et al., 2009; Ravasi et al., 2006). More recently, circular RNA (circRNAs), transcribed and spliced from exons of coding and non-coding genes, have also been discovered to be abundant in bilaterians (Barrett and Salzman, 2016; Chen, 2016; Salzman, 2016) (Fig. 1).

Although significant progress has been made in elucidating the roles of these ncRNA classes in animal development, their origin and evolution remain unclear. However, recent investigations of ncRNAs in non-bilaterian metazoans and unicellular holozoans are providing insights into their contribution to the evolution of animal multicellularity (Bråte et

al., 2015; Bråte et al., 2016; Calcino, 2015; de Mendoza et al., 2015; Gaiti et al., 2015; Grimson et al., 2008; Sebé-Pedrós et al., 2016).

Innovations in animal RNA interference pathways

Small RNA (sRNA) mediated post-transcriptional regulation, known as RNA interference (RNAi), emerged prior to the divergence of the eukaryotic supergroups (Cerutti and Casas-Mollano, 2006). Thought to have evolved as a system of immunity against parasitic exogenous RNAs (Obbard et al., 2009), RNAi pathways have been co-opted into numerous additional regulatory roles. In addition to the pan-eukaryotic exogenous siRNA and the endogenous siRNA pathways (Svoboda, 2014), most animals possess two additional RNAi pathways – miRNAs and piRNAs.

Animal miRNAs are abundant 21-23 nucleotide (nt) RNAs that share features with miRNAs in plants (Voinnet, 2009), including the single-celled green algae *Chlamydomonas reinhardtii* (Molnar et al., 2007; Zhao et al., 2007) and the multicellular brown algae *Ectocarpus siliculosus* (Cock et al., 2010; Tarver et al., 2015; Tarver et al., 2012), as well as the social amoeba *Dictyostelium discoideum* (Avesson et al., 2012). However, the biogenesis and mode of action of bilaterian miRNAs differs from that of other eukaryotic miRNAs, leading to the widespread inference that metazoan miRNAs evolved independently and, thus, were uniquely instrumental in the evolution of animal multicellularity and complexity (Ivey and Srivastava, 2010; Kosik, 2010; Mattick, 2004; Peterson et al., 2009; Shenoy and Blelloch, 2014).

The discovery and analysis of miRNAs in non-bilaterians and unicellular holozoans has yielded insights into the question of whether animal miRNAs have originated

independently of other miRNA systems. miRNAs and key bilaterian miRNA biogenic enzymes (*i.e*., Drosha and Pasha) are present in sponges and cnidarians, but absent from placozoans (Pasha and miRNAs absent), ctenophores and the sister group to the Metazoa, the choanoflagellates, leading to the initial conclusion that this miRNA system was unique to animals but lost in some lineages (Grimson et al., 2008; Liew et al., 2014; Liew et al., 2016; Maxwell et al., 2012; Moran et al., 2014; Moran et al., 2013; Moroz et al., 2014; Tarver et al., 2012; Wheeler et al., 2009).

However, there are differences between bilaterian, cnidarian and sponge miRNAs. Sponge miRNAs have no primary sequence identity with bilaterian miRNAs. In addition, the predicted pre-miRNA hairpins of the sponge *Amphimedon queenslandica* are longer and structurally variable, and appear to be more similar to plant miRNAs than animal miRNAs (Grimson et al., 2008; Tarver et al., 2012). This variability has been suggested as evidence that *Amphimedon* miRNAs are unlikely to be produced through the conventional and conserved eumetazoan miRNA biogenesis pathway (Robinson et al., 2013; Sperling et al., 2010; Tarver et al., 2015; Wheeler et al., 2009).

Despite deep conservation of many miRNAs within the Bilateria, the cnidarian *Nematostella vectensis* has only one miRNA (*miR-100*) that can be considered homologous to a bilaterian miRNA (Grimson et al., 2008). While deadenylation and translational repression of *Nematostella* target mRNAs does occur as per bilaterian miRNAs (Mauri et al., 2016), target cleavage is also commonly employed (Moran et al., 2014). This siRNA-like silencing mechanism is also common to plant miRNAs and, thus, represents either an ancestral mode of action or a convergent innovation in cnidarians and plants. Further complicating interpretations on the evolution of metazoan miRNAs is the presence of *bona fide* miRNAs and related biogenesis machinery (*i.e.*, Drosha and Pasha) in the unicellular

holozoan *Sphaeroforma arctica* (Bråte et al., 2016). This recent discovery supports the metazoan miRNA system being more ancient than initially perceived. The differences in bilaterian, cnidarian and sponge miRNA systems, and the loss of miRNAs in ctenophores, placozoans and choanoflagellates, suggests that the canonical miRNA system present in bilaterians was not essential for the evolution of metazoan multicellularity.

In bilaterians, piRNAs function in the silencing of transposons within the germline, but also have a role in directing DNA methylation of CpG islands in mammalian germlines (Aravin et al., 2008; Carmell et al., 2007; Kuramochi-Miyagawa et al., 2008), and guiding epigenetic modifications in *C. elegans* and *Drosophila* (Ashe et al., 2012; Klenov et al., 2011; Luteijn et al., 2012; Shirayama et al., 2012; Shpiz et al., 2011; Sienski et al., 2012; Wang and Elgin, 2011). piRNAs with a bilaterian-like ping-pong piRNA biogenesis signature [a bias for a 5' uracil and an adenosine at position 10 (Brennecke et al., 2007; Czech and Hannon, 2016; Gunawardane et al., 2007)] have been reported in cnidarians, sponges and ctenophores, but not in placozoans (Calcino, 2015; Grimson et al., 2008), suggesting the piRNA pathway was a regulatory feature of the last common ancestor of animals.

Long non-coding RNAs as regulatory elements of animal development

Long non-coding RNAs (lncRNAs) may be sense or antisense, intronic, and intergenic with respect to protein-coding genes (Morris and Mattick, 2014). Although they lack obvious protein-coding potential (Guttman et al., 2013; Housman and Ulitsky, 2016; Ingolia et al., 2014), lncRNAs are structurally similar to protein-coding mRNAs and are generally transcribed by RNA polymerase II from genomic loci with similar chromatin states to mRNAs (Guttman et al., 2009). lncRNAs also have CpG islands, complex splicing patterns, 5'-terminal

methylguanosine cap and poly(A) 3'-tails (Quinn and Chang, 2016). However, lncRNAs tend to be shorter, have fewer exons, and appear to be expressed at relatively low levels in comparison to mRNAs (Liu et al., 2016; Quinn and Chang, 2016). Exceptions to these trends exist, with some lncRNAs being up to 100 kb (Lyle et al., 2000) and others being abundant in restricted subnuclear locations (Hutchinson et al., 2007). Unlike most protein-coding sequences, lncRNAs are rapidly evolving and exhibit poor primary sequence similarity between species; orthologous lncRNAs are difficult to identify (Ulitsky, 2016).

Metazoan lncRNAs appear to play a diversity of cellular and developmental roles. Several lncRNAs have been shown to act as decoys that titrate away miRNAs or regulatory proteins (Quinn and Chang, 2016). Other lncRNAs act as scaffolds to bring two or more proteins into a complex or in physical proximity (Engreitz et al., 2016b; Hacisuleyman et al., 2014; Tsai et al., 2010). They also can act as guides to recruit chromatin modifying enzymes and can be required for localization of ribonucleoprotein complexes to specific targets (Quinn and Chang, 2016; Rutenberg-Schoenberg et al., 2016). Early embryonic stages appear to be a period of active transcription of lncRNAs (Bråte et al., 2015; Cabili et al., 2011; Fatica and Bozzoni, 2014; Gaiti et al., 2015; Pauli et al., 2012), perhaps to regulate maternal transcripts or transcription of cell-cycle genes (Hung et al., 2011; Pauli et al., 2011). In other stages and contexts, lncRNAs are often in low abundance, which may reflect their highly restricted expression to a particular cell type, tissue, or developmental stage (Bråte et al., 2015; Cabili et al., 2015; Cabili et al., 2011; Gaiti et al., 2015; Liu et al., 2016; Mercer et al., 2008; Pauli et al., 2012; Ponjavic et al., 2009).

The role of lncRNA in the regulation of developmental gene activity appears to be widespread amongst metazoans (Perry and Ulitsky, 2016). For instance, in *Xenopus*, a subset of lncRNAs are developmentally regulated and spatially localized in the gastrula stage

embryo (Forouzmand et al., 2016; Necsulea et al., 2014; Tan et al., 2013). In zebrafish, many lncRNAs are expressed during embryogenesis (Pauli et al., 2012) and their developmental regulatory functions have been experimentally demonstrated (Ulitsky et al., 2011). In mice, many lncRNAs are essential for normal development and survival (Anderson et al., 2016; Goff et al., 2015; Sauvageau et al., 2013). Developmentally expressed lncRNAs have also been identified in *C. elegans* (Nam and Bartel, 2012), *Drosophila* and other insects (Brown et al., 2014; Chen et al., 2016a; Jayakodi et al., 2015; Jenkins et al., 2015; Quinn et al., 2016; Wu et al., 2016; Young et al., 2012), echinoderms (Mu et al., 2016) and sponges (Bråte et al., 2015; Gaiti et al., 2015). However, given the lack of sequence identity of lncRNAs, it remains unclear if developmental lncRNAs are conserved or independently-evolved.

Evolutionary conservation of animal lncRNAs

Although our understanding of other classes of ncRNAs (*e.g.*, miRNAs) has been improved by evolutionary comparisons (Bartel, 2009), only a handful of functionally homologous bilaterian lncRNAs have been identified and analyzed (Grant et al., 2012; Heard et al., 1999; Migeon et al., 1999; Quinn et al., 2016; Ulitsky et al., 2011). These studies suggest that sequence conservation is not an essential requirement for lncRNA functionality (Pang et al., 2006). The lack of sequence similarity observed between lncRNAs across animal phyla indicates that lncRNA sequences evolve more rapidly than protein-coding sequences (Hezroni et al., 2015; Kapusta and Feschotte, 2014; Kutter et al., 2012; Pang et al., 2006; Ponting et al., 2009). These observations are consistent with negative selection acting on only short patches within lncRNAs (Washietl et al., 2014). Highly conserved elements within lncRNA sequences (micro-homologies), interspersed with longer and less conserved

stretches of nucleotide sequences, have been reported (Chen et al., 2016b; Hezroni et al., 2015; Ulitsky and Bartel, 2013), and include the *miR-7* binding site in the lncRNA *Cyrano* (Ulitsky et al., 2011), the PRC2-binding elements in the lncRNA *Xist* (Maenner et al., 2010), and short repeated sequences in the lncRNA *FIRRE* (Hacisuleyman et al., 2016). Other lncRNA features that appear to be conserved include syntenic relationships to neighboring genes (Amaral et al., 2016; Hezroni et al., 2015; Quinn et al., 2016), secondary structure (Hawkes et al., 2016; Lee et al., 2016; Novikova et al., 2012; Sanbonmatsu, 2016; Tichon et al., 2016), and specific expression patterns (Chodroff et al., 2010; Hezroni et al., 2015; Washietl et al., 2014). The conservation of genomic position of hundreds of deuterostome (*i.e.*, sea urchins and vertebrates) lncRNAs with no detectable sequence similarity (Hezroni et al., 2015) suggests that lncRNA sequences evolve rapidly yet are likely to be functionally conserved.

Akin to bilaterians, where lncRNAs have been shown to be co-expressed with multiple protein-coding genes (Necsulea et al., 2014), sponge lncRNAs appear to be important components of co-expressed developmental gene modules (Bråte et al., 2015; Gaiti et al., 2015). As gene regulatory networks and modules are central for the control and timing of animal development (Davidson and Erwin, 2006; Erwin and Davidson, 2009; Peter and Davidson, 2011), computational construction of evolutionarily conserved modules (or networks) of co-expressed homologous genes, including lncRNAs, might therefore represent a useful approach to predict lncRNA functionality across highly divergent metazoan lineages (Fig. 2). One such example may be the developmental lncRNA co-expressed, and thus presumably co-regulated, with the G protein-coupled receptor *Frizzled B* (a key component of the Wnt signaling pathway in animal development) and other regulatory genes (*e.g.*, *TGF-β*) in the sponges *Amphimedon* and *Sycon* (Bråte et al., 2015; Gaiti et al., 2015), which

diverged from each other at least 650 Mya (Erwin et al., 2011). Although shared co-expression profiles suggest homologous roles for lncRNAs in different species, the independent co-option of non-homologous lncRNAs into these networks cannot be discounted at this time.

Premetazoan origin of lncRNA regulation: A perspective from unicellular holozoans

The recent discoveries of lncRNAs in unicellular holozoan lineages closely related to metazoans has generated additional uncertainty about the evolutionary origin of lncRNAs, and suggests that elaborate lncRNA-based genome regulation is not exclusive to animals and instead is an ancient feature of the holozoan regulatory system. While lncRNAs appear to be expanded in multicellular animals (Bråte et al., 2015; Gaiti et al., 2015; Kapusta and Feschotte, 2014; Ulitsky and Bartel, 2013), several hundred lncRNAs have now been annotated in the filasterean *Capsaspora owczarzaki* and the ichthyosporean *Creolimax fragrantissima* (632 and 692, respectively) (de Mendoza et al., 2015; Sebé-Pedrós et al., 2016). These show both similarities and differences with animal lncRNAs.

Capsaspora has a repertoire of polyadenylated and, in some cases, alternatively spliced lncRNAs. These lncRNAs, despite showing no apparent homology to any animal lncRNA, have temporally dynamic and cell-stage-specific expression patterns, and can be separated into two populations based on their association with specific histone PTMs at their promoters [a high ratio of histone H3 lysine 4 monomethylation (H3K4me1) compared to trimethylation (H3K4me3) and vice versa], as found in metazoans (Marques et al., 2013; Gaiti et al., unpublished). These two populations might represent distinct conserved classes of polyadenylated lncRNA elements with diverse biological functions and, thus, may indicate

that the last common ancestor of animals and these unicellular holozoans had more complex non-coding RNA regulatory capacities than previously thought. Contrarily, despite some global similarities to animal lncRNA regulation, the population of lncRNAs described in *Creolima*x, which represents a lineage that diverged before filasterean *Capsaspora* separated from the stem leading to animals, does not show a major enrichment near developmental genes nor extensive stage-specific expression, both hallmarks of animal lncRNAs (Gaiti et al., 2015; Ulitsky and Bartel, 2013). Nonetheless, the presence of hundreds of lncRNAs in unicellular holozoans is consistent with this class of ncRNAs antedating the origin of animal multicellularity and development.

Distal enhancers appear to be a metazoan innovation

Long-range *cis*-regulatory elements, so-called distal enhancers, have been hypothesized to be one of the key contributing factors underlying the spatial and temporal coordination of cell specification and differentiation in animal development (Levine, 2010; Levine et al., 2014; Levine and Tjian, 2003; Peter and Davidson, 2011; Woolfe et al., 2005). With the advent of functional genomics assays based on next-generation sequencing (NGS), such as ChIP-seq (Robertson et al., 2007), DNase-seq (Boyle et al., 2008) and ATAC-seq (Buenrostro et al., 2013), genome-wide maps of distal enhancers have been predicted in a range of human cell lines (ENCODE Project Consortium, 2012; Heintzman et al., 2009; Heintzman et al., 2007; Kundaje et al., 2015; Rada-Iglesias et al., 2011) and bilaterian model species (Bogdanović et al., 2012; Bonn et al., 2012; Creyghton et al., 2010; Negre et al., 2011; Roy et al., 2010; Shen et al., 2012; Visel et al., 2009).

Distal enhancers are rapidly evolving (Villar et al., 2015). In mammals, their origin often involves exaptation of ancestral DNA (Villar et al., 2015) or *de novo* emergence from neutral genomic background (proto-enhancers) (Emera et al., 2016). The finding of only a couple of conserved *cis*-regulatory elements, located in the vicinity of *SOX21* and *HMX3* genes, across eumetazoans (Maeso et al., 2013; Royo et al., 2011) underscores the dynamic nature of enhancer sequence evolution.

Distal enhancers with the same combination of bilaterian histone H3 PTMs [histone H3 lysine 4 monomethylation (H3K4me1) and histone H3 lysine 27 acetylation (H3K27ac)] have been identified in the cnidarian *Nematostella* (Schwaiger et al., 2014; Technau and Schwaiger, 2015). These, along with the transcriptional cofactor p300 binding sites, were distributed relative to developmental regulatory genes as in bilaterians (Schwaiger et al., 2014; Technau and Schwaiger, 2015) (Fig. 3). In contrast, the unicellular filasterean *Capsaspora owczarzaki* lacks animal developmental promoter types I and III (Lenhard et al., 2012) and distal enhancers, despite possessing many of the core TF-TF regulatory interactions found in animals (Sebé-Pedrós et al., 2016). The regulatory genome of *Capsaspora* appears, therefore, to be more similar to yeast (Bulger and Groudine, 2011), and suggests that distal enhancers are likely to have evolved later on the stem leading to the crown metazoans. Identification of putative distal enhancers in the sponge *Amphimedon* based on histone H3 PTM co-localization patterns is consistent with these elements evolving along the metazoan stem at the transition to multicellularity (Gaiti et al., unpublished). Interestingly, promoter DNA regulatory elements to allow for context- and cell type-specific gene expression also appeared to evolve in stem metazoans after diverging from the filasterean lineage (Fernandez-Valverde and Degnan, 2016), suggesting these are also a critical component of the animal *cis*-regulatory landscape (Fig. 3).

Perhaps not surprisingly, the emergence of distal enhancers prior to metazoan cladogenesis could then explain the pervasiveness of conserved syntenic blocks in animal genomes and the absence of these blocks in their unicellular relatives (Bulger and Groudine, 2011; Duan et al., 2010; Irimia et al., 2013; Irimia et al., 2012; Putnam et al., 2007; Sebé-Pedrós et al., 2016; Srivastava et al., 2010). Distal enhancers are enriched in metazoan microsyntenic blocks, which have been conserved over 700 million years of evolution, suggesting that their genomic co-location with syntenic genes is evolutionarily constrained and extremely ancient (Irimia et al., 2013; Irimia et al., 2012; Gaiti et al., unpublished). For instance, the Islet LIM homeobox gene (*Isl*), which plays conserved roles in animal development (Liang et al., 2011; Thor and Thomas, 1997), and *Scaper* (S-phase cyclin A-associated protein in the ER) form an ancient microsyntenic unit that possesses distal enhancers in their introns in sponges and bilaterians (Irimia et al., 2012; Gaiti et al., unpublished).

The mode of action of distal enhancers in non-bilaterians is currently unknown. Regulatory information is precisely organized in the three-dimensional chromatin and further compartmentalized into smaller topologically associated domains (TADs), restricting *cis*-regulatory sequences to their target genes (Denker and de Laat, 2016; Dixon et al., 2016). Interestingly, non-bilaterians lack CCCTC-binding factor (CTCF) (Fig. 1 and 3), a factor that establishes insulators and localizes with the more ancient cohesin (Heger et al., 2012). CTCF-binding sites within TADs facilitate enhancer–promoter long-range interactions through the recruitment of cohesin, that are jointly involved in higher-order chromatin structure in bilaterians (Lee and Iyer, 2012; Merkenschlager and Odom, 2013; Seitan et al., 2013; Vietri Rudan et al., 2015). Chromatin looping of distal enhancers to their target promoters in non-bilaterian animals might therefore occur through a CTCF-independent

cohesin binding mechanism (Schwaiger et al., 2014; Technau and Schwaiger, 2015; Gaiti et al., unpublished). Alternatively, but not exclusively, RNA polymerase II and its associated transcriptional machinery may track through the intervening DNA between enhancers and promoters, and may be the mechanism of enhancer-promoter interactions in non-bilaterian animals (Li et al., 2016). The co-occupancy of distal enhancers and RNA polymerase II in both sponges and cnidarians supports this mechanism of transcriptional activation (Schwaiger et al., 2014; Technau and Schwaiger, 2015; Gaiti et al., unpublished).

Combinatorial transcriptional regulation by the non-coding genome

The different non-coding RNAs and DNAs can interact to modulate gene expression during bilaterian development (De Kumar and Krumlauf, 2016; Jens and Rajewsky, 2015; Thomson and Dinger, 2016), but these interactions are currently uncharted in non-bilaterians.

lncRNA genes act through a multiplicity of functional and molecular mechanisms. These can either be mediated by the lncRNA transcripts themselves, or involve higher-order chromatin structure, the recruitment of regulatory complexes through RNA–protein interactions that influence the expression of nearby genes, or a combination thereof. For instance, changes in the lncRNA *Blustr* transcription and splicing may affect its neighboring gene (*Sfmbt2*) expression in part by altering chromatin state and RNA polymerase occupancy at the *Sfmbt2* promoter (Engreitz et al., 2016a). The lncRNA *Xist* interacts physically with multiple regions of the X chromosome, enhancing the spread of the lncRNA, which then recruits repressor proteins that promote X inactivation (Engreitz et al., 2016b; Engreitz et al., 2013). The lncRNA *Haunt* contains potential *cis* regulatory-elements that

induce neighboring *HoxA* genes, whereas the lncRNA transcript itself appears to repress *HoxA* expression (Yin et al., 2015). Likewise, the lncRNA *Bendr* regulates the expression of a neighboring gene via DNA regulatory elements in its proximal promoter region (Engreitz et al., 2016a). The lncRNA *CONCR* interacts with and modulates the activity of the helicase DDX11 to regulate sister chromatid cohesion (Marchese et al., 2016). The inducible lncRNA *DINO* binds to and stabilizes the well-known tumor suppressor protein p53 amplifying DNA damage response (Schmitt et al., 2016). The lncRNA *NORAD* maintains genomic stability by sequestering PUMILIO proteins, which repress the stability and translation of mRNAs to which they bind (Lee et al., 2016; Tichon et al., 2016).

lncRNAs, circRNAs and miRNAs might further synergize with each other to govern metazoan developmental processes. For instance, the lncRNA *H19* and insulin-like growth factor 2 (*IGF2*) are two well-described and studied imprinted genes, that are expressed from maternal and paternal alleles, respectively (Bartolomei et al., 1991; Brannan et al., 1990; DeChiara et al., 1990; DeChiara et al., 1991; Ferguson-Smith et al., 1991). In humans, *miR-675* is embedded in the lncRNA *H19*'s first exon. By controlling the release of *miR-675*, *H19* suppresses cell proliferation in response to cellular stress during embryogenesis (Keniry et al., 2012). The process of small RNA regulation by lncRNA precursors also extends to snoRNAs (Yin et al., 2012) and piRNAs (Brennecke et al., 2007), and may well exist as a common checkpoint for small RNA production. lncRNAs have been also suggested to interfere with miRNA-mediated mRNA destabilization and, rather than competing for miRNA-binding sites within mRNAs, lncRNAs can compete for the miRNAs themselves (Jalali et al., 2013; Paraskevopoulou et al., 2013). This is illustrated by lncRNA *PTENpg1 asRNA* β that, by forming a duplex with the *PTEN* pseudogene (*PTENpg1*), acts as a decoy for *PTEN*-related miRNAs and, thus, prevents them from promoting the degradation of *PTEN* mRNA

(Johnsson et al., 2013). *lncRNA-RoR* binds miRNAs resulting in the upregulation of principal TFs involved in embryonic stem cell maintenance and differentiation (Loewer et al., 2010; Wang et al., 2013). A similar mechanism has been observed for circRNAs, where the circRNA *CDR1as* has two functions: (i) as a trans-regulator of *CDR1* mRNA through a mechanism that is regulated by AGO2- and *miR-671*-mediated cleavage (Hansen et al., 2011); and (ii) as a miRNA sink for *miR-7*, thus regulating neural gene expression (Memczak et al., 2013).

The dynamic nature of lncRNAs, as well as the RNA plasticity in interacting with diverse molecules such as DNA, RNA and protein, renders lncRNA an ideal mediator to regulate local and sequence-specific DNA methylation (Bao et al., 2015; Chalei et al., 2014; Di Ruscio et al., 2013). An interesting example resides in the above-mentioned *H19-IGF2* locus. On the maternal chromosome, an unmethylated imprinting controlled region binds CTCF and forms an insulator that prevents access of the *IGF2* promoter to downstream enhancers, which, in turn, allow the expression of the nearby lncRNA *H19* promoter. Contrarily, on the paternal chromosome, DNA methylation of the same imprinting controlled region prevents binding of CTCF. This maintains the lncRNA *H19* silenced, allowing the downstream enhancers to activate the *IGF2* promoter (Barlow and Bartolomei, 2014; Gabory et al., 2010).

The understanding of *how* and *when* such functional interactions evolved and whether they are deeply conserved features of the metazoan genome will require analyses of non-bilaterians, in which the interconnections between protein-coding genes, transcriptional co-regulators, chromatin modifying complexes, and the various types of regulatory DNA and RNA, have not yet been functionally investigated.

Concluding remarks

In summary, the non-coding portion of the genome appears to house much of the regulatory information underpinning cell-type and developmental gene expression. In addition to an expansion of transcription factor and lncRNA repertoires before the emergence of crown metazoans, there appears to have been fundamental changes in the non-coding regulatory architecture of the genome, including the origin and/or expansion of miRNAs, piRNAs, distal enhancers and promoter types for cell-type-specificity and developmental regulation. These innovations appear to have contributed an increase in the capacity to regulate spatial and temporal gene expression, which is necessary for complex multicellularity. The further expansion of these systems in eumetazoans and bilaterians is consistent with an evolutionary scenario in which *quantitative* rather than *qualitative* differences in regulatory mechanisms underpin the evolution and diversification of eumetazoan body plans (Fig. 4).

Disclosure

The authors report no conflicts of interest in this work.

Acknowledgments

We thank Degnan lab members for constructive discussions, and William Hatleberg and Gemma Richards for critical reading of the manuscript. We apologize to colleagues who have produced primary research on the topic but could not be cited due to space limitations. This work was supported by an Australian Research Council grant to BMD.

References

Amaral, P.P., Leonardi, T., Han, N., Vire, E., Gascoigne, D.K., Arias-Carrasco, R., Buscher, M., Zhang, A., Pluchino, S., Maracaja-Coutinho, V.*, et al.* (2016). Genomic positional conservation identifies topological anchor point (tap)RNAs linked to developmental loci. bioRxiv.

Anderson, K.M., Anderson, D.M., McAnally, J.R., Shelton, J.M., Bassel-Duby, R., and Olson, E.N. (2016). Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. Nature *advance online publication*.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. Molecular Cell *31*, 785-799.

Ashe, A., Sapetschnig, A., Weick, E.-M., Mitchell, J., Bagijn, Marloes P., Cording, Amy C., Doebley, A.-L., Goldstein, Leonard D., Lehrbach, Nicolas J., Le Pen, J.*, et al.* (2012). piRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans*. Cell *150*, 88-99.

Avesson, L., Reimegård, J., Wagner, E.G.H., and Söderbom, F. (2012). MicroRNAs in amoebozoa: Deep sequencing of the small RNA population in the social amoeba *Dictyostelium discoideum* reveals developmentally regulated microRNAs. RNA *18*, 1771-1782.

Bao, X., Wu, H., Zhu, X., Guo, X., Hutchins, A.P., Luo, Z., Song, H., Chen, Y., Lai, K., Yin, M.*, et al.* (2015). The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters. Cell Research *25*, 80-92.

Barlow, D.P., and Bartolomei, M.S. (2014). Genomic imprinting in mammals. Cold Spring Harbor Perspectives in Biology *6*, pii: a018382.

Barrett, S.P., and Salzman, J. (2016). Circular RNAs: Analysis, expression and potential functions. Development *143*, 1838-1847.

Bartel, D.P. (2009). MicroRNAs: Target recognition and regulatory functions. Cell *136*, 215-233.

Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991). Parental imprinting of the mouse H19 gene. Nature *351*, 153-155.

Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proceedings of the National Academy of Sciences USA *74*, 3171-3175.

Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S*., et al.* (2004). Global identification of human transcribed sequences with genome tiling arrays. Science *306*, 2242-2246.

Bogdanović, O., Fernandez-Miñán, A., Tena, J.J., de la Calle-Mustienes, E., Hidalgo, C., van Kruysbergen, I., van Heeringen, S.J., Veenstra, G.J.C., and Gómez-Skarmeta, J.L. (2012). Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Research *22*, 2043-2053.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczynski, B., Riddell, A., and Furlong, E.E.M. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nature Genetics *44*, 148-156.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311-322.

Brannan, C.I., Dees, E.C., Ingram, R.S., and Tilghman, S.M. (1990). The product of the H19 gene may function as an RNA. Molecular and Cellular Biology *10*, 28-36.

Bråte, J., Adamski, M., Neumann, R.S., Shalchian-Tabrizi, K., and Adamska, M. (2015). Regulatory RNA at the root of animals: Dynamic expression of developmental lincRNAs in the calcisponge *Sycon ciliatum*. Proceedings of the Royal Society of London B: Biological Sciences *282*, 20151746.

Bråte, J., Neumann, R.S., Fromm, B., Haraldsen, A.A.B., Grini, P., and Shalchian-Tabrizi, K. (2016). Pre-metazoan origin of animal miRNAs. bioRxiv.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. Cell *128*, 1089-1103.

Britten, R.J., and Davidson, E.H. (1969). Gene regulation for higher cells: A theory. Science *165*, 349-357.

Britten, R.J., and Davidson, E.H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. The Quarterly Review of Biology *46*, 111-138.

Brown, J.B., Boley, N., Eisman, R., May, G.E., Stoiber, M.H., Duff, M.O., Booth, B.W., Wen, J., Park, S., Suzuki, A.M*., et al.* (2014). Diversity and dynamics of the *Drosophila* transcriptome. Nature *512*, 393-399.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods *10*, 1213-1218.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327-339.

Cabili, M.N., Dunagin, M.C., McClanahan, P.D., Biaesch, A., Padovan-Merhar, O., Regev, A., Rinn, J.L., and Raj, A. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. Genome Biology *16*, 1-16.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes & Development *25*, 1915-1927.

Calcino, A. (2015). RNA interference in early branching metazoans (School of Biological Sciences, The University of Queensland).

Carmell, M.A., Girard, A., van de Kant, H.J.G., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 Is essential for spermatogenesis and repression of transposons in the mouse male germline. Developmental Cell *12*, 503-514.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C.*, et al.* (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559-1563.

Cerutti, H., and Casas-Mollano, J.A. (2006). On the origin and functions of RNA-mediated silencing: From protists to man. Current Genetics *50*, 81-99.

Chalei, V., Sansom, S.N., Kong, L., Lee, S., Montiel, J.F., Vance, K.W., and Ponting, C.P. (2014). The long non-coding RNA *Dali* is an epigenetic regulator of neural differentiation. eLife *3*, e04530.

Chen, B., Zhang, Y., Zhang, X., Jia, S., Chen, S., and Kang, L. (2016a). Genome-wide identification and developmental expression profiling of long noncoding RNAs during *Drosophila* metamorphosis. Scientific Reports *6*, 23330.

Chen, J., Shishkin, A.A., Zhu, X., Kadri, S., Maza, I., Guttman, M., Hanna, J.H., Regev, A., and Garber, M. (2016b). Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. Genome Biology *17*, 1-17.

Chen, L.-L. (2016). The biogenesis and emerging roles of circular RNAs. Nature Reviews Molecular Cell Biology *17*, 205-211.

Chodroff, R.A., Goodstadt, L., Sirey, T.M., Oliver, P.L., Davies, K.E., Green, E.D., Molnar, Z., and Ponting, C.P. (2010). Long noncoding RNA genes: Conservation of sequence and brain expression among diverse amniotes. Genome Biology *11*, R72.

Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell *12*, 1-8.

Cock, J.M., Sterck, L., Rouze, P., Scornet, D., Allen, A.E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J.-M., Badger, J.H.*, et al.* (2010). The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature *465*, 617-621.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A.*, et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences USA *107*, 21931-21936.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561-563.

Czech, B., and Hannon, G.J. (2016). One loop to rule them all: The ping-pong cycle and piRNA-guided silencing. Trends in Biochemical Sciences *41*, 324-337.

Davidson, E.H., and Erwin, D.H. (2006). Gene regulatory networks and the evolution of animal body plans. Science *311*, 796-800.

Davidson, E.H., Klein, W.H., and Britten, R.J. (1977). Sequence organization in animal DNA and a speculation on hnRNA as a coordinate regulatory transcript. Developmental Biology *55*, 69-84.

Davidson, E.H., and Peter, I.S. (2015). Genomic Control Process (Oxford: Academic Press).

De Kumar, B., and Krumlauf, R. (2016). HOXs and lincRNAs: Two sides of the same coin. Science Advances *2*, e1501402.

de Mendoza, A., Sebé-Pedrós, A., Šestak, M.S., Matejčić, M., Torruella, G., Domazet-Lošo, T., and Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proceedings of the National Academy of Sciences *110*, E4858-E4866.

de Mendoza, A., Suga, H., Permanyer, J., Irimia, M., and Ruiz-Trillo, I. (2015). Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. eLife, e08904.

DeChiara, T.M., Efstratiadis, A., and Robertsen, E.J. (1990). A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. Nature *345*, 78-80.

DeChiara, T.M., Robertson, E.J., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. Cell *64*, 849-859.

Degnan, B.M., Vervoort, M., Larroux, C., and Richards, G.S. (2009). Early evolution of metazoan transcription factors. Current Opinion in Genetics & Development *19*, 591-599.

Denker, A., and de Laat, W. (2016). The second decade of 3C technologies: Detailed insights into nuclear organization. Genes & Development *30*, 1357-1382.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G.*, et al.* (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Research *22*, 1775-1789.

Di Ruscio, A., Ebralidze, A.K., Benoukraf, T., Amabile, G., Goff, L.A., Terragni, J., Figueroa, M.E., De Figueiredo Pontes, L.L., Alberich-Jorda, M., Zhang, P*., et al.* (2013). DNMT1-interacting RNAs block gene-specific DNA methylation. Nature *503*, 371-376.

Dixon, Jesse R., Gorkin, David U., and Ren, B. (2016). Chromatin domains: The unit of chromosome organization. Molecular Cell *62*, 668-680.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F*., et al.* (2012). Landscape of transcription in human cells. Nature *489*, 101-108.

Drinnenberg, I.A., Weinberg, D.E., Xie, K.T., Mower, J.P., Wolfe, K.H., Fink, G.R., and Bartel, D.P. (2009). RNAi in budding yeast. Science *326*, 544-550.

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S., Blau, C.A., and Noble, W.S. (2010). A three-dimensional model of the yeast genome. Nature *465*, 363-367.

Emera, D., Yin, J., Reilly, S.K., Gockley, J., and Noonan, J.P. (2016). Origin and evolution of developmental enhancers in the mammalian neocortex. Proceedings of the National Academy of Sciences USA *113*, E2617-E2626.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016a). Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature *advance online publication*.

Engreitz, J.M., Ollikainen, N., and Guttman, M. (2016b). Long non-coding RNAs: Spatial amplifiers that control nuclear structure and gene expression. Nature Reviews Molecular Cell Biology *advance online publication*.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S*., et al.* (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science *341*, 1237973.

Erwin, D.H., and Davidson, E.H. (2009). The evolution of hierarchical gene regulatory networks. Nature Reviews Genetics *10*, 141-148.

Erwin, D.H., Laflamme, M., Tweedt, S.M., Sperling, E.A., Pisani, D., and Peterson, K.J. (2011). The Cambrian conundrum: Early divergence and later ecological success in the early history of animals. Science *334*, 1091-1097.

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: New players in cell differentiation and development. Nature Reviews Genetics *15*, 7-21.

Ferguson-Smith, A.C., Cattanach, B.M., Barton, S.C., Beechey, C.V., and Surani, M.A. (1991). Embryological and molecular investigations of parental imprinting on mouse chromosome 7. Nature *351*, 667-670.

Fernandez-Valverde, S.L., and Degnan, B.M. (2016). Bilaterian-like promoters in the highly compact *Amphimedon queenslandica* genome. Scientific Reports *6*, 22496.

Finnerty, J.R., and Martindale, M.Q. (1999). Ancient origins of axial patterning genes: Hox genes and ParaHox genes in the Cnidaria. Evolution & Development *1*, 16-23.

Finnerty, J.R., Paulson, D., Burton, P., Pang, K., and Martindale, M.Q. (2003). Early evolution of a homeobox gene: The parahox gene Gsx in the Cnidaria and the Bilateria. Evolution & Development *5*, 331-345.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature *391*, 806-811.

Forouzmand, E., Owens, N.D.L., Blitz, I.L., Paraiso, K.D., Khokha, M.K., Gilchrist, M.J., Xie, X., and Cho, K.W.Y. (2016). Developmentally regulated long non-coding RNAs in *Xenopus tropicalis*. Developmental Biology *pii: S0012-1606*, 30120-30128.

Gabory, A., Jammes, H., and Dandolo, L. (2010). The H19 locus: Role of an imprinted non-coding RNA in growth and development. BioEssays *32*, 473-480.

Gaiti, F., Fernandez-Valverde, S.L., Nakanishi, N., Calcino, A.D., Yanai, I., Tanurdzic, M., and Degnan, B.M. (2015). Dynamic and widespread lncRNA expression in a sponge and the origin of animal complexity. Molecular Biology and Evolution *32*, 2367-2382.

Goff, L.A., Groff, A.F., Sauvageau, M., Trayes-Gibson, Z., Sanchez-Gomez, D.B., Morse, M., Martin, R.D., Elcavage, L.E., Liapis, S.C., Gonzalez-Celeiro, M.*, et al.* (2015). Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the mouse brain. Proceedings of the National Academy of Sciences USA *112*, 6855-6862.

Grant, J., Mahadevaiah, S.K., Khil, P., Sangrithi, M.N., Royo, H., Duckworth, J., McCarrey, J.R., VandeBerg, J.L., Renfree, M.B., Taylor, W.*, et al.* (2012). Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. Nature *487*, 254-258.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193-1197.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. Science *315*, 1587-1590.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P.*, et al.* (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223-227.

Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S., and Lander, E.S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell *154*, 240-251.

Hacisuleyman, E., Goff, L.A., Trapnell, C., Williams, A., Henao-Mejia, J., Sun, L., McClanahan, P., Hendrickson, D.G., Sauvageau, M., Kelley, D.R.*, et al.* (2014). Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nature Structural & Molecular Biology *21*, 198-206.

Hacisuleyman, E., Shukla, C.J., Weiner, C.L., and Rinn, J.L. (2016). Function and evolution of local repeats in the Firre locus. Nature Communications *7*, 11021.

Hansen, T.B., Wiklund, E.D., Bramsen, J.B., Villadsen, S.B., Statham, A.L., Clark, S.J., and Kjems, J. (2011). miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. The EMBO Journal *30*, 4414-4422.

Hawkes, Emily J., Hennelly, Scott P., Novikova, Irina V., Irwin, Judith A., Dean, C., and Sanbonmatsu, Karissa Y. (2016). *COOLAIR* antisense RNAs form evolutionarily conserved elaborate secondary structures. Cell Reports *16*, 3087-3096.

Heard, E., Mongelard, F., Arnaud, D., Chureau, C., Vourc'h, C., and Avner, P. (1999). Human XIST yeast artificial chromosome transgenes show partial X inactivation center function in mouse embryonic stem cells. Proceedings of the National Academy of Sciences USA *96*, 6841-6846.

Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., and Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. Proceedings of the National Academy of Sciences USA *109*, 17507-17512.

Heger, P., Marin, B., and Schierenberg, E. (2009). Loss of the insulator protein CTCF during nematode evolution. BMC Molecular Biology *10*, 1-14.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W.*, et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genetics *39*, 311-318.

Hezroni, H., Koppstein, D., Schwartz, Matthew G., Avrutin, A., Bartel, David P., and Ulitsky, I. (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. Cell Reports *11*, 1110-1122.

Housman, G., and Ulitsky, I. (2016). Methods for distinguishing between protein-coding and long noncoding RNAs and the elusive biological purpose of translation of long noncoding RNAs. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1859*, 31-40.

Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C., Wang, P.*, et al.* (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nature Genetics *43*, 621-629.

Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B., and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. BMC Genomics *8*, 1-16.

Ingolia, Nicholas T., Brar, Gloria A., Stern-Ginossar, N., Harris, Michael S., Talhouarne, Gaëlle J.S., Jackson, Sarah E., Wills, Mark R., and Weissman, Jonathan S. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. Cell Reports *8*, 1365-1379.

Irimia, M., Maeso, I., Roy, S.W., and Fraser, H.B. (2013). Ancient *cis*-regulatory constraints and the evolution of genome architecture. Trends in Genetics *29*, 521-528.

Irimia, M., Tena, J.J., Alexis, M.S., Fernandez-Miñan, A., Maeso, I., Bogdanović, O., de la Calle-Mustienes, E., Roy, S.W., Gómez-Skarmeta, J.L., and Fraser, H.B. (2012). Extensive conservation of ancient microsynteny across metazoans due to *cis*-regulatory constraints. Genome Research *22*, 2356-2367.

Ivey, K.N., and Srivastava, D. (2010). MicroRNAs as regulators of differentiation and cell fate decisions. Cell Stem Cell *7*, 36-41.

Jalali, S., Bhartiya, D., Lalwani, M.K., Sivasubbu, S., and Scaria, V. (2013). Systematic transcriptome wide analysis of lncRNA-miRNA interactions. PLOS ONE *8*, e53823.

Jayakodi, M., Jung, J.W., Park, D., Ahn, Y.-J., Lee, S.-C., Shin, S.-Y., Shin, C., Yang, T.-J., and Kwon, H.W. (2015). Genome-wide characterization of long intergenic non-coding RNAs (lincRNAs) provides new insight into viral diseases in honey bees *Apis cerana* and *Apis mellifera*. BMC Genomics *16*, 1-12.

Jenkins, A.M., Waterhouse, R.M., and Muskavitch, M.A.T. (2015). Long non-coding RNA discovery across the genus anopheles reveals conserved secondary structures within and beyond the Gambiae complex. BMC Genomics *16*, 1-14.

Jens, M., and Rajewsky, N. (2015). Competition between target sites of regulators shapes post-transcriptional gene regulation. Nature Reviews Genetics *16*, 113-126.

Johnsson, P., Ackley, A., Vidarsdottir, L., Lui, W.O., Corcoran, M., Grander, D., and Morris, K.V. (2013). A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. Nature Structural & Molecular Biology *20*, 440-446.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L*., et al.* (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484-1488.

Kapusta, A., and Feschotte, C. (2014). Volatile evolution of long noncoding RNA repertoires: Mechanisms and biological implications. Trends in Genetics *30*, 439-452.

Keniry, A., Oxley, D., Monnier, P., Kyba, M., Dandolo, L., Smits, G., and Reik, W. (2012). The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. Nature Cell Biology *14*, 659-665.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A*., et al.* (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proceedings of the National Academy of Sciences USA *106*, 11667-11672.

King, N. (2004). The unicellular ancestry of animal development. Developmental Cell *7*, 313-325.

King, N., Hittinger, C.T., and Carroll, S.B. (2003). Evolution of key cell signaling and adhesion protein families predates animal origins. Science *301*, 361-363.

Klenov, M.S., Sokolova, O.A., Yakushev, E.Y., Stolyarenko, A.D., Mikhaleva, E.A., Lavrov, S.A., and Gvozdev, V.A. (2011). Separation of stem cell maintenance and transposon silencing functions of Piwi protein. Proceedings of the National Academy of Sciences USA *108*, 18760-18765.

Kosik, K.S. (2010). MicroRNAs and cellular phenotypy. Cell *143*, 21-26.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J*., et al.* (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317-330.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W*., et al.* (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes & Development *22*, 908-917.

Kutter, C., Watt, S., Stefflova, K., Wilson, M.D., Goncalves, A., Ponting, C.P., Odom, D.T., and Marques, A.C. (2012). Rapid turnover of long noncoding RNAs and the evolution of gene expression. PLOS Genetics *8*, e1002841.

Larroux, C., Luke, G.N., Koopman, P., Rokhsar, D.S., Shimeld, S.M., and Degnan, B.M. (2008). Genesis and expansion of metazoan transcription factor gene classes. Molecular Biology and Evolution *25*, 980-996.

Lee, B.-K., and Iyer, V.R. (2012). Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation. Journal of Biological Chemistry *287*, 30906-30913.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell *75*, 843-854.

Lee, S., Kopp, F., Chang, T.-C., Sataluri, A., Chen, B., Sivakumar, S., Yu, H., Xie, Y., and Mendell, Joshua T. (2016). Noncoding RNA *NORAD* regulates genomic stability by sequestering PUMILIO proteins. Cell *164*, 69-80.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: Emerging characteristics and insights into transcriptional regulation. Nature Reviews Genetics *13*, 233-245.

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Current Biology *20*, R754-R763.

Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: Transcription enters a new era. Cell *157*, 13-25.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147-151.

Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: Recent insights and future perspectives. Nature Reviews Genetics *17*, 207-223.

Liang, X., Song, M.-R., Xu, Z., Lanuza, G.M., Liu, Y., Zhuang, T., Chen, Y., Pfaff, S.L., Evans, S.M., and Sun, Y. (2011). Isl1 is required for multiple aspects of motor neuron development. Molecular and Cellular Neuroscience *47*, 215-222.

Liew, Y.J., Aranda, M., Carr, A., Baumgarten, S., Zoccola, D., Tambutté, S., Allemand, D., Micklem, G., and Voolstra, C.R. (2014). Identification of microRNAs in the coral *Stylophora pistillata*. PLOS ONE *9*, e91101.

Liew, Y.J., Ryu, T., Aranda, M., and Ravasi, T. (2016). miRNA repertoires of demosponges *Stylissa carteri* and *Xestospongia testudinaria*. PLOS ONE *11*, e0149080.

Liu, S.J., Nowakowski, T.J., Pollen, A.A., Lui, J.H., Horlbeck, M.A., Attenello, F.J., He, D., Weissman, J.S., Kriegstein, A.R., Diaz, A.A.*, et al.* (2016). Single-cell analysis of long non-coding RNAs in the developing human neocortex. Genome Biology *17*, 1-17.

Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.-H., Thomas, K., Park, I.H., Garber, M., Curran, M., Onder, T., Agarwal, S.*, et al.* (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nature Genetics *42*, 1113-1117.

Luteijn, M.J., van Bergeijk, P., Kaaij, L.J.T., Almeida, M.V., Roovers, E.F., Berezikov, E., and Ketting, R.F. (2012). Extremely stable Piwi-induced gene silencing in *Caenorhabditis elegans*. The EMBO Journal *31*, 3422-3430.

Lyle, R., Watanabe, D., Vruchte, D.t., Lerchner, W., Smrzka, O.W., Wutz, A., Schageman, J., Hahner, L., Davies, C., and Barlow, D.P. (2000). The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. Nature Genetics *25*, 19-21.

Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., Sanglier-Cianferani, S., Van Dorsselaer, A., Clerc, P., Avner, P.*, et al.* (2010). 2-D structure of the A region of Xist RNA and its implication for PRC2 association. PLOS Biology *8*, e1000276.

Maeso, I., Irimia, M., Tena, J.J., Casares, F., and Gómez-Skarmeta, J.L. (2013). Deep conservation of *cis*-regulatory elements in metazoans. Philosophical Transactions of the Royal Society B: Biological Sciences *368*, 20130020.

Marchese, Francesco P., Grossi, E., Marín-Béjar, O., Bharti, Sanjay K., Raimondi, I., González, J., Martínez-Herrera, Dannys J., Athie, A., Amadoz, A., Brosh, Robert M., Jr.*, et al.* (2016). A long noncoding RNA regulates sister chromatid cohesion. Molecular Cell *63*, 397-407.

Marques, A., Hughes, J., Graham, B., Kowalczyk, M., Higgs, D., and Ponting, C. (2013). Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. Genome Biology *14*, R131.

Mattick, J.S. (2004). RNA regulation: A new genetics? Nature Reviews Genetics *5*, 316-323.

Mauri, M., Kirchner, M., Aharoni, R., Ciolli Mattioli, C., van den Bruck, D., Gutkovitch, N., Modepalli, V., Selbach, M., Moran, Y., and Chekulaeva, M. (2016). Conservation of miRNA-mediated silencing mechanisms across 600 million years of animal evolution. Nucleic Acids Research, pii: gkw792.

Maxwell, E.K., Ryan, J.F., Schnitzler, C.E., Browne, W.E., and Baxevanis, A.D. (2012). MicroRNAs and essential components of the microRNA processing machinery are not encoded in the genome of the ctenophore *Mnemiopsis leidyi*. BMC Genomics *13*, 1-11.

McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A., and Gehring, W.J. (1984). A homologous protein-coding sequence in drosophila homeotic genes and its conservation in other metazoans. Cell *37*, 403-408.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M.*, et al.* (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature *495*, 333-338.

Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F., and Mattick, J.S. (2008). Specific expression of long noncoding RNAs in the mouse brain. Proceedings of the National Academy of Sciences USA *105*, 716-721.

Merkenschlager, M., and Odom, Duncan T. (2013). CTCF and cohesin: Linking gene regulatory elements with their targets. Cell *152*, 1285-1297.

Migeon, B.R., Kazi, E., Haisley-Royster, C., Hu, J., Reeves, R., Call, L., Lawler, A., Moore, C.S., Morrison, H., and Jeppesen, P. (1999). Human X inactivation center induces random X chromosome inactivation in male transgenic mice. Genomics *59*, 113-121.

Molnar, A., Schwach, F., Studholme, D.J., Thuenemann, E.C., and Baulcombe, D.C. (2007). miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. Nature *447*, 1126-1129.

Moran, Y., Fredman, D., Praher, D., Li, X.Z., Wee, L.M., Rentzsch, F., Zamore, P.D., Technau, U., and Seitz, H. (2014). Cnidarian microRNAs frequently regulate targets by cleavage. Genome Research *24*, 651-663.

Moran, Y., Praher, D., Fredman, D., and Technau, U. (2013). The evolution of microRNA pathway protein components in Cnidaria. Molecular Biology and Evolution *30*, 2541-2552.

Moroz, L.L., Kocot, K.M., Citarella, M.R., Dosung, S., Norekian, T.P., Povolotskaya, I.S., Grigorenko, A.P., Dailey, C., Berezikov, E., Buckley, K.M.*, et al.* (2014). The ctenophore genome and the evolutionary origins of neural systems. Nature *510*, 109-114.

Morris, K.V., and Mattick, J.S. (2014). The rise of regulatory RNA. Nature Reviews Genetics *15*, 423-437.

Mu, C., Wang, R., Li, T., Li, Y., Tian, M., Jiao, W., Huang, X., Zhang, L., Hu, X., Wang, S.*, et al.* (2016). Long non-coding RNAs (lncRNAs) of sea cucumber: Large-scale prediction, expression profiling, non-coding network construction, and lncRNA-microRNA-gene interaction analysis of lncRNAs in *Apostichopus japonicus* and *Holothuria glaberrima* during LPS challenge and radial organ complex regeneration. Marine Biotechnology *18*, 485-499.

Nam, J.W., and Bartel, D.P. (2012). Long noncoding RNAs in *C. elegans*. Genome Research *22*, 2529-2540.

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker, J.C., Grutzner, F., and Kaessmann, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature *505*, 635-640.

Negre, N., Brown, C.D., Ma, L., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R.*, et al.* (2011). A *cis*-regulatory map of the *Drosophila* genome. Nature *471*, 527-531.

Novikova, I.V., Hennelly, S.P., and Sanbonmatsu, K.Y. (2012). Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. Nucleic Acids Research *40*, 5034-5051.

Obbard, D.J., Gordon, K.H.J., Buck, A.H., and Jiggins, F.M. (2009). The evolution of RNAi as a defence against viruses and transposable elements. Philosophical Transactions of the Royal Society B: Biological Sciences *364*, 99.

Okazaki, Y., and al., e. (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature *420*, 563-573.

Pang, K.C., Frith, M.C., and Mattick, J.S. (2006). Rapid evolution of noncoding RNAs: Lack of conservation does not mean lack of function. Trends in Genetics *22*, 1-5.

Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, N., Reczko, M., Maragkakis, M., Dalamagas, T.M., and Hatzigeorgiou, A.G. (2013). DIANA-LncBase: Experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. Nucleic Acids Research *41*, D239-D245.

Pauli, A., Rinn, J.L., and Schier, A.F. (2011). Non-coding RNAs as regulators of embryogenesis. Nature Reviews Genetics *12*, 136-149.

Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A.*, et al.* (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genome Research *22*, 577-591.

Perry, R.B.-T., and Ulitsky, I. (2016). The functions of long noncoding RNAs in development and stem cells. Development *143*, 3882-3894.

Peter, Isabelle S., and Davidson, Eric H. (2011). Evolution of gene regulatory networks controlling body plan development. Cell *144*, 970-985.

Peterson, K.J., Dietrich, M.R., and McPeek, M.A. (2009). MicroRNAs and metazoan macroevolution: Insights into canalization, complexity, and the Cambrian explosion. BioEssays *31*, 736-747.

Pombo, A., and Dillon, N. (2015). Three-dimensional genome architecture: Players and mechanisms. Nature Reviews Molecular Cell Biology *16*, 245-257.

Ponjavic, J., Oliver, P.L., Lunter, G., and Ponting, C.P. (2009). Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. PLOS Genetics *5*, e1000617.

Ponting, C.P., Oliver, P.L., and Reik, W. (2009). Evolution and functions of long noncoding RNAs. Cell *136*, 629-641.

Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V.V.*, et al.* (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science *317*, 86-94.

Quinn, J.J., and Chang, H.Y. (2016). Unique features of long non-coding RNA biogenesis and function. Nature Reviews Genetics *17*, 47-62.

Quinn, J.J., Zhang, Q.C., Georgiev, P., Ilik, I.A., Akhtar, A., and Chang, H.Y. (2016). Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. Genes & Development *30*, 191-207.

Quiring, R., Walldorf, U., Kloter, U., and Gehring, W.J. (1994). Homology of the eyeless gene of *Drosophila* to the Small eye gene in mice and Aniridia in humans. Science *265*, 785-789.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Ravasi, T., Suzuki, H., Pang, K.C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M.C., Gongora, M.M.*, et al.* (2006). Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Research *16*, 11-19.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. Nature *403*, 901-906.

Richards, G.S., and Degnan, B.M. (2009). The dawn of developmental signaling in the Metazoa. Cold Spring Harbor Symposia on Quantitative Biology *74*, 81-90.

Richter, D.J., and King, N. (2013). The genomic and cellular foundations of animal origins. Annual Review of Genetics *47*, 509-537.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A.*, et al.* (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nature Methods *4*, 651-657.

Robinson, J.M., Sperling, E.A., Bergum, B., Adamski, M., Nichols, S.A., Adamska, M., and Peterson, K.J. (2013). The Identification of microRNAs in calcisponges: Independent evolution of microRNAs in basal metazoans. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution *320*, 84-93.

Rokas, A. (2008). The origins of multicellularity and the early history of the genetic toolkit for animal development. Annual Review of Genetics *42*, 235-251.

Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F.*, et al.* (2010). Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. Science *330*, 1787-1797.

Royo, J.L., Maeso, I., Irimia, M., Gao, F., Peter, I.S., Lopes, C.S., D'Aniello, S., Casares, F., Davidson, E.H., Garcia-Fernández, J.*, et al.* (2011). Transphyletic conservation of developmental regulatory state in animal evolution. Proceedings of the National Academy of Sciences USA *108*, 14186-14191.

Rutenberg-Schoenberg, M., Sexton, A.N., and Simon, M.D. (2016). The properties of long noncoding RNAs that regulate chromatin. Annual Review of Genomics and Human Genetics *17*, 69-94.

Salzman, J. (2016). Circular RNA expression: Its potential regulation and function. Trends in Genetics *32*, 309-316.

Sanbonmatsu, K.Y. (2016). Towards structural classification of long non-coding RNAs. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1859*, 41-45.

Sauvageau, M., Goff, L.A., Lodato, S., Bonev, B., Groff, A.F., Gerhardinger, C., Sanchez-Gomez, D.B., Hacisuleyman, E., Li, E., Spence, M.*, et al.* (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. eLife *2*, e01749.

Schmitt, A.M., Garcia, J.T., Hung, T., Flynn, R.A., Shen, Y., Qu, K., Payumo, A.Y., Peres-da-Silva, A., Broz, D.K., Baum, R.*, et al.* (2016). An inducible long noncoding RNA amplifies DNA damage signaling. Nature Genetics *48*, 1370-1376.

Schwaiger, M., Schonauer, A., Rendeiro, A.F., Pribitzer, C., Schauer, A., Gilles, A.F., Schinko, J.B., Renfer, E., Fredman, D., and Technau, U. (2014). Evolutionary conservation of the eumetazoan gene regulatory landscape. Genome Research *24*, 639-650.

Sebé-Pedrós, A., Ballaré, C., Parra-Acero, H., Chiva, C., Tena, Juan J., Sabidó, E., Gómez-Skarmeta, José L., Di Croce, L., and Ruiz-Trillo, I. (2016). The dynamic regulatory genome of *Capsaspora* and the origin of animal multicellularity. Cell *165*, 1224-1237.

Sebé-Pedrós, A., and de Mendoza, A. (2015). Transcription factors and the origin of animal multicellularity. In Evolutionary Transitions to Multicellular Life: Principles and mechanisms (Dordrecht: Springer Netherlands), pp. 379-394.

Sebé-Pedrós, A., de Mendoza, A., Lang, B.F., Degnan, B.M., and Ruiz-Trillo, I. (2011). Unexpected repertoire of metazoan transcription factors in the unicellular holozoan *Capsaspora owczarzaki*. Molecular Biology and Evolution *28*, 1241-1254.

Sebé-Pedrós, A., Roger, A.J., Lang, F.B., King, N., and Ruiz-Trillo, I. (2010). Ancient origin of the integrin-mediated adhesion and signaling machinery. Proceedings of the National Academy of Sciences USA *107*, 10142-10147.

Sebé-Pedrós, A., Zheng, Y., Ruiz-Trillo, I., and Pan, D. (2012). Premetazoan origin of the Hippo signaling pathway. Cell Reports *1*, 13-20.

Seitan, V.C., Faure, A.J., Zhan, Y., McCord, R.P., Lajoie, B.R., Ing-Simmons, E., Lenhard, B., Giorgetti, L., Heard, E., Fisher, A.G*., et al.* (2013). Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Research *23*, 2066-2077.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V*., et al.* (2012). A map of the *cis*-regulatory sequences in the mouse genome. Nature *488*, 116-120.

Shenoy, A., and Blelloch, R.H. (2014). Regulation of microRNA function in somatic stem cell proliferation and differentiation. Nature Reviews Molecular Cell Biology *15*, 565-576.

Shirayama, M., Seth, M., Lee, H.-C., Gu, W., Ishidate, T., Conte, D., Jr., and Mello, Craig C. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. Cell *150*, 65-77.

Shpiz, S., Olovnikov, I., Sergeeva, A., Lavrov, S., Abramov, Y., Savitsky, M., and Kalmykova, A. (2011). Mechanism of the piRNA-mediated silencing of *Drosophila* telomeric retrotransposons. Nucleic Acids Research *39*, 8703-8711.

Sienski, G., Dönertas, D., and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and Maelstrom and its impact on chromatin state and gene expression. Cell *151*, 964-980.

Slack, J.M.W., Holland, P.W.H., and Graham, C.F. (1993). The zootype and the phylotypic stage. Nature *361*, 490-492.

Sperling, E.A., Robinson, J.M., Pisani, D., and Peterson, K.J. (2010). Where's the glass? Biomarkers, molecular clocks, and microRNAs suggest a 200-Myr missing Precambrian fossil record of siliceous sponge spicules. Geobiology *8*, 24-36.

Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U*., et al.* (2010). The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature *466*, 720-726.

Svoboda, P. (2014). Renaissance of mammalian endogenous RNAi. FEBS Letters *588*, 2550-2556.

Tan, M.H., Au, K.F., Yablonovitch, A.L., Wills, A.E., Chuang, J., Baker, J.C., Wong, W.H., and Li, J.B. (2013). RNA sequencing reveals a diverse and dynamic repertoire of the *Xenopus tropicalis* transcriptome over development. Genome Research *23*, 201-216.

Tarver, J.E., Cormier, A., Pinzón, N., Taylor, R.S., Carré, W., Strittmatter, M., Seitz, H., Coelho, S.M., and Cock, J.M. (2015). microRNAs and the evolution of complex multicellularity: Identification of a large, diverse complement of microRNAs in the brown alga *Ectocarpus*. Nucleic Acids Research *43*, 6384-6398.

Tarver, J.E., Donoghue, P.C.J., and Peterson, K.J. (2012). Do miRNAs have a deep evolutionary history? BioEssays *34*, 857-866.

Technau, U., and Schwaiger, M. (2015). Recent advances in genomics and transcriptomics of cnidarians. Marine Genomics *24, Part 2*, 131-138.

Thomson, D.W., and Dinger, M.E. (2016). Endogenous microRNA sponges: Evidence and controversy. Nature Reviews Genetics *17*, 272-283.

Thor, S., and Thomas, J.B. (1997). The *Drosophila* islet gene governs axon pathfinding and neurotransmitter identity. Neuron *18*, 397-409.

Tichon, A., Gil, N., Lubelsky, Y., Havkin Solomon, T., Lemze, D., Itzkovitz, S., Stern-Ginossar, N., and Ulitsky, I. (2016). A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. Nature Communications *7*, 12209.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science *329*, 689-693.

Ulitsky, I. (2016). Evolution to the rescue: Using comparative genomics to understand long non-coding RNAs. Nature Reviews Genetics *17*, 601-614.

Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: Genomics, evolution, and mechanisms. Cell *154*, 26-46.

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell *147*, 1537-1550.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, Duncan T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Reports *10*, 1297-1309.

Villar, D., Berthelot, C., Aldridge, S., Rayner, Tim F., Lukk, M., Pignatelli, M., Park, Thomas J., Deaville, R., Erichsen, Jonathan T., Jasinska, Anna J.*, et al.* (2015). Enhancer evolution across 20 ammalian species. Cell *160*, 554-566.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., and Chen, F. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854-858.

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. Cell *136*, 669-687.

Wang, S.H., and Elgin, S.C.R. (2011). *Drosophila* Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proceedings of the National Academy of Sciences USA *108*, 21164-21169.

Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., Wu, M., Xiong, J., Guo, X., and Liu, H. (2013). Endogenous miRNA sponge *lincRNA-RoR* regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Developmental Cell *25*, 69-80.

Warner, J.R., Soeiro, R., Birnboim, H.C., Girard, M., and Darnell, J.E. (1966). Rapidly labeled HeLa cell nuclear RNA. Journal of Molecular Biology *19*, 349-361.

Washietl, S., Kellis, M., and Garber, M. (2014). Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. Genome Research *24*, 616-628.

Wheeler, B.M., Heimberg, A.M., Moy, V.N., Sperling, E.A., Holstein, T.W., Heber, S., and Peterson, K.J. (2009). The deep evolution of metazoan microRNAs. Evolution & Development *11*, 50-68.

Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., and Kelly, K. (2005). Highly conserved non-coding sequences are associated with vertebrate development. PLOS Biology *3*, e7.

Wu, Y., Cheng, T., Liu, C., Liu, D., Zhang, Q., Long, R., Zhao, P., and Xia, Q. (2016). Systematic identification and characterization of long non-coding RNAs in the silkworm, *Bombyx mori*. PLOS ONE *11*, e0147147.

Yin, Q.F., Yang, L., Zhang, Y., Xiang, J.F., Wu, Y.W., Carmichael, G.G., and Chen, L.L. (2012). Long noncoding RNAs with snoRNA ends. Molecular Cell *48*, 219-230.

Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., Zhao, Y., Shen, B., Huang, X., Zhu, H.*, et al.* (2015). Opposing roles for the lncRNA *Haunt* and its genomic locus in regulating *HOXA* gene activation during embryonic stem cell differentiation. Cell Stem Cell *16*, 504-516.

Young, R.S., Marques, A.C., Tibbit, C., Haerty, W., Bassett, A.R., Liu, J.L., and Ponting, C.P. (2012). Identification and properties of 1,119 candidate lincRNA loci in the *Drosophila melanogaster* genome. Genome Biology and Evolution *4*, 427-442.

Zhao, T., Li, G., Mi, S., Li, S., Hannon, G.J., Wang, X.-J., and Qi, Y. (2007). A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. Genes & Development *21*, 1190-1203.

Figure 1 – Comparison of major non-coding regulatory systems in animals and their unicellular relatives. Cladogram representing known phylogenetic relationships between animals and their unicellular cousins. The presence or absence of the major classes of small RNAs (miRNAs, siRNAs, piRNAs) along with lncRNAs, circRNAs, distal enhancers and the architectural protein CTCF, is indicated. Check mark and dash symbols indicate the presence or absence of the components in these taxa, respectively. Circle indicates the information is not available at the time of writing. Yellow background highlights the animal kingdom. [a]Evidence for putative *bona fide* miRNAs and the presence of the miRNA-processing pathway components (*i.e.*, Drosha and Pasha) have been recently reported in the ichthyosporean *Sphaeroforma arctica* (Bråte et al., 2016). [b]While siRNAs have been lost in *S. cerevisiae*, they are present in other yeast species, including *S. pombe* (Drinnenberg et al., 2009). [c]CTCF is present in nematodes, but secondarily lost in *C. elegans* (Heger et al., 2009).

Figure 2 – Do metazoan long non-coding RNAs operate in evolutionarily conserved developmental co-expression modules (or networks)?. Although lncRNAs are important components of co-expressed gene modules from early-branching non-bilaterian phyla to complex vertebrates (Bråte et al., 2015; Gaiti et al., 2015; Necsulea et al., 2014), the investigation of lncRNA roles in an evolutionary framework has been drastically limited by the lack of the ability to detect significant sequence similarities and functional studies. A conceptual model is illustrated of how evolutionarily conserved networks of co-expressed orthologous genes could infer the potential biological function of specific lncRNAs in divergent species that lack sequence conservation.

Figure 3 – Distal enhancers may be a metazoan innovation. (A) A schematic representation of the presence or absence of distal enhancers in *Capsaspora*, *Amphimedon*, *Nematostella*,

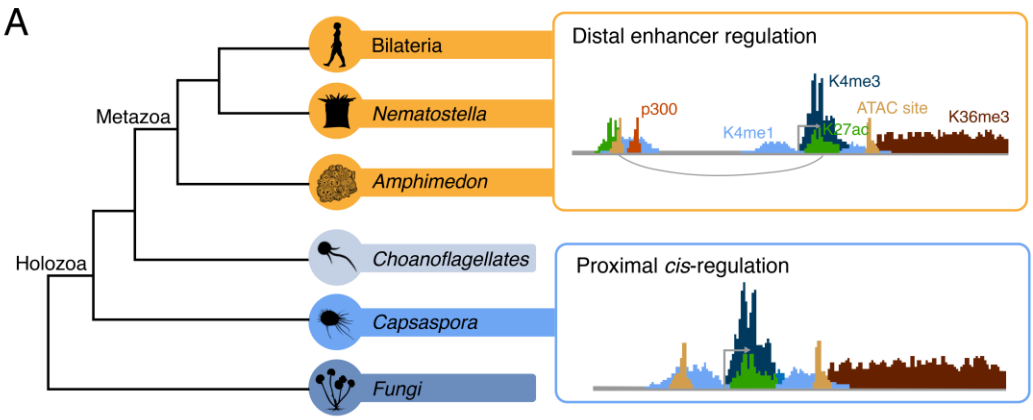
and bilaterians, together with the presence or absence of the typical chromatin signatures associated with animal enhancers [the transcriptional cofactor p300, histone 3 lysine 4 monomethylation (H3K4me1), histone 3 lysine 27 acetylation (H3K27ac), and ATAC site]. Adapted from (Sebé-Pedrós et al., 2016). (B) Regulatory elements, including enhancers and promoters, are engaged in multiple long-range interactions with many other regions through a looping mechanism, in which the transcriptional machinery is loaded at the enhancers and then reaches the promoter due to a physical interaction, facilitated by CTCF and cohesin (Li et al., 2016). All chromatin interactions are created and maintained in a hierarchy of 3D chromatin architectures, including topologically associated domains (TADs) (Pombo and Dillon, 2015). The mechanism of enhancer-promoter interaction in non-bilaterian animals lacking the architectural protein CTCF is unknown (Heger et al., 2012).

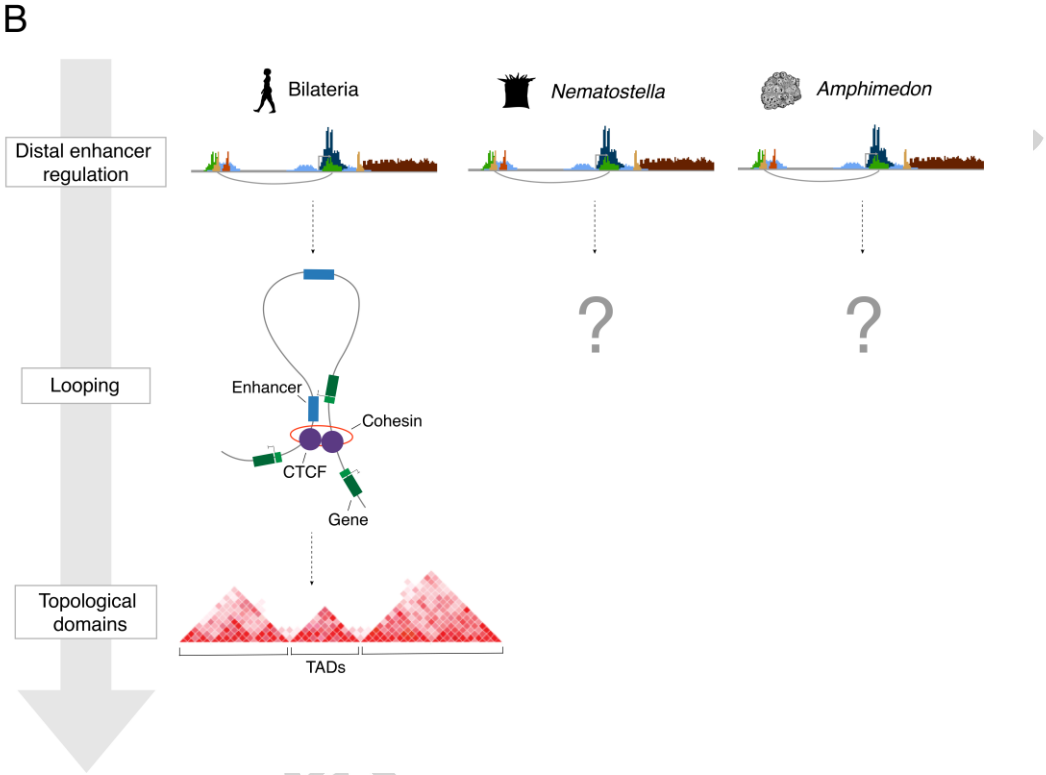
Figure 4 – Major steps in the evolution of the animal regulatory genome. The phylogenetic relationship of representative animal lineages and unicellular holozoans is shown. Highlighted are the major genomic innovations that correlate with the emergence and early diversification of animals. Some components of the metazoan regulatory landscape may predate the split of the metazoan and holozoan lineages, including a subset of core TF-TF regulatory interactions and lncRNAs. While the latter have been recently identified in unicellular relatives of animals, the evolutionary origin of lncRNAs as a group is still unclear. With a complex gene regulatory landscape already in place at the dawn of animals, the expansion of TFs and signaling pathways components, *cis*-regulatory DNA and non-coding RNAs, appear to underlie the morphological and functional diversification of eumetazoan animals. In bilaterians, the emergence of the architectural protein CTCF further allows more complex enhancer-promoter interactions. [a]Evidence for putative *bona fide* miRNAs and the

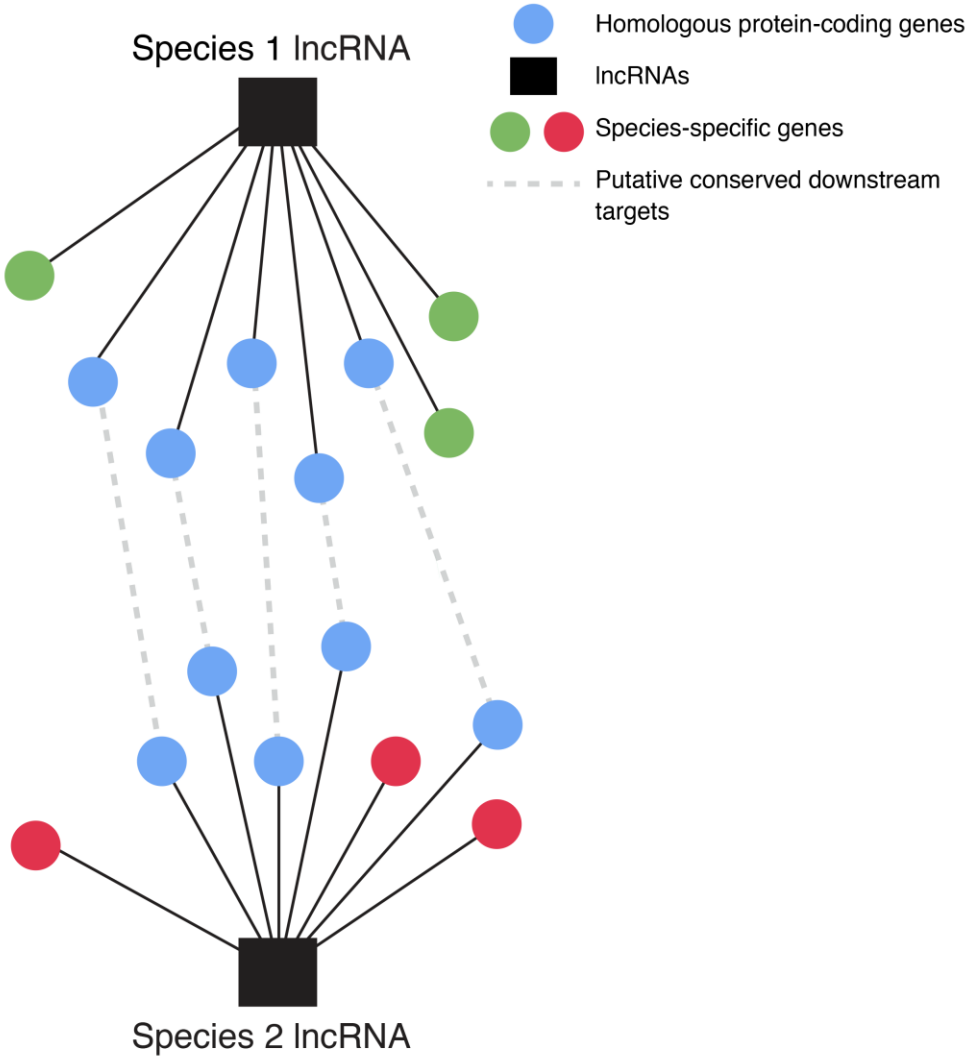
presence of the miRNA-processing pathway components (*i.e.*, Drosha and Pasha) have been recently reported in the ichthyosporean *Sphaeroforma arctica* (Bråte et al., 2016).

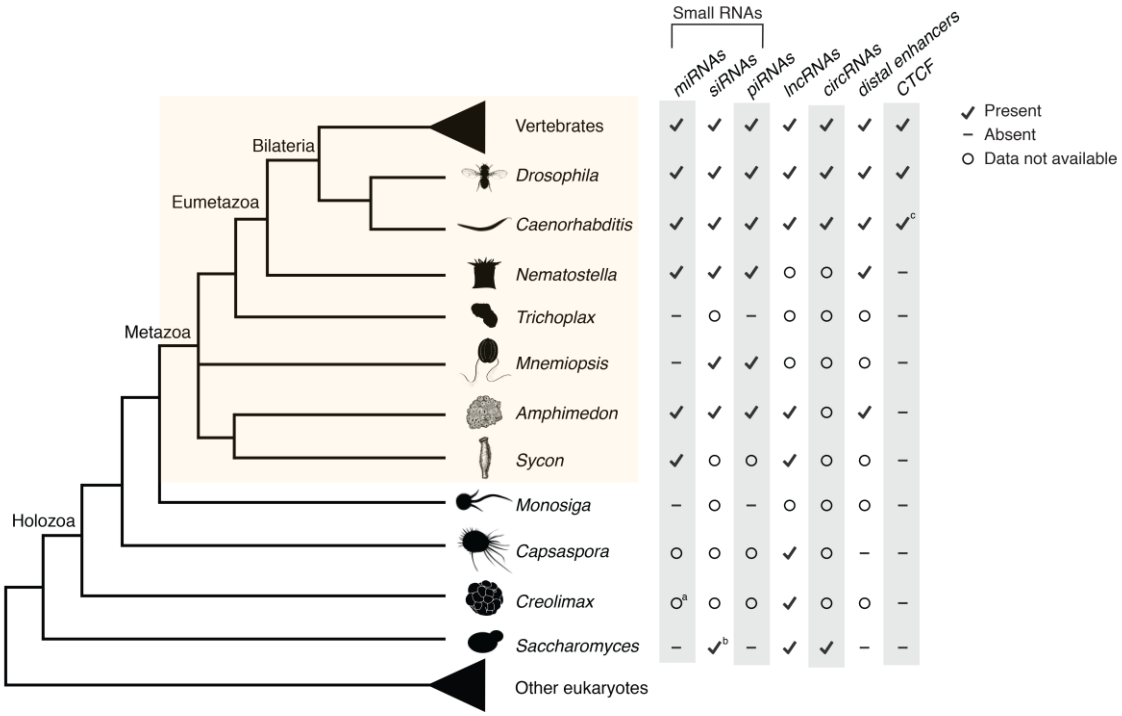
Highlights

- The non-coding genome is essential for multicellularity
- Review of non-coding RNA and DNA in non-bilaterian metazoans and unicellular holozoans.
- Most non-coding regulatory mechanisms antedate the evolution of metazoans.
- piRNAs and distal enhancers appear to be metazoan innovations.
- The CTCF/cohesin system appears to be restricted to bilaterians.






Species 1 lncRNA
Homologous protein-coding genes
lncRNAs
Species-specific genes
Putative conserved downstream targets
Species 2 lncRNA


Small RNAs
miRNAs
siRNAs
piRNAs
lncRNAs
circRNAs
distal enhancers
CTCF
Present
Absent
Data not available
Bilateria
Eumetazoa
Metazoa
Holozoa
Vertebrates
Drosophila
Caenorhabditis
Nematostella
Trichoplax
Mnemiopsis
Amphimedon
Sycon
Monosiga
Capsaspora
Creolimax
Saccharomyces
Other eukaryotes