"Hypothesis for the modern RNA world": a pervasive non-coding RNA-based genetic regulation is a prerequisite for the emergence of multicellular complexity

Irma Lozada-Chávez · Peter F. Stadler · Sonja J. Prohaska

Abstract The transitions to multicellularity mark the most pivotal and distinctive events in life's history on Earth. Although several transitions to "simple" multicellularity (SM) have been recorded in both bacterial and eukaryotic clades, transitions to "complex" multicellularity (CM) have only happened a few times in eukaryotes. A large number of cell types (associated with large body size), increased energy consumption per gene expressed, and an increment of non-protein-coding DNA positively correlate with CM. These three factors can indeed be understood as the causes and consequences of the regulation of gene expression. Here, we discuss how a vast expansion of non-protein-coding RNA (ncRNAs) regulators rather than large numbers of novel protein regulators can easily contribute to the emergence of CM. We also propose that the evolutionary advantage of RNA-based gene regulation derives from the robustness of the RNA structure that makes it easy to combine genetic drift with functional exploration. We describe a model

· I. Lozada-Chávez
Computations Evo-Devo Group, Department of Computer Science, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany

Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany · Max Planck Institute for Mathematics in the Science, Inselstraße 22, D-04103 Leipzig, Germany
Tel.: +49-341-9716694
Fax: +49-341-9716679
E-mail: ilozada@bioinf.uni-leipzig.de
· P.F. Stadler
Bioinformatics Group, Department of Computer Science, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany

Fraunhofer Institut für Zelltherapie und Immunologie – IZI Perlickstraße 1, D-04103 Leipzig, Germany · Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria · Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe NM 87501, USA
· S.J. Prohaska
Computations Evo-Devo Group, Department of Computer Science, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany

Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany

which aims to explain how the evolutionary dynamic of ncRNAs becomes dominated by the accessibility of advantageous mutations to innovate regulation in complex multicellular organisms. The information and models discussed here outline the hypothesis that pervasive ncRNA-based regulatory systems, only capable of being expanded and explored in higher eukaryotes, are prerequisite to complex multicellularity. Thereby, regulatory RNA molecules in Eukarya have allowed intensification of morphological complexity by stabilizing critical phenotypes and controlling developmental precision. Although the origin of RNA on early Earth is still controversial, it is becoming clear that once RNA emerged into a protocellular system, its relevance within the evolution of biological systems has been greater than we previously thought.

1 Introduction

For the first 2.5 billion years (Ga) of life on Earth, the size of most species was generally much smaller than 1 mm, and rarely was this exceeded [20][106]. Prokaryotes, with few exceptions, have remained unicellular organisms optimizing their size [20][113], gene and protein content [102] as well as the flexibility of their protein-based gene regulation [80] in order to maximize metabolisms that drive Earth's biogeochemical cycles [33][37][64]. Plants, animals, fungi, and protozoan seaweeds, on the other hand, are multicellular organisms that dominate the terrestrial landscapes and the oceans since one billion years ago with a remarkable diversity in genotypic and phenotypic complexity [67][70]. *A multicellular organism is a collection of self-organized cells that express different phenotypes, despite having the same genotype, in response to the specialization of tasks to perform a cooperative physiological division of labour within an economic organization.*

2 How Difficult is the Transition to Multicellularity?

At least 25 independent transitions to multicellularity have been recorded during the evolution of cellular complexity on Earth [118][16][43][70]. The transition to multicellularity has been repeatedly promoted from unicellular and colonial ancestors [20][60][67][121][52][16][43][91]. In several bacterial and eukaryotic organisms, these transitions are an inducible response to environmental stimuli such as predation and starvation [60][18][98][16][82][140]. Surprisingly, multicellurarity can be reverted back to an unicellular state in several bacterial lineages [134][18]. A special case are *defectors* (i.e., mutant cell lineages that selfishly improve their own fitness and fail to cooperate with the other cell types of the organism) in vertebrates, where non-viral transmissible cancerous cells could in effect become independently evolving unicellular colonies [107][97][8][128].

The frequent origination and spread of multicellularity suggests that (1) selection favoring this transition is pervasive across organisms and time, (2) the genetic and developmental obstacles to this transition are relatively "easy" to overcome, and (3) adaptive mechanisms that control defectors and stabilize the transition

are widely available in natural populations [114][43]. According to Grosberg and Strathmann (2007), all these pieces of empirical evidence together support the idea that evolution of muticellularity can be considered itself a significant but minor transition, and that cellular diversity can evolve easily when functionally called for by selective advantages. Nevertheless, multicellularity shows two states that presumably are not just extremes of a continuous spectrum but are fundamentally different: *simple* or *complex* [67][118][16][70]. Increase of organismal size, diversity of cell types, division of labor and functional specialization are interrelated reflections of multicellular complexity; however, all these factors are differentially represented between simple and complex multicellular organisms, as described below (Table 1). Thus, we here support the hypothesis that transitions to *complex multicellularity* have required, in addition to the advent of the eukaryotic cell and several other key factors, the development of a *pervasive* non-protein-coding RNA-based gene regulation. Non-protein-coding RNAs (ncRNAs) and proteins have been used as regulatory molecules in prokaryotes and eukaryotes, and conventional wisdom holds that increased complexity requires more regulators. Nevertheless, we here polarize this major evolutionary transition to the contrary: ***pervasive* non-coding RNA-based regulatory systems are prerequisite to complex multicellularity**. We are going to argue (1) why *pervasive* non-coding RNA-based regulatory systems can only be supported and easily explored in higher eukaryotes, and (2) how a vast expansion of regulatory ncRNAs rather than large numbers of novel protein regulators can easily contribute to the emergence of complex multicelullarity.

Simple multicellular organisms (SMOs) include filaments, balls or sheets of cells that arise either *via* mitotic division from a single progenitor with the offspring sticking together (e.g., from an aquatic origin) or when several solitary cells aggregate to form a colony (e.g., from a terrestrial origin). They form a coherent and reproducible morphology by cell-cell adhesion, and differentiation of somatic and reproductive cells is common [139][16][43]. However, complex differentiation patterns and intercellular signaling are limited, so that every cell lies in direct contact with the environment during active metabolism [70]. SMOs are found in both multiple eukaryotic lineages (such as chlorophyceae, dictyostelia and oomycetes) and in some eubacterial clades, e.g., cyanobacteria, myxobacteria and actinobacteria [60][118][16]. In fact, the first signs of cell differentiation come from fossils of filamentous and mat-forming cyanobacteria-like organisms that diverged once between 2.4 and 2.1 Ga [132].

Complex multicellularity, on the other hand, is limited to Eukarya where it arose independently in at least six clades (Fig. 1): once for the eumetazoan animals [67], but multiple times (with possible secondary losses) in embryophytic land plants, florideophyte red algae, stramenopile brown algae (from the order Laminariales), basidiomycete, and ascomycete fungi [17][118][29][99][91][43][70]. Complex multicellularity arose relatively late in the history of life, probably less than 1000 million years (Ma) ago [10], and left an extended fossil record during the Ediacaran and Cambrian periods ($\sim$600 Ma ago) [106][70]. *Complex multicellular organisms* (CMOs) show not only evidence of genes involved in cell-cell and cell-matrix adhesion but also a diverse "toolkit" of genes associated with developmental and cell-death programs comprising intercellular signaling, specialization of cell types, and multiple tissue differentiation patterns mediated by complex regulatory networks. This genetic toolkit (including protein-based gene regulation) has been the

product of evolutionary innovations, tinkering and expansions of genetic material from ancestral unicellular organisms [67][68][118][29][127][38][126]. Interestingly, the "multicellular genetic toolkit" corresponds only to a few hundred genes from a few dozen gene families [118][35][38][126][70] that belong to the 3,000 novel gene families diverged from the last common ancestor of eukaryotes [72].

3 The Eukaryotic Cell as a Source for Complex Multicellularity

Large organismal size, the origin of endosymbiotic energy production, and a "passive" increase of non-protein-coding DNA (ncDNA) in the genome may have predisposed the eukaryotic cell as substrate for complex multicellularity.

3.1 Increase of Body Size and Cell Differentiation

Organismal size positively correlates with the number of cell types in CMOs [20][17][33][16][133][89] (Table 1). Once size had increased, the putative advantages of this change would follow [16]. Knoll (2011) proposed that large three-dimensional sizes of CMOs could have been enhanced by a positive feedback cycle from the availability of ambient oxygen (pO_2). The gradual increase of pO_2 in the oceans and atmosphere about 2.5 Ga ago [54], and the exposure to pO_2 during the transition of life from water to land [50][16][70], would have increased the permissible size of diffusion-limited multicellular organisms [106]. Oxygen may not have started all CMOs, but it would have imposed severe constraints on the evolution of macroscopic organisms with high energy demands. A larger size, in turn, would have allowed larger surface-to-interior gradients of oxygen, nutrients, signaling molecules, bulk transport, as well as the formation of reactive oxygen species (in response to environmental cues) which would have been capable of inducing cell differentiation [14][1][76][70]. Environmental cues selecting cell differentiation more than just larger organismal size should have been crucial for the transition to CMOs, given that the presence of developmentally differentiated cell types in a colony (or in an organism) is what makes it truly multicellular [139].

3.2 The Acquisition of an Endosymbiotic Energy Production

In general terms, a cell type is a cell with a discrete pattern of gene expression driving a distinct morphological or functional cellular shape in comparison to other cell types with the same organismal genotype. Gene expression of cell types not only needs the involvement of signal(s) to start and maintain differentiation and a considerable diversity of regulatory elements to control expression, but it also needs a powerful source of energy to express both conserved and novel genes in a combinatorial manner. Thus, the massive difference in mean genome size between prokaryotes and eukaryotes is most revealing in terms of the energy available to transcribe and translate their genes [113][74]. Whereas the energetic cost of possessing genes is trivial ($\sim$2%), the cost of expressing them as RNA transcripts and proteins is not: protein synthesis consumes most ($\sim$75%) of the cell's total energy budget [136][74]. For example, if a bacterial genome is increased tenfold in size, it

could still be replicated, but there is no current known mechanism (such as the number of regulatory proteins or ribosomes, carbon metabolism, respiratory chain or giant polyploids) that can circumvent the energetic barrier to express ten times as many proteins [113][74]. The origin of eukaryotes, however, entails a bioenergetic innovation that is key to sustain multicellular life. Lane and Martin (2010) elegantly argue that the endosymbiosis that gave rise to mitochondria restructured the distribution of DNA in relation to bioenergetic membranes. By enabling oxidative phosphorylation across a wide area of internal membranes, mitochondrial genes permitted a remarkable 200,000-fold expansion of genome size compared to bacteria. Thereby, mitochondrial power expanded the genotype that an eukaryotic cell could express, inherit, and evolve by four to six orders of magnitude, affording the cell the possibility (but not the necessity) of becoming complex.

3.3 A ncRNA-Based Regulatory Network is Hidden Within ncDNA

Large-scale tandem and block duplications have been extensively reported among eukaryotes, hinting at polyploidy in their ancestry [42]. Furthermore, fundamental changes in gene structure, such as the advent of introns, allowed the expansion of the eukaryotic proteome by alternative splicing [100]. In contrast to prokaryotes [71], however, there is no clear relationship between eukaryotic complexity and either genome size (C-value) or the number of protein-coding genes (G-value) (Fig. 1). In higher eukaryotes, indeed, the vast majority of nuclear DNA is non-protein-coding (ncDNA) [85][84][130]. Intragenic ncDNA is either intronic sequence or untranslated region (UTR), whereas intergenic ncDNA is composed (in different proportions across species) of repetitive transposable elements from class I (e.g., LINEs, SINEs and LTRs) and class II (e.g., MITEs), simple sequence repeats as well as segmental and pseudogene duplications [42][84]. The progressive expansion of ncDNA in higher eukaryotic organisms is thought to be a consequence of the reduced efficiency of selection acting against the passive accumulation of "mutationally hazardous" DNA in taxa experiencing elevated magnitudes of random genetic drift [84]. This effect may be explained because of reduced effective population sizes (invertebrates and vascular plants $\sim 10^5 - 10^6$, and vertebrates $\sim 10^4 - 10^5$; whilst prokaryotes $\sim 10^8$, and unicellular eukaryotes and fungi $\sim 10^7$) [85][136], reduced recombination in large genomes, and a mutational bias toward insertions of large segments of DNA [83][84]. The extent and ways by which these ncDNA components contribute to the phenotype of eukaryotic species are still being elucidated [63][42]. Nevertheless, some ncDNA components retain several regulatory elements (in particular *cis*-regulatory ones), and a considerable proportion of them encodes for a huge diversity of ncRNAs (i.e., functional RNA molecules that are not translated into a protein), at least a substantial fraction of which is thought to have regulatory functions.

Certainly, regulatory ncRNAs were first found in prokaryotes and, for instance, they currently represent $\sim 2\%$ of the total number of genes in the unicellular *Escherichia coli*, which represents almost half of the genes encoding for protein regulators ($\sim 5\%$) [137]. There is increasing evidence of (non-coding) antisense transcripts in bacteria for which regulatory functions are at least suspected [125]. More than 100 types of regulatory ncRNAs have been identified throughout the bacterial kingdom, and riboswitches are one of the best known examples. A particular class

of riboswitches, namely those responding to coenzyme thiamine pyrophosphate, has also been identified in some eukaryotic lineages [22][15]; however, it exhibits an uncertain evolutionary origin in Eukarya [129][15]. With the exception of small nucleolar RNAs (snoRNA) present in Archaea and Eukarya, there is no homologous relationship between the regulatory ncRNAs found in prokaryotes and eukaryotes. Furthermore, the physiological role of bacterial regulatory ncRNAs has been evolutionary driven to mediate rapid responses to changing environmental conditions by modulating specific metabolic pathways or stress cues, like pathogenesis and SOS response [116][137].

4 ncRNAs Dominate the Genotype and Phenotype of Complex Multicellular Organisms

In contrast to prokaryotes, regulatory ncRNAs are encoded basically everywhere in the eukaryotic genome, and in particular they cover the ncDNA regions which dominate the genotypes of CMOs. Regulatory ncRNAs are transcribed from pseudogenes, they are also produced from protein-coding loci both by alternative splicing and as independent transcripts (e.g., anti-sense RNAs, enhancer RNAs, intronic transcripts, UTR associated RNAs, repetitive elements), and from their own "intergenic" genes [66][111][135][115][19]. It appears generally accepted by now that eukaryotic ncDNA is pervasively transcribed [77][12][62][27][131], but see [6] and [7] for a dissenting opinion. Accordingly, it has been extensively reported that regulatory ncRNAs show a differential, widespread and complex transcription that gives rise to a considerable number of functional ncRNA families (Table 2) [61][58].

Different classes of small and large ncRNAs have been reported as regulating a larger number of both species-specific and deeply conserved cellular processes in tissue identity and stem cell self-renewal and differentiation through well defined mechanisms within the major eukaryotic clades [93][40][111][25][101][73] (Fig. 2). Furthermore, ncRNAs have been shown to regulate almost every level of gene expression, including the activation and repression of homeotic genes and the targeting of chromatin-remodeling complexes (i.e., epigenesis) [3][26][105]. Indeed, recent transcriptome analyses and different experimental approaches in cell development provide strong evidence that perturbations in ncRNA regulation are involved in complex developmental disorders, including cancers and neurological diseases in mammals; for review see [31] and [112]. Nevertheless, the loss of ncRNA function rarely results in a lethal phenotype. One of the counter-examples is the elimination of microRNA-1-2 (a miRNA expressed in skeletal muscle in vertebrates), which results in a lethal phenotype with defects to cardiac morphogenesis, electrical conduction, and cell cycle control in mouse [142]. With the exception again in mouse [11], however, the genetic inactivation of the central ncRNA processing enzyme Dicer in vertebrates does not dramatically affect cell differentiation and gene expression patterning [28][41][48]. All these lines of evidence in conjunction with the evolutionary trends of these regulators (described below) support the hypothesis that *regulatory ncRNAs are primarily moderators of complex multicellular phenotypes on Earth.*

5 Why Can ncRNAs Quickly and Selectively Acquire a Regulatory Functionality in Complex Multicellular Organisms?

The observations that i) energy costs constrain the evolution of gene expression in any organism, ii) ncRNAs could be fixed and even conserved on eukaryotic genomes by nearly neutral forces (perhaps without any advantage or disadvantage for the organism in the beginning), and that iii) possible *pervasive* transcription of new ncRNAs could perturb existing regulatory networks, raise the question how critical changes of gene regulation have arisen during the evolution of complex multicellularity. Wagner (2005) suggested at least three scenarios that may operate at the same time to promote any gene expression change. First, changes in messenger RNA (mRNA) or protein half-life may contribute significantly to reduce the cost of controlling the cellular concentration of a gene product, given that half-lives are energetically less constrained than synthesis rates. Second, a substantial influx of regulatory mutations may balance the total energy consumption by increasing the synthesis of some gene products and decreasing the synthesis of others. Finally, substantial changes in mRNA and protein synthesis rates can only go to fixation in a large population (like in bacteria and unicellular eukaryotes) when they provide an advantage sufficiently large to overcome the cost and thus the effect of selection opposing it. Ohta's *Near Neutral Theory* shows, however, that slightly deleterious mutations can be fixed by genetic drift [104]. This effect allows to transcripts without an adaptive effect to persist and spread in a population despite its small energetic cost, and to accumulate mutations. The mutation-selection balance is shifted, furthermore, in favour of genetic drift in small populations, as we expect for the higher multicellular organisms. Indeed, apparently non-functional pseudogenes (i.e., degenerate copies of functional protein-coding transcripts) continue to be expressed in measurable numbers from the human genome [143]. New potential regulatory changes can then open the possibility (but again not the necessity) to become positively selected by exploring and canalizing functional phenotypic novelties.

ncRNA-based regulatory networks have shown their evolutionary potential to guide new critical phenotypes. First, ncRNAs may allow the cell to exert faster control of gene expression and thus to improve adaptation to environmental conditions in a specific space and time through low energetic costs. A major distinction between protein- and RNA-based gene regulations can be found in their mode of action. *Transcription factors* (TFs) recognize sequence-specific *cis*-motifs in the regulatory upstream regions of target genes (i.e., the promoter), and regulate transcription when bound to their promoters. After transcription, in contrast, *micro RNAs* (miRNA) recognize complementary *cis*-acting sites in the 3' UTR of the mRNA, and ultimately control the protein translation of the mRNA by promoting its degradation once bound to a target site [93] (Table 2). On the basis of common sequence motifs and structural features, *small nucleolar RNAs* (snoRNAs) bind to a transcript *via* base complementarity to regulate alternative splicing at sequences subject to RNA editing [69]. Additionally, a large number of different regulatory functions have been found for *long non-coding RNAs* (lncRNA) which often do not appear to show dependence on particular secondary structures. The control of mRNA populations by lncRNA includes the organization of nuclear bodies, a direct involvement in epigenetic mechanisms and transcriptional regulation as well as the inhibition of several receptors [92][138][26][65]. Furthermore, snoRNAs

and lncRNA can be also processed into small regulatory RNAs similar to miR-NAs [138][123]. Therefore, *precision in genic output (as measured by the variation in number and half-life of mRNA molecules) is often achieved by ncRNAs rather than by protein regulators.* ncRNAs could confer robustness (i.e., invariance of the resulting phenotype in the face of perturbation) to regulatory networks by preventing unwanted ectopic protein molecules, buffering fluctuations in expression levels or reducing transcriptional noise [55][124][5][45].

Unlike protein-coding sequences, moreover, *ncRNAs do not necessarily exhibit a conserved sequence to perform their functions.* In most cases, indeed, the biogenesis or function of the RNA molecule is only possible if the molecule folds into a characteristic two- and three-dimensional structure *via* formation of intra-molecular base pairs into "stems" or "helices" (Fig. 3b). The disruption of these paired regions through mutations in the primary sequence may result in conformational changes of the structure and can compromise the function of the RNA molecule; however, *compensatory mutations* that replace one type of base pair by another one in the paired regions of the molecule can restore its functional conformation (Fig. 3c). Under the influence of various selection forces, such as purifying (speed up the loss of mutant alleles), stabilizing (stabilize the frequency of an allele in a population), or positive (promote faster fixation of an advantageous alleles in a population), a RNA molecule may accumulate nucleotide double-substitutions (i.e., covariations) to maintain a structural-functional class or may store mutations to explore a new one.

The effect of mutations on structured RNAs is quite well understood. On the one hand, *RNAs are robust against mutations in the sense that a large fraction of mutations does not affect the folding* (Fig. 3c). These neutral mutants form extended neutral networks in sequence space on which drift leads to a diffusion-like evolutionary dynamics. *Another large fraction of mutations leads to dramatic structural changes*, so that any frequent structure is realized in the vicinity of any arbitrary chosen sequence, an effect termed *shape space covering* [122]. Taken together, these two competing effects imply that drift leads to a rapid exploration of novel variants at the fringes of the neutral networks [56][57]. In small populations, slightly detrimental mutations at ncDNA regions can also be fixed by genetic drift [103], so that the small energy cost of transcribing non-functional ncRNAs does not lead to their rapid elimination from the genome. Furthermore, second order effects, such as an increased probability to produce offspring with a detrimental mutation in a gene that was fitness-neutral in the parent, cannot be selected because drift dominates the very small effective fitness effect. The first order effect of an advantageous mutation is readily selectable, however. Thus, non-functional transcripts can "wait" for rare advantageous mutations to place them under stabilizing selection. The evolutionary dynamics thereby becomes dominated by the accessibility of advantageous mutations [39]. *This structural accessibility of the RNA molecule could explain in part the* **structural and functional convergence** *of the regulatory miRNA-like RNA class in plants, animals and fungi, which is in congruency with an independent origin of their multicellular development* [75][32].

5.1 Evolution and Selection of Regulatory ncRNAs in Eukaryotes: The Case of microRNAs

Functional regulatory ncRNAs, even those under strong stabilizing selection, may exhibit rapid evolution at the sequence level. Thus, some subclasses are clearly under purifying selection acting predominantly on the secondary structure in order to keep their regulatory function [110][95][23][87][111][96]. Furthermore, the genetic structure involved in the transcription and function of some regulatory ncRNAs, such as introns and splice sites, is also subject of stabilizing selection [53][119]. Hence several models have been suggested to describe how selection may be driving the birth-and-death evolutionary dynamics of some ncRNAs [55][24][78][109][32].

In particular, comparative analyses of miRNA homologs in diverse animal and plant species have revealed both long-term maintenance and taxa-specific occurrence of miRNA families [36][51][81][32]. In animals, it has been proposed that new miRNAs could evolve by point mutations from existing hairpin structures in the genome that are transcribed at low levels and in specific cell types. This process is followed by selection against inadequate miRNA/mRNA pairing and expression modifications, and simultaneously is preserving many of the neutral or advantageous targets that increase the expression of the miRNA (without being highly deleterious to the organism) [24][78][109][32]. Thus, animal miRNAs and fungal *microRNA-like RNAs* (milRNAs) have exploited the robustness of the hairpin structure both to maintain the regulatory function and to diversify their activity by targeting imperfectly complementary sequences [78][75]. Because miRNA/target pairing is almost perfectly complementary and extends over twice as many nucleotides in plants, when compared to animals and fungi, the evolution of new plant miRNAs has been proposed to be driven by inverted duplication of target gene sequences (i.e., pre-miRNA-encoding regions) [2]. Thus, newly formed expressed RNA hairpins may experience successive selective sweeps until they reach an evolutionarily optimized form, allowing a one-to-one miRNA/target relationship to be established, and then be incorporated into new or existing regulatory networks [90][32].

Both models have been supported by experiments showing that unlike highly conserved-ancient plant and animal miRNAs, evolutionary young miRNAs are typically expressed at low levels, processed imprecisely, lack targets, and display patterns of neutral variation [4][24][90][81][32]. These trends suggest that young MIRNA loci tend to evolve neutrally, possibly helping to maintain negative pleiotropic effects at low levels, until compensatory and advantageous mutations have emerged [24][32]. Exceptional evidence for evolutionary optimization of young pre-miRNA stem-loop structures has been recently presented for two miRNA-encoding loci in *Arabidopsis thaliana* [90] as well as for five miRNA-encoding loci in *Drosophila melanogaster* [81]. In both works, the authors demonstrated that pervasive variation occurs at miRNA-encoding loci, and that the structural variation among alleles suggests non-random evolution of a thermoresistant substructure in the miRNA precursor, which presumably impacts the processing of the mature form. In *A. thaliana*, for instance, miR856 shows a weak signature of a selective sweep and miR824 displays signs of balancing selection; whereas the polymorphism pattern of miR310/311/312/313 in *D. melanogaster* is indicative of hitchhiking under positive selection.

Detailed knowledge of evolutionary trends of regulatory non-coding transcripts is currently restricted to the much better understood microRNAs. Even though some genome projects have estimated the proportion and conservation of some ncRNA families, their extension and evolutionary patterns on a large phylogenetic scale are still being determined. Part of the methodological problem lies in the characterization of ncRNA evolution by large inhomogeneous variations in both sequence and structure [94] (Fig. 3a and c), which hampers their discovery and functional description. The assessment of a large compendium of ncRNAs in basal eukaryotes would help us to understand the origin and evolution of ncRNA-based gene regulation in Eukarya. Interestingly, *both genomic and functional similarities among regulatory ncRNAs might be the result of convergent evolution due to duplication and transposition mechanisms, similarly processing enzymes, and comparable selective pressures across species, but it might also reflect either the divergent evolution from a common RNA ancestor or the transition from one ncRNA type to another* [123].

5.2 Why and How Can a Vast Expansion of RNA Regulators Contribute to the Emergence of Complex Multicellularity Rather than Large Numbers of Novel Protein Regulators?

Many families of regulatory proteins that contribute to the complexity of developmental programs have been conserved, expanded and innovated throughout the evolution of eukaryotes. They are, however, under more stringent constraints than regulatory ncRNAs. Novel transcription factors, for instance, require the independent acquisition of DNA binding sites in all target promoters, while additional paralogs are forced to rapidly diverge from an existing regulatory module to avoid deleterious effects at the time of acquiring a new target or domain of expression (e.g., in a new tissue) [24]. Furthermore, for a new protein to come under positive selection, the mRNA must be transcribed (easy with pervasive transcription) and the messenger must be translated (maybe also easy, but protein synthesis is energy intensive). Hence a neutral coding gene should be selected against more strongly than a neutral non-translated transcript. Indeed, proteincoding genes can lose their coding capacity and hang around long enough as pseudogenes to become regulatory ncRNAs [141][46]. One of the best documented examples is the regulatory ncRNA *Xist*, which is exclusively present in eutherian mammals and its regulatory role is involved in the initiation process of X chromosome inactivation (Table 2). *Xist* evolved by pseudogenization of the protein coding gene *Lnx3*, which is conserved in all vertebrate classes, except Eutheria. Both *Xist* and *Lnx3* are thus present in a mutually exclusive manner at a syntenic locus in all vertebrates [34].

Conversely, multiple aspects of ncRNA function and evolution, such as a) *leaky* repression of transcripts to allow "fine-tuning" of gene expression, b) reduction of phenotypic variation by buffering genetic noise or avoiding unwanted ectopic molecules, c) a presumably regulatory increase in lineages of CMOs over geologic time, and d) the rarely dramatic impact in a phenotype due to loss of function, suggest that *regulatory ncRNAs act as key players in canalizing genetic programs* [55][109]. In this context, *canalization* can be visualized as the process of formation of *virtual* phenotypic (relatively) invariant canals in which developmental programs flow [55]. If the main function of ncRNA regulation is to stabilize gene

expression levels, then almost paradoxically, ncRNAs may increase the heritability of a phenotypic trait through natural selection by decreasing its expression variability about the mean, i.e., making robust the phenotype (a result) by canalizing the genetic program (a cause). Hence, *regulatory ncRNAs could be the evolutionary innovatory instrument to canalize development such that phenotypic variation decreases over geologic time at the cost of increasing developmental precision and consequently enhancing the morphological complexity* [109].

Here, we suggest a model that explains why novel RNA regulators with advantageous functions are readily accessible in multicellular organisms. Consider a gene that is expressed at some homogeneous level x in an organism. The expression level x will be selected to optimize the organism's fitness. Now suppose that the organism has two cell types. Figure 4 shows the fitness landscape as function of the expression levels in the two cell types $f(x_1, x_2)$. As long as the regulation of the gene is identical in both cell types, only the points on the median line ($x_1 = x_2$) are accessible. If a regulatory ncRNA is innovated in one of the two cell types (regulator + in the Fig. 4), then it can modify the expression levels of the gene. Evolution can now adjust the organism's fitness to optimize $f(x_1, x_2)$ without the constraint that $x_1 = x_2$. Mathematically, we know that $max f(x_1, x_2) \geq max f(x, x)$ simply because we now optimize over a larger domain. In general, the optimum for $f(x_1, x_2)$ will be reached for some point $x_1 \neq x_2$ so that the fitness difference $f(x_1, x_2) - max f(x, x)$ will generically be strictly positive for a non-empty $set A$ of expression values x_1 and x_2. If the new regulatory ncRNA takes the system into this $set A$ with a small energetic cost (S_{cost}), then it has a "window of opportunity" to be placed under stabilizing selection. Thus, this new regulatory ncRNA could make the phenotypic traits influenced by the regulated gene much more "evolvable". If the organism's fitness depends on the heritability of those phenotypic traits and, accordingly, also on the expression levels of that gene, then the new regulatory ncRNA becomes subject to positive selection (S^*) to optimize $f(x_1, x_2)$ (relationship shown in Fig. 4 as $S_{cost} + S^*$). If the fitness depends continuously on the expression levels, moreover, $set A$ touches the maximal value on the median line and hence is reachable from (x, x) by either a relative increase or a relative decrease of x_1 over x_2. In summary, for every novel RNA regulatory molecule, there is a "window of opportunity" for being placed under stabilizing selection if it can be used to disentangle expression patterns of other functionally relevant genes in different cell types. We argue that this mechanism makes advantageous innovations easily accessible in differentiated multicellular organisms. The resulting refinement in regulation in turn could lead to the innovation of further cell types, opening the same "window of opportunity" again for the incorporation of further protein and/or RNA regulators. *Since ncRNAs are much easier to generate from genomic DNA and energetically cheaper to produce than functional proteins in eukaryotes, we conclude that complex multicellular organisms should be able to rapidly accumulate ncRNA regulators –as we observe indeed in evolutionary history.*

This avenue to increasing cellular complexity is much less attractive in single-cell organisms, and in particular prokaryotes with their small size, in which gene products are distributed much more homogeneously. The mechanism of Fig. 4 critically depends on the availability of at least two cell types with the same genetic background and the same ancestral pattern of gene expression from the parental cell, a situation simply not available either in unicellular organisms or in SMOs

organized in colonies during certain life stages. Solitary cells thus do not generically provide an opportunity for the invasion of novel regulators, at least not when they are already well-adapted to their environment *(i.e., a lack of selective pressure or trade-offs)*[43]. The smaller size and hence the expected larger effective population size, in concert with strong dependence of genome size on energy constraints, also implies a much larger selective pressure against non-functional transcripts. Accordingly, prokaryotes should also have a much smaller repertoire of nearly neutral transcripts that persists for much shorter time scales, and thus have a much smaller chance to stumble across a functional improvement or innovation. All these constraints have placed protein regulators, instead of ncRNAs, to "easily" fulfill the selective advantages to dominate the regulation of phenotypes in prokaryotes. Global protein regulators, acting in conjunction with cis-regulatory elements, protein and RNA co-regulators, and specific environmental signals, control the morphogenesis and metabolic complexity of bacterial species [88][79]. Accordingly, bacterial protein regulators are highly flexible in switching their mode of action between global and local targeting and have a propensity to spawn homologs with similar function at short evolutionary time scales [86][80][79]. Furthermore, the total number of transcription factors positively correlates with genome size and life style [21][86]. Therefore, each bacterial species has evolved its own set of transcription factors, suggesting that the emergence of distinct repertoires of protein regulators is a crucial step for the adaptation to new environments and to control the bacterial phenotypic variance.

6 Role of RNA Throughout Evolution of Life on Earth

RNA is one of the three major biopolymers that manages the fluxes of biological information of life on Earth. The "RNAworld" hypothesis proposes that the biochemical, structural, and catalytic properties of RNA once were sufficient to encode the genotype, to decode the phenotype, and to drive evolution in the early state of life on Earth. As a corollary, RNA has been replaced to a large extent by DNA and proteins on the way to contemporary life. Its capabilities, however, have been preserved in the core of the most fundamental and highly conserved biological processes across the three domains of life [59]. Furthermore, the *pervasive* regulatory role of RNA in the three domains of life, particularly in eukaryotes, has been recently recognized (Fig. 2).

We suggest here that non-coding RNA-based genetic regulation is a prerequisite for the emergence of multicellular complexity. The evolution of complex multicellularity is based upon the successive innovation and refinement of cell and tissue types, requiring a flexible layer of compartment-specific gene regulation. We have argued here that novel regulators that refine gene expression patterns are easily selected for, and that non-coding transcripts are entirely capable of fulfilling such roles at all levels of cellular regulation. Since they are less costly than peptides, non-functional transcripts are available as evolutionary raw material for much longer time spans, and preferably in smaller populations where genetic drift is stronger. Thus, non-coding transcripts have much more time and much more widespread opportunities to find an adaptable role. Functional ncRNAs, therefore, are integrated into the genomes of complex multicellular organisms at a faster pace than paralogous proteins. Consequently, regulatory ncRNAs become the driving

force behind increasingly complex modes of regulation. Accordingly, it would not be hard to expect that the total number of regulatory ncRNAs could be higher than the total number of protein regulators in CMOs. If so, regulatory ncRNAs could substantially contribute to the theorical and practical definition of simple and complex multicellularity. The origin of RNA on early Earth is still controversial [117]. Nevertheless, it is becoming clear that once RNA emerged into a protocellular system, its relevance within the central dogma of biology has been greater than we have thought.

References

1. Aguirre, J., Ros-Momberg, M., Hewitt, D., Hansberg, W.: Reactive oxygen species and development in microbial eukaryotes. Trends Microbiol **13**(3), 111–8 (2005). DOI 10.1016/j.tim.2005.01.007
2. Allen, E., Xie, Z., Gustafson, A.M., Sung, G.H., Spatafora, J.W., Carrington, J.C.: Evolution of microrna genes by inverted duplication of target gene sequences in arabidopsis thaliana. Nat Genet **36**(12), 1282–90 (2004). DOI 10.1038/ng1478
3. Amaral, P.P., Dinger, M.E., Mercer, T.R., Mattick, J.S.: The eukaryotic genome as an rna machine. Science **319**(5871), 1787–9 (2008). DOI 10.1126/science.1155472
4. Axtell, M.J., Snyder, J.A., Bartel, D.P.: Common functions for diverse small rnas of land plants. Plant Cell **19**(6), 1750–69 (2007). DOI 10.1105/tpc.107.051706
5. Baek, D., Villn, J., Shin, C., Camargo, F.D., Gygi, S.P., Bartel, D.P.: The impact of micrornas on protein output. Nature **455**(7209), 64–71 (2008). DOI 10.1038/nature07242
6. van Bakel, H., Nislow, C., Blencowe, B.J., Hughes, T.R.: Most "dark matter" transcripts are associated with known genes. PLoS Biol **8**(5), e1000,371 (2010). DOI 10.1371/journal.pbio.1000371
7. van Bakel, H., Nislow, C., Blencowe, B.J., Hughes, T.R.: Response to 'the reality of pervasive transcription'. PLoS Biology **9**(7), e1001,102 (2011)
8. Banfield, W.G., Woke, P.A., MacKay, C.M., Cooper, H.L.: Mosquito transmission of a reticulum cell sarcoma of hamsters. Science **148**(3674), 1239–1240 (1965)
9. Bell, G., Mooers, A.O.: Size and complexity among multicellular organisms. Biological Journal of the Linnean Society **60**(3), 345–363 (1997)
10. Benton, M.J., Ayala, F.J.: Dating the tree of life. Science **300**(5626), 1698–700 (2003). DOI 10.1126/science.1077795
11. Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., Hannon, G.J.: Dicer is essential for mouse development. Nat Genet **35**(3), 215–7 (2003). DOI 10.1038/ng1253
12. Berretta, J., Morillon, A.: Pervasive transcription constitutes a new level of eukaryotic genome regulation. EMBO Rep **10**(9), 973–82 (2009). DOI 10.1038/embor.2009.181
13. Bistis, G.N., Perkins, D.D., Read, N.D.: Different cell types in neurospora crassa. Fungal Genetics Newsletter pp. 17–19 (2003)
14. Blackstone, N.W.: Redox control and the evolution of multicellularity. Bioessays **22**(10), 947–53 (2000). DOI 10.1002/1521-1878(200010)22:10<947::AID-BIES10>3.0.CO;2-W
15. Bocobza, S.E., Aharoni, A.: Switching the light on plant riboswitches. Trends Plant Sci **13**(10), 526–33 (2008). DOI 10.1016/j.tplants.2008.07.004
16. Bonner, J.T.: The origins of multicellularity. Integrative Biology Issues News and Reviews **1**(1), 27–36 (1998)
17. Bonner, J.T.: Perspective: the size-complexity rule. Evolution **58**(9), 1883–90 (2004)
18. Branda, S.S., Gonzlez-Pastor, J.E., Ben-Yehuda, S., Losick, R., Kolter, R.: Fruiting body formation by bacillus subtilis. Proc Natl Acad Sci U S A **98**(20), 11,621–6 (2001). DOI 10.1073/pnas.191384198

19. Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., Rinn, J.L.: Integrative annotation of human large intergenic noncoding rnas reveals global properties and specific subclasses. Genes Dev **25**(18), 1915–27 (2011). DOI 10.1101/gad.17446611

20. Carroll, S.B.: Chance and necessity: the evolution of morphological complexity and diversity. Nature **409**(6823), 1102–9 (2001). DOI 10.1038/35059227

21. Cases, I., de Lorenzo, V., Ouzounis, C.A.: Transcription regulation and environmental adaptation in bacteria. Trends Microbiol **11**(6), 248–53 (2003)

22. Cheah, M.T., Wachter, A., Sudarsan, N., Breaker, R.R.: Control of alternative rna splicing and gene expression by eukaryotic riboswitches. Nature **447**(7143), 497–500 (2007). DOI 10.1038/nature05769

23. Chen, C.L., Zhou, H., Liao, J.Y., Qu, L.H., Amar, L.: Genome-wide evolutionary analysis of the noncoding rna genes and noncoding dna of paramecium tetraurelia. RNA **15**(4), 503–14 (2009). DOI 10.1261/rna.1306009

24. Chen, K., Rajewsky, N.: The evolution of gene regulation by transcription factors and micrornas. Nat Rev Genet **8**(2), 93–103 (2007). DOI 10.1038/nrg1990

25. Christodoulou, F., Raible, F., Tomer, R., Simakov, O., Trachana, K., Klaus, S., Snyman, H., Hannon, G.J., Bork, P., Arendt, D.: Ancient animal micrornas and the evolution of tissue identity. Nature **463**(7284), 1084–8 (2010). DOI 10.1038/nature08744

26. Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., Chang, H.Y.: Genomic maps of long noncoding rna occupancy reveal principles of rna-chromatin interactions. Mol Cell **44**(4), 667–78 (2011). DOI 10.1016/j.molcel.2011.08.027

27. Clark, M.B., Amaral, P.P., Schlesinger, F.J., Dinger, M.E., Taft, R.J., Rinn, J.L., Ponting, C.P., Stadler, P.F., Morris, K.V., Morillon, A., Rozowsky, J.S., Gerstein, M.B., Wahlestedt, C., Hayashizaki, Y., Carninci, P., Gingeras, T.R., Mattick, J.S.: The reality of pervasive transcription. PLoS Biol **9**(7), e1000,625; discussion e1001,102 (2011). DOI 10.1371/journal.pbio.1000625

28. Cobb, B.S., Nesterova, T.B., Thompson, E., Hertweck, A., O'Connor, E., Godwin, J., Wilson, C.B., Brockdorff, N., Fisher, A.G., Smale, S.T., Merkenschlager, M.: T cell lineage choice and differentiation in the absence of the rnase iii enzyme dicer. J Exp Med **201**(9), 1367–73 (2005). DOI 10.1084/jem.20050572

29. Cock, J.M., Sterck, L., Rouz, P., Scornet, D., Allen, A.E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J.M., Badger, J.H., Beszteri, B., Billiau, K., Bonnet, E., Bothwell, J.H., Bowler, C., Boyen, C., Brownlee, C., Carrano, C.J., Charrier, B., Cho, G.Y., Coelho, S.M., Colln, J., Corre, E., Da Silva, C., Delage, L., Delaroque, N., Dittami, S.M., Doulbeau, S., Elias, M., Farnham, G., Gachon, C.M., Gschloessl, B., Heesch, S., Jabbari, K., Jubin, C., Kawai, H., Kimura, K., Kloareg, B., Kpper, F.C., Lang, D., Le Bail, A., Leblanc, C., Lerouge, P., Lohr, M., Lopez, P.J., Martens, C., Maumus, F., Michel, G., Miranda-Saavedra, D., Morales, J., Moreau, H., Motomura, T., Nagasato, C., Napoli, C.A., Nelson, D.R., Nyvall-Colln, P., Peters, A.F., Pommier, C., Potin, P., Poulain, J., Quesneville, H., Read, B., Rensing, S.A., Ritter, A., Rousvoal, S., Samanta, M., Samson, G., Schroeder, D.C., Sgurens, B., Strittmatter, M., Tonon, T., Tregear, J.W., Valentin, K., von Dassow, P., Yamagishi, T., Van de Peer, Y., Wincker, P.: The ectocarpus genome and the independent evolution of multicellularity in brown algae. Nature **465**(7298), 617–21 (2010). DOI 10.1038/nature09016

30. Condorelli, G., Dimmeler, S.: Micrornas: components of an integrated system controlling cardiac development, physiology, and disease pathogenesis. Cardiovasc Res **79**(4), 551–2 (2008). DOI 10.1093/cvr/cvn189

31. Costa, F.F.: Non-coding rnas: new players in eukaryotic biology. Gene **357**(2), 83–94 (2005). DOI 10.1016/j.gene.2005.06.019

32. Cuperus, J.T., Fahlgren, N., Carrington, J.C.: Evolution and functional diversification of mirna genes. Plant Cell **23**(2), 431–42 (2011). DOI 10.1105/tpc.110.082784

33. DeLong, J.P., Okie, J.G., Moses, M.E., Sibly, R.M., Brown, J.H.: Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. Proc Natl Acad Sci U S A **107**(29), 12,941–5 (2010). DOI 10.1073/pnas.1007783107

34. Duret, L., Chureau, C., Samain, S., Weissenbach, J., Avner, P.: The xist rna gene evolved in eutherians by pseudogenization of a protein-coding gene. Science **312**(5780), 1653–5 (2006). DOI 10.1126/science.1126316

35. Erwin, D.H.: Early origin of the bilaterian developmental toolkit. Philos Trans R Soc Lond B Biol Sci **364**(1527), 2253–61 (2009). DOI 10.1098/rstb.2009.0038

36. Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangl, J.L., Carrington, J.C.: High-throughput

sequencing of arabidopsis micrornas: evidence for frequent birth and death of mirna genes. PLoS One **2**(2), e219 (2007). DOI 10.1371/journal.pone.0000219

37. Falkowski, P.G., Fenchel, T., Delong, E.F.: The microbial engines that drive earth's biogeochemical cycles. Science **320**(5879), 1034–9 (2008). DOI 10.1126/science.1153213

38. Floyd, S.K., Bowman, J.L.: The ancestral developmental tool kit of land plants. International Journal of Plant Sciences **168**(1), 1–35 (2007)

39. Fontana, W., Schuster, P.: Continuity in evolution: on the nature of transitions. Science **280**(5368), 1451–5 (1998)

40. Gangaraju, V.K., Lin, H.: Micrornas: key regulators of stem cells. Nat Rev Mol Cell Biol **10**(2), 116–25 (2009). DOI 10.1038/nrm2621

41. Giraldez, A.J., Cinalli, R.M., Glasner, M.E., Enright, A.J., Thomson, J.M., Baskerville, S., Hammond, S.M., Bartel, D.P., Schier, A.F.: Micrornas regulate brain morphogenesis in zebrafish. Science **308**(5723), 833–8 (2005). DOI 10.1126/science.1109020

42. Gregory, T.R.: Synergy between sequence and size in large-scale genomics. Nat Rev Genet **6**(9), 699–708 (2005). DOI 10.1038/nrg1674

43. Grosberg, R.K., Strathmann, R.R.: The evolution of multicellularity: a minor major transition? Annu. Rev. Ecol. Evol. Syst. **38**, 621–654 (2007)

44. Gruber, A.R., Kilgus, C., Mosig, A., Hofacker, I.L., Hennig, W., Stadler, P.F.: Arthropod 7sk rna. Mol Biol Evol **25**(9), 1923–30 (2008). DOI 10.1093/molbev/msn140

45. Guo, H., Ingolia, N.T., Weissman, J.S., Bartel, D.P.: Mammalian micrornas predominantly act to decrease target mrna levels. Nature **466**(7308), 835–40 (2010). DOI 10.1038/nature09267

46. Guo, X., Zhang, Z., Gerstein, M.B., Zheng, D.: Small rnas originated from pseudogenes: cis- or trans-acting? PLoS Comput Biol **5**(7), e1000,449 (2009). DOI 10.1371/journal.pcbi.1000449

47. Hampl, V., Hug, L., Leigh, J.W., Dacks, J.B., Lang, B.F., Simpson, A.G., Roger, A.J.: Phylogenomic analyses support the monophyly of excavata and resolve relationships among eukaryotic "supergroups". Proc Natl Acad Sci U S A **106**(10), 3859–64 (2009). DOI 10.1073/pnas.0807880106

48. Harfe, B.D., McManus, M.T., Mansfield, J.H., Hornstein, E., Tabin, C.J.: The rnaseiii enzyme dicer is required for morphogenesis but not patterning of the vertebrate limb. Proc Natl Acad Sci U S A **102**(31), 10,898–903 (2005). DOI 10.1073/pnas.0504834102

49. Hedges, S.B.: The origin and evolution of model organisms. Nat Rev Genet **3**(11), 838–49 (2002). DOI 10.1038/nrg929

50. Hedges, S.B., Blair, J.E., Venturi, M.L., Shoe, J.L.: A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol Biol **4**, 2 (2004). DOI 10.1186/1471-2148-4-2

51. Heimberg, A.M., Sempere, L.F., Moy, V.N., Donoghue, P.C., Peterson, K.J.: Micrornas and the advent of vertebrate morphological complexity. Proc Natl Acad Sci U S A **105**(8), 2946–50 (2008). DOI 10.1073/pnas.0712259105

52. Herron, M.D., Hackett, J.D., Aylward, F.O., Michod, R.E.: Triassic origin and early radiation of multicellular volvocine algae. Proc Natl Acad Sci U S A **106**(9), 3254–8 (2009). DOI 10.1073/pnas.0811205106

53. Hiller, M., Findeiss, S., Lein, S., Marz, M., Nickel, C., Rose, D., Schulz, C., Backofen, R., Prohaska, S.J., Reuter, G., Stadler, P.F.: Conserved introns reveal novel transcripts in drosophila melanogaster. Genome Res **19**(7), 1289–300 (2009). DOI 10.1101/gr.090050.108

54. Holland, H.D.: The oxygenation of the atmosphere and oceans. Philos Trans R Soc Lond B Biol Sci **361**(1470), 903–15 (2006). DOI 10.1098/rstb.2006.1838

55. Hornstein, E., Shomron, N.: Canalization of development by micrornas. Nat Genet **38 Suppl**, S20–4 (2006). DOI 10.1038/ng1803

56. Huynen, M.A.: Exploring phenotype space through neutral evolution. J Mol Evol **43**(3), 165–9 (1996)

57. Huynen, M.A., Stadler, P.F., Fontana, W.: Smoothness within ruggedness: the role of neutrality in adaptation. Proc Natl Acad Sci U S A **93**(1), 397–401 (1996)

58. Jacquier, A.: The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small rnas. Nat Rev Genet **10**(12), 833–44 (2009). DOI 10.1038/nrg2683

59. Joyce, G.F.: The antiquity of rna-based evolution. Nature **418**(6894), 214–21 (2002). DOI 10.1038/418214a

60. Kaiser, D.: Building a multicellular organism. Annu Rev Genet **35**, 103–23 (2001). DOI 10.1146/annurev.genet.35.102401.090145

61. Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermller, J., Hofacker, I.L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., Tammana, H., Gingeras, T.R.: Rna maps reveal new rna classes and a possible function for pervasive transcription. Science **316**(5830), 1484–8 (2007). DOI 10.1126/science.1138341

62. Kapranov, P., St Laurent, G., Raz, T., Ozsolak, F., Reynolds, C.P., Sorensen, P.H., Reaman, G., Milos, P., Arceci, R.J., Thompson, J.F., Triche, T.J.: The majority of total nuclear-encoded non-ribosomal rna in a human cell is 'dark matter' un-annotated rna. BMC Biol **8**, 149 (2010). DOI 10.1186/1741-7007-8-149

63. Kazazian Jr, H.H.: Mobile elements: drivers of genome evolution. Science **303**(5664), 1626–32 (2004). DOI 10.1126/science.1089670

64. Kempes, C.P., Dutkiewicz, S., Follows, M.J.: Growth, metabolic partitioning, and the size of microorganisms. Proc Natl Acad Sci U S A **109**(2), 495–500 (2012). DOI 10.1073/pnas.1115585109

65. Kim, E.D., Sung, S.: Long noncoding rna: unveiling hidden layer of gene regulatory networks. Trends Plant Sci **17**(1), 16–21 (2012). DOI 10.1016/j.tplants.2011.10.008

66. Kim, V.N., Han, J., Siomi, M.C.: Biogenesis of small rnas in animals. Nat Rev Mol Cell Biol **10**(2), 126–39 (2009). DOI 10.1038/nrm2632

67. King, N.: The unicellular ancestry of animal development. Dev Cell **7**(3), 313–25 (2004). DOI 10.1016/j.devcel.2004.08.010

68. King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K.J., Zuzow, R., Dirks, W., Good, M., Goodstein, D., Lemons, D., Li, W., Lyons, J.B., Morris, A., Nichols, S., Richter, D.J., Salamov, A., Sequencing, J.G., Bork, P., Lim, W.A., Manning, G., Miller, W.T., McGinnis, W., Shapiro, H., Tjian, R., Grigoriev, I.V., Rokhsar, D.: The genome of the choanoflagellate monosiga brevicollis and the origin of metazoans. Nature **451**(7180), 783–8 (2008). DOI 10.1038/nature06617

69. Kishore, S., Stamm, S.: Regulation of alternative splicing by snornas. Cold Spring Harb Symp Quant Biol **71**, 329–34 (2006). DOI 10.1101/sqb.2006.71.024

70. Knoll, A.H.: The multiple origins of complex multicellularity. Annual Review of Earth and Planetary Sciences **39**, 217–239 (2011)

71. Konstantinidis, K.T., Tiedje, J.M.: Trends between gene content and genome size in prokaryotic species with larger genomes. Proc Natl Acad Sci U S A **101**(9), 3160–5 (2004). DOI 10.1073/pnas.0308653100

72. Koonin, E.V., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Krylov, D.M., Makarova, K.S., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Rogozin, I.B., Smirnov, S., Sorokin, A.V., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J., Natale, D.A.: A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol **5**(2), R7 (2004). DOI 10.1186/gb-2004-5-2-r7

73. Kutter, C., Schb, H., Stadler, M., Meins Jr, F., Si-Ammour, A.: Microrna-mediated regulation of stomatal development in arabidopsis. Plant Cell **19**(8), 2417–29 (2007). DOI 10.1105/tpc.107.050377

74. Lane, N., Martin, W.: The energetics of genome complexity. Nature **467**(7318), 929–34 (2010). DOI 10.1038/nature09486

75. Lee, H.C., Li, L., Gu, W., Xue, Z., Crosthwaite, S.K., Pertsemlidis, A., Lewis, Z.A., Freitag, M., Selker, E.U., Mello, C.C., Liu, Y.: Diverse pathways generate microrna-like rnas and dicer-independent small interfering rnas in fungi. Mol Cell **38**(6), 803–14 (2010). DOI 10.1016/j.molcel.2010.04.005

76. Lesser, M.P.: Oxidative stress in marine environments: biochemistry and physiological ecology. Annu Rev Physiol **68**, 253–78 (2006). DOI 10.1146/annurev.physiol.68.040104.110001

77. Li, L., Wang, X., Stolc, V., Li, X., Zhang, D., Su, N., Tongprasit, W., Li, S., Cheng, Z., Wang, J., Deng, X.W.: Genome-wide transcription analyses in rice using tiling microarrays. Nat Genet **38**(1), 124–9 (2006). DOI 10.1038/ng1704

78. Liu, N., Okamura, K., Tyler, D.M., Phillips, M.D., Chung, W.J., Lai, E.C.: The evolution and functional diversification of animal microrna genes. Cell Res **18**(10), 985–96 (2008). DOI 10.1038/cr.2008.278

79. Lozada-Chvez, I., Angarica, V.E., Collado-Vides, J., Contreras-Moreira, B.: The role of dna-binding specificity in the evolution of bacterial regulatory networks. J Mol Biol **379**(3), 627–43 (2008). DOI 10.1016/j.jmb.2008.04.008

80. Lozada-Chvez, I., Janga, S.C., Collado-Vides, J.: Bacterial regulatory networks are extremely flexible in evolution. Nucleic Acids Res **34**(12), 3434–45 (2006). DOI 10.1093/nar/gkl423

81. Lu, J., Shen, Y., Wu, Q., Kumar, S., He, B., Shi, S., Carthew, R.W., Wang, S.M., Wu, C.I.: The birth and death of microrna genes in drosophila. Nat Genet **40**(3), 351–5 (2008). DOI 10.1038/ng.73

82. Lürling, M., Van Donk, E.: Grazer-induced colony formation in scenedesmus acutus (chlorophyceae): ecomorph expression at different temperatures. Journal of Phycology **35**(6), 1120–1126 (1999)

83. Lynch, M.: The origins of eukaryotic gene structure. Mol Biol Evol **23**(2), 450–68 (2006). DOI 10.1093/molbev/msj050

84. Lynch, M., Bobay, L.M., Catania, F., Gout, J.F., Rho, M.: The repatterning of eukaryotic genomes by random genetic drift. Annu Rev Genomics Hum Genet **12**, 347–66 (2011). DOI 10.1146/annurev-genom-082410-101412

85. Lynch, M., Conery, J.S.: The origins of genome complexity. Science **302**(5649), 1401–4 (2003). DOI 10.1126/science.1089370

86. Madan Babu, M., Teichmann, S.A., Aravind, L.: Evolutionary dynamics of prokaryotic transcriptional regulatory networks. J Mol Biol **358**(2), 614–33 (2006). DOI 10.1016/j.jmb.2006.02.019

87. Marques, A.C., Ponting, C.P.: Catalogues of mammalian long noncoding rnas: modest conservation and incompleteness. Genome Biol **10**(11), R124 (2009). DOI 10.1186/gb-2009-10-11-r124

88. Martnez-Antonio, A., Collado-Vides, J.: Identifying global regulators in transcriptional regulatory networks in bacteria. Curr Opin Microbiol **6**(5), 482–9 (2003)

89. McCarthy, M.C., Enquist, B.J.: Organismal size, metabolism and the evolution of complexity in metazoans. Evolutionary Ecology Research **7**(5), 681 (2005)

90. de Meaux, J., Hu, J.Y., Tartler, U., Goebel, U.: Structurally different alleles of the ath-mir824 microrna precursor are maintained at high frequency in arabidopsis thaliana. Proc Natl Acad Sci U S A **105**(26), 8994–9 (2008). DOI 10.1073/pnas.0803218105

91. Medina, M., Collins, A.G., Taylor, J.W., Valentine, J.W., Lipps, J.H., Amaral-Zettler, L., Sogin, M.L.: Phylogeny of opisthokonta and the evolution of multicellularity and complexity in fungi and metazoa. International Journal of Astrobiology **2**(3), 203–211 (2003)

92. Mercer, T.R., Dinger, M.E., Mattick, J.S.: Long non-coding rnas: insights into functions. Nat Rev Genet **10**(3), 155–9 (2009). DOI 10.1038/nrg2521

93. Millar, A.A., Waterhouse, P.M.: Plant and animal micrornas: similarities and differences. Funct Integr Genomics **5**(3), 129–35 (2005). DOI 10.1007/s10142-005-0145-2

94. Mosig, A., Zhu, L., Stadler, P.F.: Customized strategies for discovering distant ncrna homologs. Brief Funct Genomic Proteomic **8**(6), 451–60 (2009). DOI 10.1093/bfgp/elp035

95. Mourier, T., Carret, C., Kyes, S., Christodoulou, Z., Gardner, P.P., Jeffares, D.C., Pinches, R., Barrell, B., Berriman, M., Griffiths-Jones, S., Ivens, A., Newbold, C., Pain, A.: Genome-wide discovery and verification of novel structured rnas in plasmodium falciparum. Genome Res **18**(2), 281–92 (2008). DOI 10.1101/gr.6836108

96. Mu, X.J., Lu, Z.J., Kong, Y., Lam, H.Y., Gerstein, M.B.: Analysis of genomic variation in non-coding elements using population-scale sequencing data from the 1000 genomes project. Nucleic Acids Res **39**(16), 7058–76 (2011). DOI 10.1093/nar/gkr342

97. Murgia, C., Pritchard, J.K., Kim, S.Y., Fassati, A., Weiss, R.A.: Clonal origin and evolution of a transmissible cancer. Cell **126**(3), 477–87 (2006). DOI 10.1016/j.cell.2006.05.051

98. Nedelcu, A.M.: The evolution of self during the transition to multicellularity. Adv Exp Med Biol **738**, 14–30 (2012). DOI 10.1007/978-1-4614-1680-7_2

99. Niklas, K.J.: The evolution of plant body plansa biomechanical perspective. Annals of Botany **85**(4), 411–438 (2000)

100. Nilsen, T.W., Graveley, B.R.: Expansion of the eukaryotic proteome by alternative splicing. Nature **463**(7280), 457–63 (2010). DOI 10.1038/nature08909

101. Nodine, M.D., Bartel, D.P.: Micrornas prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes Dev **24**(23), 2678–92 (2010). DOI 10.1101/gad.1986710

102. Ochman, H., Davalos, L.M.: The nature and dynamics of bacterial genomes. Science **311**(5768), 1730–3 (2006). DOI 10.1126/science.1119966

103. Ohta, T.: Slightly deleterious mutant substitutions in evolution. Nature **246**(5428), 96–8 (1973)

104. Ohta, T.: The nearly neutral theory of molecular evolution. Annual Review of Ecology and Systematics **23**, 263–286 (1992)

105. Pauli, A., Rinn, J.L., Schier, A.F.: Non-coding rnas as regulators of embryogenesis. Nat Rev Genet **12**(2), 136–49 (2011). DOI 10.1038/nrg2904

106. Payne, J.L., Boyer, A.G., Brown, J.H., Finnegan, S., Kowalewski, M., Krause Jr, R.A., Lyons, S.K., McClain, C.R., McShea, D.W., Novack-Gottshall, P.M., Smith, F.A., Stempien, J.A., Wang, S.C.: Two-phase increase in the maximum size of life over 3.5 billion years reflects biological innovation and environmental opportunity. Proc Natl Acad Sci U S A **106**(1), 24–7 (2009). DOI 10.1073/pnas.0806314106

107. Pearse, A.M., Swift, K.: Allograft theory: transmission of devil facial-tumour disease. Nature **439**(7076), 549 (2006). DOI 10.1038/439549a

108. Peterlin, B.M., Brogie, J.E., Price, D.H.: 7sk snrna: a noncoding rna that plays a major role in regulating eukaryotic transcription. Wiley Interdiscip Rev RNA **3**(1), 92–103 (2012). DOI 10.1002/wrna.106

109. Peterson, K.J., Dietrich, M.R., McPeek, M.A.: Micrornas and metazoan macroevolution: insights into canalization, complexity, and the cambrian explosion. Bioessays **31**(7), 736–47 (2009). DOI 10.1002/bies.200900033

110. Ponjavic, J., Ponting, C.P., Lunter, G.: Functionality or transcriptional noise? evidence for selection within long noncoding rnas. Genome Res **17**(5), 556–65 (2007). DOI 10.1101/gr.6036807

111. Ponting, C.P., Oliver, P.L., Reik, W.: Evolution and functions of long noncoding rnas. Cell **136**(4), 629–41 (2009). DOI 10.1016/j.cell.2009.02.006

112. Prasanth, K.V., Spector, D.L.: Eukaryotic regulatory rnas: an answer to the 'genome complexity' conundrum. Genes Dev **21**(1), 11–42 (2007). DOI 10.1101/gad.1484207

113. Ranea, J.A., Grant, A., Thornton, J.M., Orengo, C.A.: Microeconomic principles explain an optimal genome size in bacteria. Trends Genet **21**(1), 21–5 (2005). DOI 10.1016/j.tig.2004.11.014

114. Ratcliff, W.C., Denison, R.F., Borrello, M., Travisano, M.: Experimental evolution of multicellularity. Proc Natl Acad Sci U S A **109**(5), 1595–600 (2012). DOI 10.1073/pnas.1115323109

115. Ren, B.: Transcription: Enhancers make non-coding rna. Nature **465**(7295), 173–4 (2010). DOI 10.1038/465173a

116. Repoila, F., Darfeuille, F.: Small regulatory non-coding rnas in bacteria: physiology and mechanistic aspects. Biol Cell **101**(2), 117–31 (2009). DOI 10.1042/BC20070137

117. Robertson, M.P., Joyce, G.F.: The origins of the rna world. Cold Spring Harb Perspect Biol **4**(5) (2012). DOI 10.1101/cshperspect.a003608

118. Rokas, A.: The origins of multicellularity and the early history of the genetic toolkit for animal development. Annu Rev Genet **42**, 235–51 (2008). DOI 10.1146/annurev.genet.42.110807.091513

119. Rose, D., Hiller, M., Schutt, K., Hackermller, J., Backofen, R., Stadler, P.F.: Computational discovery of human coding and non-coding transcripts with conserved splice sites. Bioinformatics **27**(14), 1894–900 (2011). DOI 10.1093/bioinformatics/btr314

120. Roush, S., Slack, F.J.: The let-7 family of micrornas. Trends Cell Biol **18**(10), 505–16 (2008). DOI 10.1016/j.tcb.2008.07.007

121. Ruiz-Trillo, I., Burger, G., Holland, P.W., King, N., Lang, B.F., Roger, A.J., Gray, M.W.: The origins of multicellularity: a multi-taxon genome initiative. Trends Genet **23**(3), 113–8 (2007). DOI 10.1016/j.tig.2007.01.005

122. Schuster, P., Fontana, W., Stadler, P.F., Hofacker, I.L.: From sequences to shapes and back: a case study in rna secondary structures. Proc Biol Sci **255**(1344), 279–84 (1994). DOI 10.1098/rspb.1994.0040

123. Scott, M.S., Ono, M.: From snorna to mirna: Dual function regulatory non-coding rnas. Biochimie **93**(11), 1987–92 (2011). DOI 10.1016/j.biochi.2011.05.026

124. Selbach, M., Schwanhusser, B., Thierfelder, N., Fang, Z., Khanin, R., Rajewsky, N.: Widespread changes in protein synthesis induced by micrornas. Nature **455**(7209), 58–63 (2008). DOI 10.1038/nature07228

125. Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermller, J., Reinhardt, R., Stadler, P.F., Vogel, J.: The primary transcriptome of the major human pathogen helicobacter pylori. Nature **464**(7286), 250–5 (2010). DOI 10.1038/nature08756

126. Specht, C.D., Bartlett, M.E.: Flower evolution: the origin and subsequent diversification of the angiosperm flower. Annu. Rev. Ecol. Evol. Syst. **40**, 217–243 (2009)

127. Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N.H., Stanke, M., Adamska, M., Darling, A., Degnan, S.M., Oakley, T.H., Plachetzki, D.C., Zhai, Y., Adamski, M., Calcino, A., Cummins, S.F., Goodstein, D.M., Harris, C., Jackson, D.J., Leys, S.P., Shu, S., Woodcroft, B.J., Vervoort, M., Kosik, K.S., Manning, G., Degnan, B.M., Rokhsar, D.S.: The amphimedon queenslandica genome and the evolution of animal complexity. Nature **466**(7307), 720–6 (2010). DOI 10.1038/nature09201
128. Strathmann, R.R.: From metazoan to protist via competition among cell lineages. Evol. Theory **10**, 67–70 (1991)
129. Sudarsan, N., Barrick, J.E., Breaker, R.R.: Metabolite-binding rna domains are present in the genes of eukaryotes. RNA **9**(6), 644–7 (2003)
130. Taft, R.J., Pheasant, M., Mattick, J.S.: The relationship between non-protein-coding dna and eukaryotic complexity. Bioessays **29**(3), 288–99 (2007). DOI 10.1002/bies.20544
131. Tisseur, M., Kwapisz, M., Morillon, A.: Pervasive transcription - lessons from yeast. Biochimie **93**(11), 1889–96 (2011). DOI 10.1016/j.biochi.2011.07.001
132. Tomitani, A., Knoll, A.H., Cavanaugh, C.M., Ohno, T.: The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. Proc Natl Acad Sci U S A **103**(14), 5442–7 (2006). DOI 10.1073/pnas.0600999103
133. Valentine, J.W., Collins, A.G., Meyer, C.P.: Morphological complexity increase in metazoans. Paleobiology pp. 131–142 (1994)
134. Velicer, G.J., Kroos, L., Lenski, R.E.: Loss of social behaviors by myxococcus xanthus during evolution in an unstructured habitat. Proc Natl Acad Sci U S A **95**(21), 12,376–80 (1998)
135. Voinnet, O.: Origin, biogenesis, and activity of plant micrornas. Cell **136**(4), 669–87 (2009). DOI 10.1016/j.cell.2009.01.046
136. Wagner, A.: Energy constraints on the evolution of gene expression. Mol Biol Evol **22**(6), 1365–74 (2005). DOI 10.1093/molbev/msi126
137. Waters, L.S., Storz, G.: Regulatory rnas in bacteria. Cell **136**(4), 615–28 (2009). DOI 10.1016/j.cell.2009.01.043
138. Wilusz, J.E., Sunwoo, H., Spector, D.L.: Long noncoding rnas: functional surprises from the rna world. Genes Dev **23**(13), 1494–504 (2009). DOI 10.1101/gad.1800909
139. Wolpert, L., Szathmry, E.: Multicellularity: evolution and the egg. Nature **420**(6917), 745 (2002). DOI 10.1038/420745a
140. Yang, Z., Kong, F., Yang, Z., Zhang, M., Yu, Y., Qian, S.: Benefits and costs of the grazer-induced colony formation in microcystis aeruginosa. Annales de Limnologie–International Journal of Limnology **45**, 203–208 (2009)
141. Yano, Y., Saito, R., Yoshida, N., Yoshiki, A., Wynshaw-Boris, A., Tomita, M., Hirotsune, S.: A new role for expressed pseudogenes as ncrna: regulation of mrna stability of its homologous coding gene. J Mol Med (Berl) **82**(7), 414–22 (2004). DOI 10.1007/s00109-004-0550-3
142. Zhao, Y., Ransom, J.F., Li, A., Vedantham, V., von Drehle, M., Muth, A.N., Tsuchihashi, T., McManus, M.T., Schwartz, R.J., Srivastava, D.: Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking mirna-1-2. Cell **129**(2), 303–17 (2007). DOI 10.1016/j.cell.2007.03.030
143. Zheng, D., Frankish, A., Baertsch, R., Kapranov, P., Reymond, A., Choo, S.W., Lu, Y., Denoeud, F., Antonarakis, S.E., Snyder, M., Ruan, Y., Wei, C.L., Gingeras, T.R., Guig, R., Harrow, J., Gerstein, M.B.: Pseudogenes in the encode regions: consensus annotation, analysis of transcription, and evolution. Genome Res **17**(6), 839–51 (2007). DOI 10.1101/gr.5586307
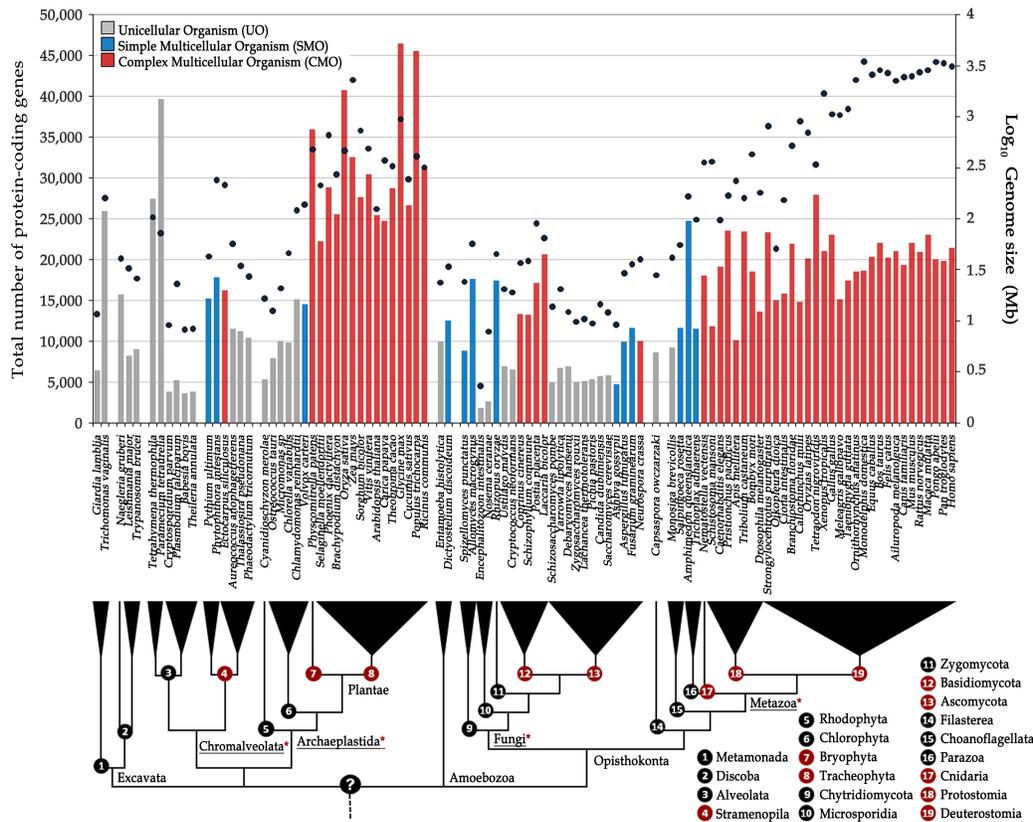
Fig. 1 Multiple origins of CM and C-value and G-value paradoxes among the 100 Eukarya with complete genome sequences. The phylogenetic relationships for higher taxa were obtained from [49] and [47]; main eukaryotic clades with independent origins of CM are underlined and marked with *. Red circles indicate taxa harboring at least one CMO with a complete sequence genome. There is no clear relationship between multicellularity and genome size, although CMOs tend to have larger genomes (black dots). This observation is, however, at least in part biased by the overrepresentation of vertebrates and flowering plants. Comparable proportions of protein-coding genes (bars) are found among unicellular species, SMOs and CMOs. The total number of nuclear protein-coding genes and the non-organellar genome size are depicted in Mega base (Mb) pairs for each genome project (with 7X or greater fold coverage), and were obtained from the reports of their latest publication and updated databases; when necessary, the gene content was corrected by counting the coding sequences for proteins (CDS) from the corresponding annotation of their best gene models.
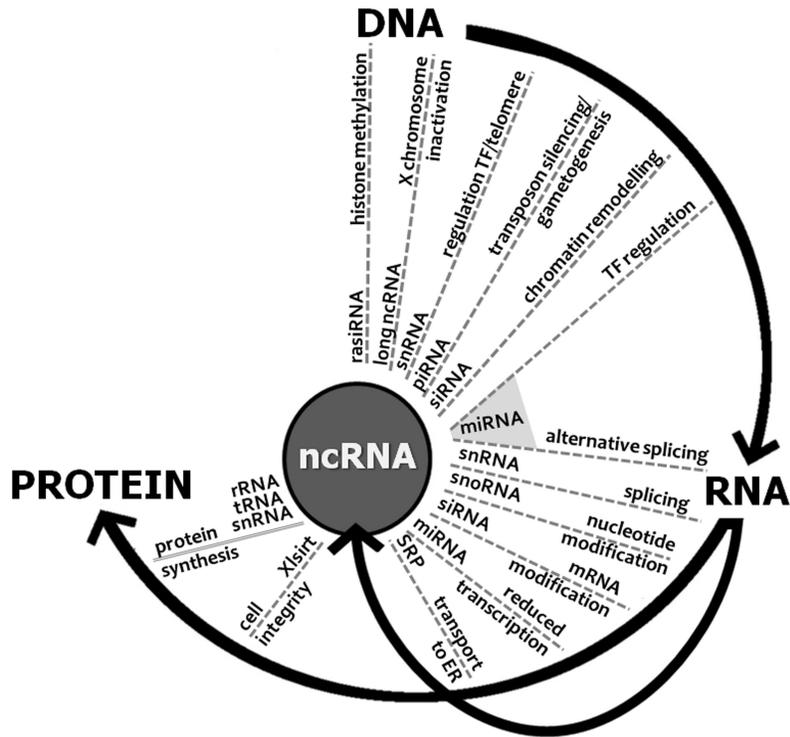
Fig. 2 Regulatory functions of ncRNAs are involved in all the stages of the central dogma of biology. Genetic information flows from DNA (the genotype) to RNA (with thick black lines). RNA (the phenotype) then decodes proteins (from a mRNA) and/or ncRNAs (from a transcript or a mRNA). Regulatory ncRNAs control gene expression (shown in dashed lines), whereas structural ncRNAs (e.g., ribosomal and transfer RNAs) are involved in protein synthesis (shown in thin doubled-line). ncRNA symbology: mi, micro; nc, non-coding; pi, piwi; rasi, repeat-associated small interfering; si, short interfering; sn, small nuclear; sno, small nucleolar; r, ribosomal; t, transfer; SRP, signal recognition particle; TF, transcription factor. Figure is modified with kind permission from [30].
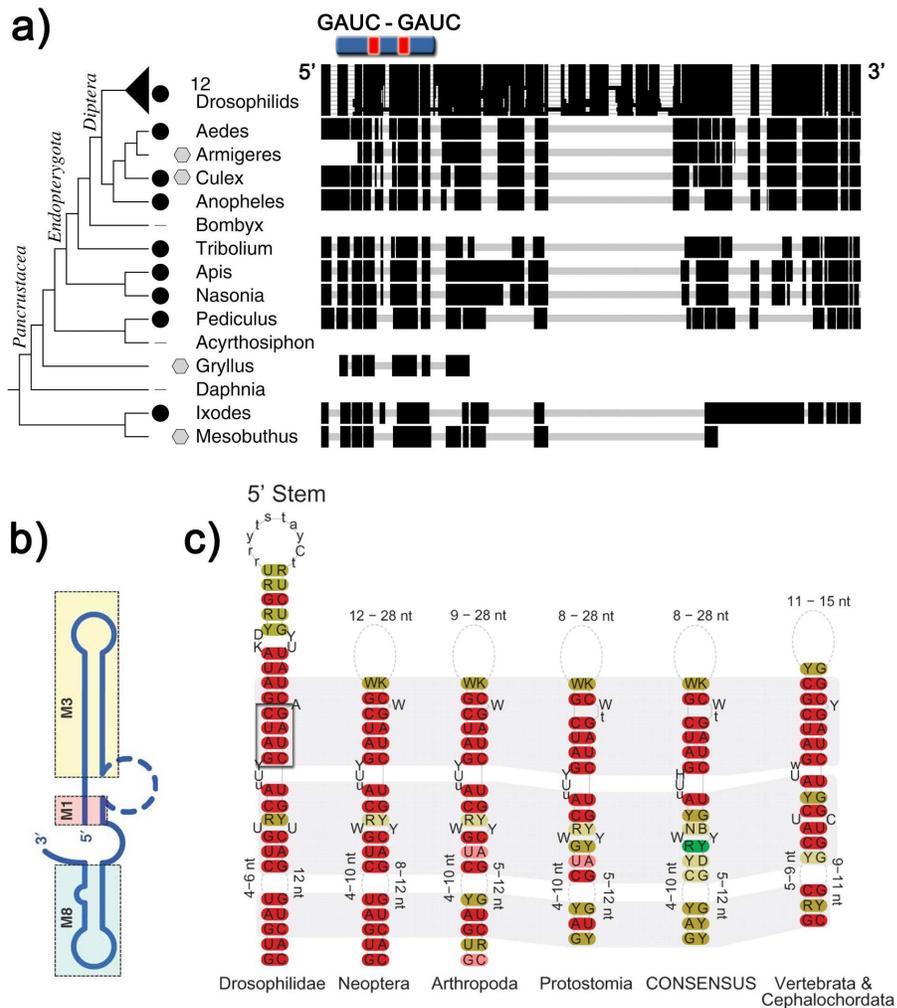
Fig. 3 Comparison of the 5' stems of the regulatory 7SK small nuclear RNAs. **a)** Phylogenetic distribution of 7SK candidate sequences in Arthropoda. Figure a) is adapted with kind permission of [44]. A black dot indicates a match in the genomic sequence; the hexagons refer to partial ESTs. Aligned blocks are shown in black, gray bars indicate gaps in the alignment, and missing sequence data adjacent to EST regions appear white. **b)** Simplified cartoon model of the functional secondary structure of 7SK RNA (based on [108]) showing the position of the M3 stem located at the 5' stem region. The position of the M3 stem region is shown by a blue block at the top of the alignment in **a)** highlighting the highly conserved motif GAUC-GAUC. **c)** Structural models for the M3 stems in several eumetazoan clades. The consensus structural models for Drosophilidae, Neoptera and Arthropoda are based on the alignment in **a)**. Conserved regions are colored in red; two and three compensatory mutations are shown in ochre and green. Lower case letters denote a deletion in some sequences. Helices are highlighted by a gray background. The open rectangle at the M3 stem for Drosophilidae shows the highly conserved motif GAUC-GAUC. Hairpin loops (with variable size and no clear consensus folds) are drawn as dashed ellipses. Figure c) is reproduced with kind permission of [94].
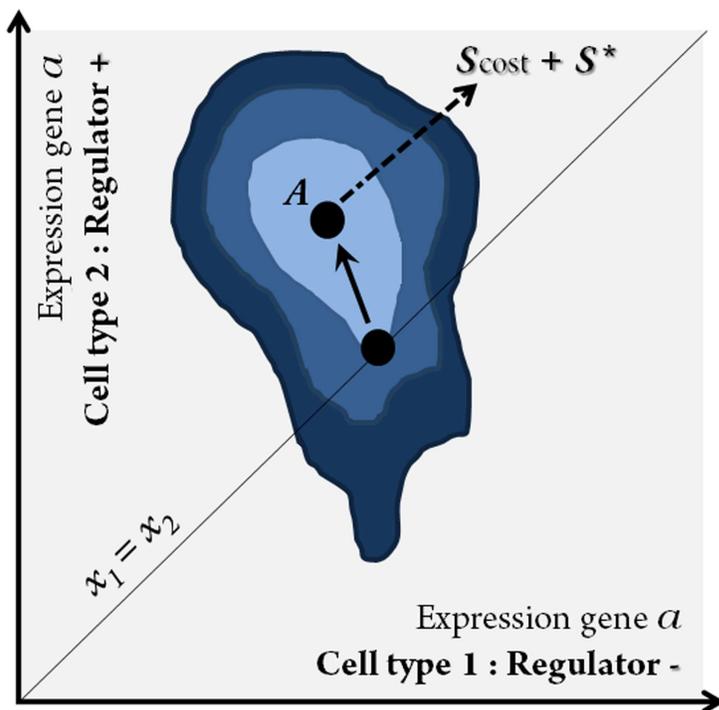
Fig. 4 Model for the acquisition of ncRNA regulators in CMOs. The figure shows the fitness landscape of gene a as a function of its expression levels in two cell types $f(x_1, x_2)$. As long as the regulation of the gene a is identical in both cell types, only the points on the median line $(x_1 = x_2)$ are accessible. If a regulatory ncRNA is innovated (regulator +) in one of the cell types, then it can modify the expression levels of gene a in that compartment. If the new regulatory ncRNA takes the system into a new $setA$ of expression values with a small energetic cost (S_{cost}), then it has a "window of opportunity" to be placed under stabilizing selection. Thus, this new regulatory ncRNA could make the phenotypic traits influenced by the regulated gene a much more "evolvable". If the organism's fitness depends on the heritability of those traits and, accordingly, on the expression levels of gene a, then the new regulatory ncRNA becomes subject to positive selection (S^*) to optimize $f(x_1, x_2)$. The resulting refinement in regulation in turn could lead to the innovation of further cell types, opening the same "window of opportunity" again for the incorporation of further regulators. See text for a detailed description.

Table 1 Some key characters that collectively underpin multicellularity in model organisms

Lineage	Cell types#	miRNA families$	Model species	Genome size (Mb)&	Protein genes&
Simple multicellularity					
Cyanobacteria	1 - 3	na*	*Nostoc punctiforme*	9	7,432
Myxobacteria	1 - 3	na	*Myxococcus xanthus*	9	7,388
Actinobacteria	1 - 3	na	*Streptomyces coelicolor*	9	7,825
Amoebozoa	1 - 3	2	*Dictyostelium discoideum*	34	12,500
Chlorophyta	1 - 5	nd**	*Volvox carteri*	138	14,520
Zygomycota	1 - 3	nd	*Rhizopus oryzae*	45	17,467
Parazoa	4 - 16	nd	*Trichoplax adhaerens*	98	11,514
Complex multicellularity					
Stramenophila	4 - 14	nd	*Laminaria dentigera*	nd	nd
Rhodophyta	6 - 14	nd	*Farlowia mollis*	nd	nd
Bryophyta	11 - 26	nd	*Physcomitrella patens*	480	35,938
Tracheophyta	5 - 44	148 - 491	*Arabidopsis thaliana*	125	25,498
Basidiomycota	9	nd	*Coprinus cinereus*	37	13,342
Ascomycota	5 - 28	2 - 88	*Neurospora crassa*	40	10,082
Cnidaria	3 - 22	nd	*Nematostella vectensis*	357	18,000
Protostomia	10 - 70	190 - 238	*Drosophila melanogaster*	180	13,601
Deuterostomia	100 - 250	132 - 1,424	*Homo sapiens*	3,100	21,494

Data compiled from [9][13][118]
$ *Data compiled from RFAM database version 10.0 (http://rfam.sanger.ac.uk/)*
& *Data compiled according to the methods described in Fig. 1*
* *na: not applicable (there are not miRNAs in prokaryotes)*
** *nd: not determined yet*

Table 2 A brief description of some regulatory ncRNA types in animals

ncRNA type	Size (nts)	Example	Functional & evolutionary relevance of the example	Reference
micro RNA (miRNA)	~20-25	*Let-7*: pre-miRNA, ~70 nts; mature miRNA, ~22 nts (metazoans)	*Nematode*: sequentially controls developmental timing of larva transitions. *Fly*: regulates the timing of neuromuscular junction formation in the abdomen and cell-cycle in the wing. *Vertebrates*: promotes terminal differentiation in development and tumor suppression.	[120]
Small nuclear RNA (snRNA)	~50-350	*7SK*: ~330 nts (metazoans)	1) Mediates the inhibition of the general transcription elongation factor P-TEFb by the HEXIM1 protein and thereby represses transcript elongation by Pol II. The 5' M3 stem of 7SK is essential to bind the HEXIM1 protein (see Fig. 3). 2) Suppresses the deaminase activity of APOBEC3C and sequesters this enzyme in the nucleolus.	[108]
Small nucleolar RNA (snoRNA)	~60-300	*HBII-52*: ~80 nts, C/D box (human)	*Brain-specific*. Controls the alternative Splicing of the Serotonin Receptor 2C. Its abrogation is involved in the origin of Prader-Willi syndrome.	[69]
Long non-coding RNA (lncRNA)	>200-100,000	*Xist*: ~17 Kb (mammals)	Is expressed exclusively in the X Inactivation Center (XIC) of the inactive X chromosome. With other ncRNAs ensure that only one X chromosome is active in female mammals.	[34]