RNA Biology 6:5 495-502; November/December 2009; © 2009 Landes Bioscience

Ancestral RNA

The RNA biology of the eukaryotic ancestor

Lesley J. Collins* and Xiaowei Sylvia Chen

Allan Wilson Centre for Molecular Ecology and Evolution; Institute for Molecular BioSciences; Massey University; Palmerston North, New Zealand

Our knowledge of RNA biology within eukaryotes has exploded over the last five years. Within new research we see that some features that were once thought to be part of multicellular life have now been identified in several protist lineages. Hence, it is timely to ask which features of eukaryote RNA biology are ancestral to all eukaryotes. We focus on RNA-based regulation and epigenetic mechanisms that use small regulatory ncRNAs and long ncRNAs, to highlight some of the many questions surrounding eukaryotic ncRNA evolution.

©2009 Landes Bioscience. Do not distribute.

Key words: eukaryote ancestor, RNA infrastructure, RNAi, epigenetics

Submitted: 10/28/08

Revised: 07/12/09

Accepted: 07/18/09

Previously published online: www.landesbioscience.com/journals/rnabiology/article/9551

*Correspondence to:
Lesley J. Collins; Email: L.J.Collins@massey.ac.nz

Introduction

Over the last five years we have witnessed an exponential growth in the interest in the non-protein-coding-RNA biology (ncRNAs) of the eukaryotic cell. Many processing features of ncRNA have of course been known for decades (e.g., snoRNA modification of rRNA, or snRNA cleavage of mRNA during splicing), and now gene regulation by RNA (e.g., RNAi) is no less accepted. However, beneath the obvious RNA families found throughout eukaryotes, new types of ncRNA-based processing and regulation are not only coming to light but are being found to be more commonplace than previously imagined. It is the distribution of these ncRNA-based complexes throughout eukaryotes that raises an important question; how many of these features are inherent to all eukaryotes and thus, can be termed ancestral?

Before looking back in time towards ancestral eukaryotes, it is often helpful to look at the big picture of today's 'RNA world'. Eukaryotic RNA biology is a network with interactions that change not only temporally during the life cycle of the cell, but also spatially as RNA moves between sub-regions of the nucleus to the cytoplasm, and sometimes back to the nucleus again. This network has been described as the 'RNA infrastructure',[1] and infers that each RNA-based machinery does not stand alone in the cell, but is connected to other RNA-based machineries. It has been long known that transcription, splicing and polyadenylation in eukaryotes were linked processes.[2,3] However, only recently can we see that the proteins that are shared between these machineries allow the processes of the RNA-infrastructure to be linked. For example, in mammals the Exon Junction Complex (EJC) is deposited on the spliced transcript after the second catalytic reaction, and this protein complex not only allows the transcript to be checked within the nucleus, but it then interacts with the RNA export machinery to get the transcript to the cytoplasm for translation (reviewed in ref. 1). The RNA infrastructure goes beyond the transcription-to-translation processes and with it we are investigating various RNA-based machineries.

Recently there has been a significant increase in RNA-biology research of the protist lineages. A number of ncRNA families that we expected to find throughout eukaryotes have been further characterized [e.g., snRNAs and snoRNAs from *Paramecium tetraurelia*[4] and *Giardia lamblia* (hereafter Giardia)[5,6]], enabling in the case of the snoRNAs further studies of expression strategies.[7] Research of other RNA families in protists is also enabling

us to find RNA-based processes that were once thought to occur only in multicellular organisms such as animals and plants [one example being microRNAs (miRNAs) which will be discussed later]. What is even more interesting is that the inverse also applies, with commonplace characteristics from some protists now being characterized in multicellular organisms. One example is the vast amounts of antisense transcripts known to occur in Giardia[8,9] being recently characterized in humans[10] and mouse.[11]

We have some prior information about the RNA-biology of the last common ancestor of extant eukaryotes (hereafter called the ancestral eukaryote, clearly there would have been earlier 'proto-eukaryotes' that have no descendents in the modern world). Earlier studies (reviewed in refs. 12 and 13) determined the presence of the 'processing' RNA families involved in the path from transcription to translation (i.e., processing of mRNA, tRNA and rRNA). RNP complexes such as RNase P are found in both eukaryotes and prokaryotes (although with structural differences and different protein cofactors) and thus, are easily considered to be ancestral to all kingdoms. However, other ncRNA families have only been found throughout eukaryotes with the small nuclear RNAs (snRNAs), the catalytic RNAs of splicing as an example. Once it was established that not only were there many ancestral snRNA-proteins, but other proteins from throughout the splicing cycle were also ancestral, we could see that major splicing as a mechanism was part of the ancestral eukaryote.[14] Recent studies have shown that minor splicing (once thought to be part of multicellular organisms only) may also be ancestral to eukaryotes, but has been lost in some lineages.[15,16] Another example is RNase MRP, a ribonucleoprotein that has roles in processing rRNA and is evolutionarily related to RNase P.[17] RNase MRP was initially characterised from animals and yeast, and has since been found in protists, but again showing loss in some lineages.[17,18]

In this review, we will focus less on the processing ncRNAs, but more on RNA-based regulation and epigenetics, because recent studies are revealing more about how these processes have evolved. Although many of them are often lumped under the generic term 'RNA interference' or RNAi, it has become clear that there are many different ways that ncRNAs can interact with host genes to upregulate, or to downregulate expression, or silence translation or guide methylation (just as some examples). It is perhaps not so clear as to which mechanisms are specific to multicellularity and which can be traced back earlier to the eukaryotic ancestor. Recent studies are bringing understanding of the ancestral eukaryotic state, but they are also raising many questions, which the move into ncRNA research from protist species is beginning to answer.

Ancestral RNA-Based Regulation

There are now many classes of ncRNA responsible for the regulation of transcription and translation processes but most fall into three larger groups, miRNA-like, siRNA-like and piRNA-like. miRNAs were once thought to only be found in multicellular eukaryotes but they have since been characterized in such diverse organisms as Excavates (e.g., Giardia and *Trichomoans vaginalis*),[19] ciliates (*Tetrahymena thermophila* and *Paramecium tetraurelia*),[20,21] trypanosomes (*Trypanosoma brucei*),[22] slime molds (*Dictyostelium discoideum*),[23] as well as single-celled green algae (*Chlamydomonas reinhardtii*).[24] It is not the object of this review to describe in detail how these regulatory RNAs work as it has been covered elsewhere.[21,25,26] Instead we will concentrate on how they are thought to have evolved.

Beginning with miRNAs, it was originally thought that animal miRNAs downregulated their targets largely by translational repression, whereas plant miRNAs used post-transcriptional gene silencing (reviewed in ref. 27). However, later studies (reviewed in ref. 27) showed that not only can a single miRNA downregulate expression of hundreds of its target genes,[28] but that some miRNAs use alternative methods of downregulation, such as mRNA cleavage or accelerated deadenylation of the polyA tail. It was also discovered that animal miRNAs can induce translational upregulation, and that some plant miRNAs can function as translational inhibitors. This shows that we cannot assume that all miRNAs within an organism will regulate in the same way. This is also an indication that although miRNAs are often grouped under the one term 'miRNA' and often maintain their simple secondary structure and binding characteristics, the mechanism by which a specific miRNA operates becomes the primary classifier. This mechanistic factor may also be key in understanding how these miRNAs evolved.

There are at present, two primary models of miRNA evolution; the de novo inventing model, where new miRNAs arise from innovative random sequences, and the duplication model where whole genome duplications have permitted expansion of existing miRNA families (reviewed in ref. 29). Within vertebrates at least, studies have indicated that miRNA numbers underwent a major expansion and this has by some researchers been connected to cellular complexity.[30,31] Opinions have also been divided over the mechanism of this expansion because whole genome duplication results from two rounds of genome duplication that occurred in early vertebrates, appear not to account for all of the differences seen when miRNAs are studied phylogenetically.[29] Many new animal miRNAs for example, have been suggested to arise from accumulation of sequence mutations and not from gene or genome duplication.[25,32] Gu et al.[29] take a combinatorial approach in that small scale genomic rearrangements and local duplication as well as whole genome duplications contributed to miRNA evolution within vertebrates. However, they make a strong point that deep phylogenetic analysis (even to deep within vertebrates) is difficult due not only to the short sequence lengths, but also to the complex nucleotide substitution patterns enforced by secondary structure constraints. This statement does not bode well for comparing sequences throughout eukaryotes and especially between such diverged groups as animals and excavates (e.g., *Giardia lamblia* and *Trichomonas vaginalis*), and is an issue that we will return to in the discussion.

However, we can begin to consider ncRNA based pathways in the same manner as we have seen previously with proteins,

©2009 Landes Bioscience. Do not distribute.

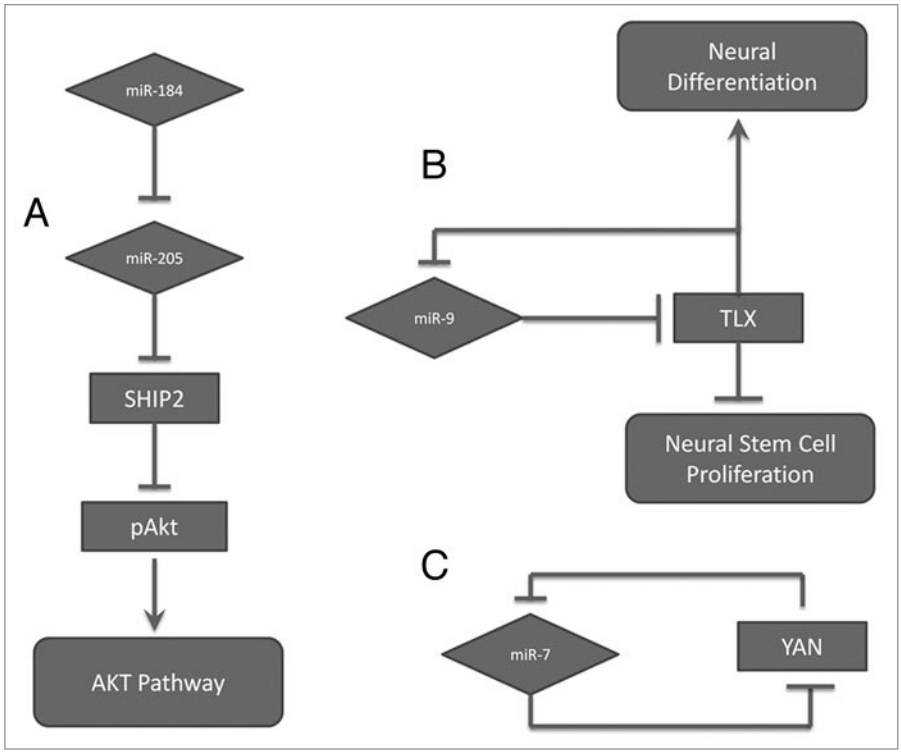

Figure 1. Network diagrams showing two examples of miRNA regulation. (A) In epithelial cells miR-184 negatively regulates another miRNA, miR-205. Interfering with miR-205 function leads to a coordinated damping of the Akt signaling pathway via SHIP2 induction and associated with a marked increase in keratinocyte apoptosis and cell death.[33] (B) miR-9 on the other hand is expressed in neurogenic areas of the brain and directly suppresses the nuclear receptor TLX; thus accelerating neural differentiation while negatively regulating stem cell proliferation.[34] (C) In Drosophila miR-7 negatively represses the protein YAN which in turn negatively represses miR-7, forming a double negative feedback loop. Network diagrams are drawn in a style taken from Zhang and Su;[27] miRNAs are represented by diamonds, proteins by rectangles and processes by rounded rectangles.

and from this we see much variation in how miRNAs regulate their targets. **Figure 1A** illustrates an example where in epithelial cells miR-184 regulates a second miRNA (miR-205) to maintain levels of the SHIP2 protein.[33] Interference with these levels can lead to increased apoptosis and cell death, suggesting that miR-205 may be a factor in squamous cell carcinomas.[33] **Figure 1B** illustrates another example where one miRNA can have a large cellular effect, as in neural cells miR-9 regulates TLX maintaining a balance between neural stem cell proliferation and differentiation.[34] Current research is finding that more and more miRNAs are subjected to feedback from their target genes. This tight regulation is very necessary as mis-expression of miRNAs frequently mimics loss-of-function phenotypes for their targets.[25] **Figure 1C** shows a well known example where the *Drosophila melanogaster* (hereafter Drosophila) miR-7 is repressed by the ETS domain transcription factor YAN (reviewed in ref. 25). YAN is itself repressed by miR-7 forming a double negative feedback loop. The RNA/target/regulation details shown in these network diagrams (reviewed in ref. 27) can fit into the RNA infrastructure framework, which we can then use in future studies to look at whether some of the linkages we see between miRNAs, genes and proteins are ancestral.

Given the close relationship between miRNAs and their targets it is not surprising that they show signs of co-evolving. The selection pressures on the evolution of the target genes would of course differ depending on whether the miRNA binds to its mRNA target in the coding region (as seen primarily in plants) or the 3' UTR (as seen primarily in animals). There is also a suggestion that genes co-expressed with a certain miRNA may have evolved to selectively avoid sites complementary with the miRNA-seed region, i.e., avoiding similarity with the miRNA will ensure they will not be silenced by it.[27] This is interesting as it suggests that the evolution of a gene that is not regulated by a particular miRNA will still be affected by it.

As with miRNAs there are many subclasses of siRNAs that can be processed either as double-stranded transcripts (generated by Dicer), or as sense-antisense pairs (e.g., bidirectional promoter produced).[25] Double-stranded RNAs are in general processed into 21–24 nt siRNAs which then 'programs' the RISC complex to degrade the target mRNA.[35] siRNAs can also be divided into those that are endogenously encoded (endo-siRNAs) from the genome (sense-antisense siRNAs) and those that are exogenous, exo-siRNAs (e.g., viral-induced siRNAs). Both miRNAs and siRNAs play an important role in viral defence especially in plants where they have been well characterised.[25,35,36]

The role of RNAi processes in protecting against viruses and transposable element has led to the idea that these ncRNAs are involved in an 'arms race' between host and parasite.[37] This could not only drive the rapid evolution of the ncRNA and their associated proteins but also the entire RNAi pathway (reviewed in ref. 37). Due to the seemingly 'ancestral' nature of transposons and viruses, we could assume that such a defence evolved early in eukaryotes and can work in a number of ways. For example, in Drosophila it has been shown[38] that it is not enough for siRNA-based viral defence to act locally at the site of infection, but the spread of dsRNAs to uninfected sites is essential for effective antiviral immunity. Upon infection, dsRNAs released from the lysis of infected cells and are taken up into uninfected cells to generate a virus-specific immunity. This is a different system from plants and nematodes which do not have such a highly efficient mechanism for dsRNA uptake.[38]

Although at first siRNAs especially in animals were mostly considered to be from foreign DNA or RNA, recent studies using deep-sequencing have characterised many endogenously encoded siRNAs (reviewed in refs. 39 and 40). Presently there appears to be three main groups of

endo-siRNAs which have been studied mostly in plants and Drosophila. The first are RDR2-dependent siRNAs which are preferentially associated with transposons, retroelements and repetitive DNA but also appear to correspond to regions containing DNA methylation.[40] Similar to the piRNAs (which are covered later) the RDR2-dependent siRNAs occur in the soma as well as the germline in Drosophila and exhibit little strand or sequence bias due to having dsRNA as their precursors.[39] This is not an issue in plants which do not have a soma/germline separation. The second group are trans-acting or ta-siRNAs which regulate target genes other than their originating loci.[40] The third group are termed natural antisense or nat-siRNAs, which are produced from loci with overlapping bidirectional transcripts with the best known examples being involved in abiotic stress response[41] or are bacterial pathogen induced.[42]

RNA-based silencing has been studied in some protist groups. RNAi has been used in trypanosomes to study other aspects of its biology for some years now.[43] The existence of these RDR2-type siRNAs is consistent with this pathway of RNA-silencing emerging early in eukaryotic evolution.[44] As yet there is little known about other endogenous siRNAs in protists other than the existence of bidirectionally-transcribed antisense products in Giardia.[9]

The third major class of small regulatory ncRNAs we will discuss are piRNAs (PIWI-interacting RNAs). Typically 25–30 nt in length, binding to the Piwi family of proteins (PIWI, Argonaute3 and Aubergine), they are not produced by dicing, and were initially found in mammal and Drosophila germline cells where they have important roles in transposon defense.[21,45,46] piRNAs can be involved in a 'ping-pong' amplification cycle, where primary antisense piRNAs identify then cleave targets which generate secondary sense piRNAs (reviewed in ref. 21). However, recently a Drosophila ovary somatic piRNA pathway has been characterised[46,47] involving only the PIWI protein (not Argonaute3 or Aubergine), where primary piRNAs derived from the *flamenco* transposons were loaded directly onto PIWI and not further amplified. These two studies show two distinct pathways are responsible for silencing different transposons classes in Drosophila ovaries.[47]

A similar type of germline amplification process can be seen in ciliates such as *Tetrahymena thermophila* (hereafter Tetrahymena) and *Paramecium tetraurelia* although not strictly a with a "ping-pong" cycle.[21] piRNA-type RNAs (scan RNAs or scnRNAs) are produced during the reorganization of the macronucleus during sexual development when some exons can become 'scrambled'.[21] In Tetrahymena ~6000 IES internal eliminated sequences consisting of transposon-like and other repeats are targeted for removal by RNA-directed heterochromatin marking.[20] Kurth and Kazufumi[20] describe how the scnRNAs can be derived from what they term 'promiscuous transcription' of the micronuclear genome. Since scnRNAs are not specific for the eliminated genes they need to be selected for IES specificity. Although the actual molecular mechanism is not as yet known, the scnRNAs pair with either DNA or RNA from the parental macronucleus to be sorted. The selected scnRNAs then move to the newly developed macronucleus where they induce heterochromatin formation on the IES prior to elimination. In a way we could regard the micronucleus as a type of 'germline cell' and we find it interesting that piRNA type ncRNAs are associated with germline retention in such diverse eukaryotes as ciliates and Drosophila.

From the recent miRNA, siRNA and piRNA studies we have described here, we can see the trend on how ncRNAs are utilized to undertake complex coordinating actions not only in multicellular eukaryotes but in single-celled protists. We have no doubt that RNAi as a general mechanism was likely to be a feature of the ancestral eukaryote since we find RNA-based regulation across eukaryotes. Analysis of eukaryote-specific proteins (or Eukaryotic Signature Proteins) has shown that the ancestral eukaryote was compartmentalized[48] and that complex RNA-based machinery like the major spliceosome was present.[14] It is expected that ncRNA was also a part of the ancestral eukaryote's cellular control. A question we must work on now is to understand how such an ancestral mechanism arose and what modern components it contained.

When investigating any RNA process under the RNA-infrastructure model we must include the essential proteins that either make up the mechanisms or transport ncRNAs. The obvious proteins for the small regulatory RNAs are the Dicer and Argonaute proteins. Some eukaryotes have a single Dicer protein (e.g., mammals and nematodes) that is involved in the processing of both miRNAs and siRNAs.[25] However, multiple Dicer proteins are found in Drosophila and plants, and it is still not clear as to whether multiple copies of the proteins specialize in either different types of processing, or perhaps in some other way, to ensure tissue or temporal specificity. Recent studies[49-51] have revealed that the key proteins (e.g., Dicer, Argonaute and RNA-dependent RNA polymerase—RDRP) and small RNA components of RNAi (e.g., miRNAs and siRNAs) are largely conserved throughout eukaryotic evolution, leaving little doubt that these proteins were present in the eukaryotic ancestor. However, until we know more about how they function in a wider range of eukaryotes we cannot as yet determine if the ancestral proteins functioned in the same way as the ones we see today.

Argonaute proteins can often be found in cytoplasmic compartments such as Golgi bodies and the Endoplasmic Reticulum, and miRISC factors may become anchored in P-bodies.[25] P-bodies (also known as GW bodies) are cytoplasmic granules where translationally inactive mRNAs can accumulate.[52,53] P-bodies are not essential for miRNA repression,[54] but P-body formation requires an intact miRNA pathway[55] and there is increasing evidence that small regulatory ncRNAs bring about degradation or translational arrest in P-bodies.[55] Other dynamic compartments in the nucleus such as Cajal bodies are involved in miRNA and siRNA biogenesis, as well as their more studied roles in snRNA and snoRNA biogenesis.[55] Since the movement of RNA-based complexes through cellular sub-compartments is in effect a form of spatial and temporal regulation,[1] it will be interesting to see if future studies will show similar compartmentalisations in protists where certain

©2009 Landes Bioscience. Do not distribute.

"common" compartments are not present (e.g., Golgi bodies and peroxisomes are absent from *Giardia lamblia*, but some Golgi features are performed by the ER or nuclear envelope[56]).

Ancestral RNA-Based Epigenetics

In eukaryotes, the fine control of chromatin architecture is one of the characteristics of gene regulation. Chromatin configuration is altered through DNA methylation and a set of histone modifications including acetylation, deacetylation, methylation, phosphorylation and carbonylation.[57] These chromatin modifications contribute to the "epigenetic memory" of the cell defined as heritable changes in gene function that occur without changes to the DNA sequence.[58,59] During the last decade, evidence of ncRNA mediated epigenetic control has increased dramatically (reviewed in refs. 12, 59–61). There has especially been an increase in the number and type of long ncRNAs (typically >200 nt) involved in gene regulation (reviewed in ref. 62). However, although some long ncRNAs act as cis-acting silencers, many other ncRNAs including some snoRNAs, miRNAs and piRNAs have been known to act as trans-acting regulators of site specific modification and imprinted gene-silencing (reviewed in ref. 61).

Genomic imprinting in humans and mouse governs ~80 genes in 25 clusters, resulting in the expression of only one of the two parental alleles.[61,63] One well known example involves two ncRNAs: rox1 and rox2. In Drosophila they are directly involved in dosage compensation leading to ~2x transcriptional activity on the male X chromosome.[64] The rox1 and rox2 RNAs spread along the X chromosome and recruit the histone deacetylation protein complex, which generates an open chromatin conformation to facilitate active transcription.[65,66] Another well known example from mammals is where dosage compensation is achieved by inactivation of one of the two X chromosomes in females during development.[67] Two RNAs: Xist (17 kb) and its antisense transcript Tsix are involved in X-chromosome inactivation in XX females. Xist RNA coats the X chromosome in cis and triggers extensive histone methylation on the future inactive X chromosome,[67] and Tsix is required to restrict Xist activity on the future active X chromosome.[68]

Recent studies in mammals have revealed more complex regulatory network of epigenetic control which involves interplay of ncRNAs of various lengths. The potential roles of small regulatory ncRNAs in X chromosome inactivation (XiRNAs) have been investigated.[69] Dicer-dependent XiRNAs sized between 24–42 nt were found to be produced from both Xist and Tsix transcripts. These XiRNAs are required for histone 3, lysine 27, trimethylation along the future inactive X chromosome, and methylation of CpG island of Xist promoter region in the future active X chromosome.[70] Although XiRNAs are produced as an effect of Dicer, RNAi is not directly involved in X chromosome inactivation, instead the effect is more to do with steady-state levels of the Xist RNA.[71] Several studies have shown that knocking-out Dicer leads to a general global methylation defect by affecting the level of de novo methyltransferases.[72-74] Adding to this complex network is another ncRNA: RepA, which has been found to mediate the heterochromatic configuration of the Xist promoter through recruiting the Polycomb repressive protein complex PRC2.[75]

Other examples of long ncRNA involved in epigenetics are Air, which regulates imprinting of a cluster of genes on mouse chromosome 17 (reviewed in ref. 63), and HOTAIR which epigenetically silences *HoxD* gene expression.[76,77] Genomic projects have recently revealed over a thousand long ncRNAs that are conserved throughout mammals,[78] and the rise of deep sequencing technology is likely to aid in investigating these ncRNAs throughout animals and then throughout eukaryotes.

At the protein level, comparative genomic studies have shown that core histone modification proteins such as histone methylases, demethylases and SWI2/SNF2 ATPases, appear to diversify through proliferating paralogous families followed by acquisition of novel domains. This then results in lineage-specific diversity in epigenetic marks.[79] These major histone modification proteins and the mechanism of chromatin remodelling are found across all lineages of eukaryotes including protists such as Giardia and Trichomonas.[79,80]

To date many chromatin modifications have been shown to be directed by short and long ncRNAs. In plants, the siRNA-directed DNA methylation is involved in heterochromatin formation.[81,82] In fission yeast, siRNAs direct heterochromatin formation via RNA-induced transcriptional silencing or RNA-directed RNA polymerase pathway.[82] In animals, siRNAs provide initial scaffolding of heterochromatin[83] which is inherited during the cell cycle.[59] In addition, piRNAs and promoter targeted small RNAs which mediate transcriptional gene silencing,[84] are known to target upstream of chromatin modification.[84,85] Despite the individual variances in these pathways, they all share several key protein components including Argonaute, PIWI, RDRP and Dicer. Many RNA-directed epigenetic regulation events thus appear to be sharing protein and RNA components with the RNAi pathway if not dependent on the latter. We can also note that evolution of ncRNAs by duplication could allow epigenetic states such as methylation and imprinting, between the two copies to differ.[29] Since there are instances where a single trans-acting siRNA may have ~2,300 predicted gene targets,[61] this type of duplication could possibly result in a significant change in phenotype.

Figure 2 shows a selection of RNA-regulated epigenetic pathways in animals, plants, insects and other eukaryotes. Protein homologues are found in most eukaryotes, and thus are very likely to be present in the ancestral eukaryote, and are highlighted. Most pathways indicated here ultimately lead to methylation of DNA and/or histones. Although methylations are directed by different methylases across various species, the methylation events are common, and therefore could also be an ancestral feature of eukaryotes. While most studies have been conducted on major model organisms, little is known about the epigenetic regulation in the many protist lineages of eukaryotes. However, recently long ncRNAs have been found in a study of the malaria parasite *Plasmodium falciparum*, where sterile sense and antisense RNAs were found to

©2009 Landes Bioscience. Do not distribute

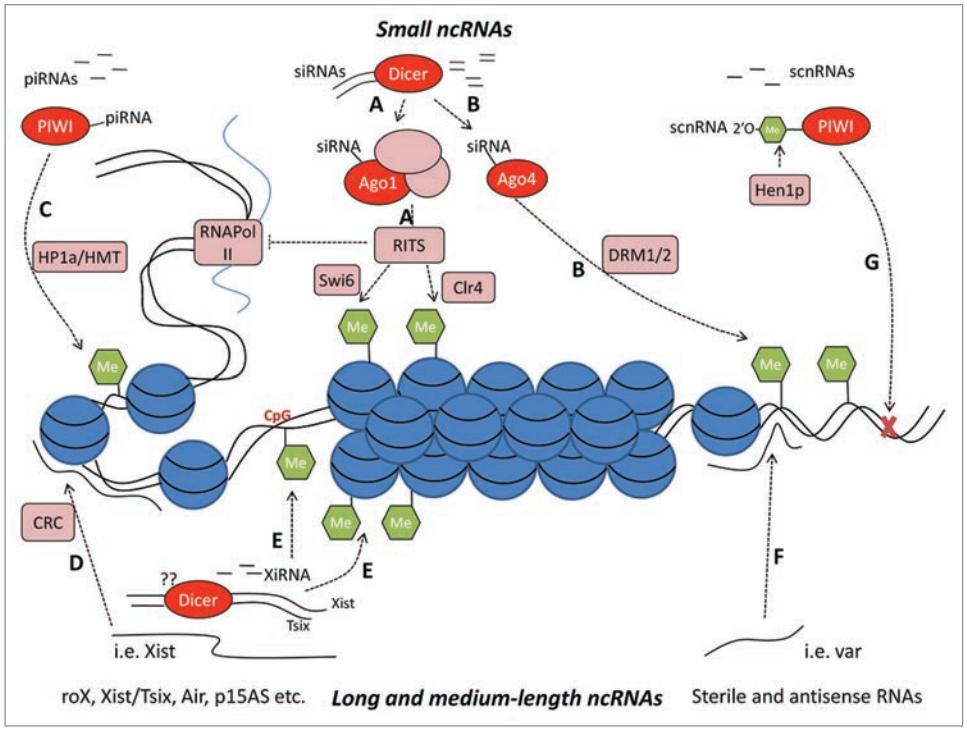

Figure 2. Epigenetic pathways regulated by ncRNAs across eukaryotes. For example: (A) In fission yeast, the RNAi machinery functions to recruit the RNA-induced initiation of transcriptional gene silencing complex (RITS) and subsequently induces heterochromatin formation either through targeting nascent transcripts or further recruiting histone methyltransferases (Swi6, Clr4). (B) In plants, siRNAs associate with Ago4 protein and can direct DNA methylation by domains rearranged methyltransferase 1/2 (DRM1/2). (C) piRNAs bound to PIWI protein can induce heterochromatin formation in Drosophila through recruitment of DNA methyltransferase HP1a or direct DNA methylation by HMT in mammalian germ cells. (D) Long ncRNAs such as the mammalian Xist coats the X chromosome in cis and recruits chromatin repressor complex (CRC) to the future inactive X chromosome during X-chromosome inactivation in females. (E) The Xist RNA can also be processed into small XiRNAs through complementary binding with Tsix ncRNA followed by possible action of Dicer. XiRNAs are involved in in histone methylation of the future inactive X chromosome and inhibition of Xist transcription in the future active X chromosome through CpG island methylation. (F) Chromatin associated sense and antisense ncRNAs transcribed from the *var* gene family in Plasmodium are involved in chromatin remodeling and switching of surface antigen expression. (G) PIWI binding scnRNAs in Tetrahymena are 2'-O methylated by the methyltransferase Hen1p, and are required for DNA elimination.

be transcribed from the *var* virulence gene family and coated chromatin in a similar way as the animal Xist RNA and the Drosophila rox RNAs.[86] The regulation of the VSP genes involved in antigenic switching in Giardia has also been thought to be epigenetically regulated.[80] Subsequently there has been the identification of key RNAi proteins,[49,79] some miRNAs[19] and a little on the regulatory mechanism of the VSP genes,[50] but nothing is known as yet about chromatin modifications in Giardia. Of what we do know it is likely that the eukaryotic ancestor contained some type of chromatin-modified regulation but much more work needs to be done to define it more than that.

Discussion

Looking back down an evolutionary pathway allows us to analyze how a mechanism changes due to different selection pressures and this can impact on phenotype. These pressures may be cellular (the change of an auxiliary protein) environmental (responses to a lack of a cofactor normally found in a previous diet), or even larger scale (the movement from free-living to parasitic lifestyle or single cell to multicellular composition). It is feasible that ncRNA-based networks may allow the cell to respond more rapidly to such changes but as yet we do not know the 'rules' by which the different classes of ncRNAs evolve. Unlike protein-coding sequences, ncRNAs can change their nucleotide sequence so long as essential protein binding and RNA secondary structure is conserved. However, this premise is largely based on studies of the well known ncRNAs (e.g., tRNAs, rRNAs, RNaseP RNA). Studies are beginning to emerge on miRNA evolution,[29-32] where length and complementarity to target sequences may be important factors, but will the same rules apply for the evolution of long ncRNAs? With our expansion of ncRNA knowledge, we can no longer lump all ncRNAs under one set of evolutionary ideas.

It was once surmised that some ncRNAs such as miRNAs evolved 'for' the developmental regulation of multicellular eukaryotes. However, Darwinian models of evolution state that a new function cannot be selected to fulfil a future role, but instead existing RNAs or proteins can be co-opted or recruited into new roles.[1] It is clear that there are many roles that regulatory ncRNAs play in eukaryotes, and many more than what could be covered here. The obvious questions are how many of these roles were ancestral to eukaryotes, and which roles have evolved due to developmental and/or multicellular complexity.

Thus, due to characterization of complex ncRNA-machinery in many protist lineages, single-celled eukaryotes can no longer be termed 'primitive', at least at the ncRNA level. The level of developmental complexity seen in multicellular organisms can be seen to some extent in the complex life-cycles that many protists display and we have as yet to uncover what (if any) RNA biology is behind this complexity.

We know that in the ancestral eukaryote, we find ancestral proteins (i.e., Eukaryotic Signature proteins that are conserved throughout eukaryotes[48]), thus if RNAi was ancient could we expect them to be largely regulated by ancient RNAs? We cannot assume that the most well known systems (discovered in multicellular organisms because their RNA biology has been the most studied) is the most ancestral. It will be the investigation of the most crucial parts of the mechanisms that will give us the clues. How the different mechanisms link will also give an idea as to the ancestral state. One question that may take a few more years to answer is the following: for each ncRNA mechanism described here, does any modern eukaryote contain an 'ancestral' type system? The lack of conservation of a single mechanism of RNA regulation does not necessitate that the ancestral eukaryote contained multiple systems. This could have been the case but it is also a possibility that RNA regulation is ideally suited to be integrated into solving co-expressional regulation issues and has been utilized in different ways in different eukaryotes then evolved as each of those eukaryotes changed due to environmental and/or multicellular changes. It is therefore likely that there is no ancestral-type system still present in today's eukaryotes since such mechanisms have been shown to be rapidly evolving. Investigations of non-model protist ncRNA-based processing will elucidate a clearer picture of the distribution and lineage-specific characteristics of RNA regulatory systems we find in common. It will also enable us to compare newly-discovered protist ncRNAs with those from model eukaryotes and see how different ncRNAs can evolve.

Over the years we have become used to the RNA world concept where RNA was both the storage and catalytic molecule before being replaced to a large extent by DNA and proteins. Although our studies of especially small ncRNAs such as miRNAs show that ncRNAs are continuously evolving, we raise the open question of whether any of the RNA infrastructure linkages go all the way back to the RNA world and thus are 'relics', as opposed to being 'ancestral' to extant eukaryotes. Relics we could surmise as being either still present or lost from prokaryotes, and could reflect on RNA-based regulation within the earliest stages of life. Processes such as transcription and translation remain largely the same in eukaryotes and prokaryotes (when we look at this on a grand scale level) so we can surmise that the RNA-based functions involving the transfer of message by mRNA, the structural processing by rRNA and even the transfer of cofactors (if you could consider amino acids cofactors) by tRNAs. Regulation of RNAs by RNAs thus appears to be a natural progression of this idea. With the characterisation of regulatory RNAs in prokaryotes (reviewed in ref. 87) this idea can be investigated further.

From what we have seen over the last few years, as more deep-sequencing projects analyse ncRNAs from non-model organisms, we will find novel RNA regulatory mechanisms, but we will likely also find new twists on mechanisms we thought we knew well. Without a doubt the next few years hold a bounty of surprises in the modern RNA world.

©2009 Landes Bioscience. Do not distribute

Acknowledgements

The authors send many thanks to David Penny for the critical reading of this manuscript and support in this work. This work was supported financially by the Allan Wilson Centre for Molecular Ecology and Evolution and the Institute for Molecular BioSciences at Massey University.

References

1. Collins LJ, Penny D. The RNA infrastructure: dark matter of the eukaryotic cell? Trends Genet 2009; 25:120-8.
2. Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ, Nogues G. Multiple links between transcription and splicing. RNA 2004; 10:1489-98.
3. Millevoi S, Loulergue C, Dettwiler S, Karaa SZ, Keller W, Antoniou M, Vagner S. An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. EMBO 2006; 25:4854-64.
4. Chen CL, Zhou H, Liao JY, Qu LH, Amar L. Genome-wide evolutionary analysis of the noncoding RNA genes and noncoding DNA of *Paramecium tetraurelia*. Rna 2009; 15:503-14.
5. Chen XS, Rozhdestvensky TS, Collins LJ, Schmitz J, Penny D. Combined experimental and computational approach to identify non-protein-coding RNAs in the deep-branching eukaryote *Giardia intestinalis*. Nucleic Acids Res 2007; 35:4619-28.
6. Chen XS, White WT, Collins LJ, Penny D. Computational identification of four spliceosomal snRNAs from the deep-branching eukaryote *Giardia intestinalis*. PLoS ONE 2008; 3:3106.
7. Dieci G, Preti M, Montanini B. Eukaryotic snoRNAs: A paradigm for gene expression flexibility. Genomics 2009.
8. Elmendorf HG, Singer SM, Nash TE. The abundance of sterile transcripts in *Giardia lamblia*. Nucleic Acids Res 2001; 29:4674-83.
9. Teodorovic S, Walls CD, Elmendorf HG. Bidirectional transcription is an inherent feature of *Giardia lamblia* promoters and contributes to an abundance of sterile antisense transcripts throughout the genome. Nucleic Acids Res 2007; 35:2544-53.
10. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW. The antisense transcriptomes of human cells. Science 2008; 322:1855-7.
11. Werner A, Carlile M, Swan D. What do natural antisense transcripts regulate? RNA Biol 2009; 6:43-8.
12. Amaral PP, Dinger ME, Mercer TR, Mattick JS. The eukaryotic genome as an RNA machine. Science 2008; 319:1787-9.
13. Costa FF. Non-coding RNAs: Lost in translation? Gene 2007; 386:1-10.
14. Collins L, Penny D. Complex spliceosomal organization ancestral to extant eukaryotes. Mol Biol Evol 2005; 22:1053-66.
15. Davila Lopez M, Rosenblad MA, Samuelsson T. Computational screen for spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal components. Nucleic Acids Res 2008; 36:3001-10.
16. Russell AG, Charette JM, Spencer DF, Gray MW. An early evolutionary origin for the minor spliceosome. Nature 2006; 443:863-6.
17. Woodhams MD, Stadler PF, Penny D, Collins LJ. RNase MRP and the RNA processing cascade in the eukaryotic ancestor. BMC Evol Biol 2007; 7:13.
18. Piccinelli P, Rosenblad MA, Samuelsson T. Identification and analysis of ribonuclease P and MRP RNA in a broad range of eukaryotes. Nucleic Acids Res 2005; 33:4485-95.
19. Chen XS, Collins LJ, Biggs PJ, Penny D. microRNAs of *Giardia lamblia* and *Trichomonas vaginalis* from mass sequencing. Genome Evolution and Biology 2010:(Accepted).
20. Kurth HM, Mochizuki K. Non-coding RNA: A bridge between small RNA and DNA. RNA Biol 2009; 6:138-40.
21. Malone CD, Hannon GJ. Small RNAs as guardians of the genome. Cell 2009; 136:656-68.
22. Ngo H, Tschudi C, Gull K, Ullu E. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. Proc Natl Acad Sci USA 1998; 95:14687-92.
23. Martens H, Novotny J, Oberstrass J, Steck TL, Postlethwait P, Nellen W. RNAi in Dictyostelium: the role of RNA-directed RNA polymerases and double-stranded RNase. Mol Biol Cell 2002; 13:445-53.
24. Casas-Mollano JA, Rohr J, Kim EJ, Balassa E, van Dijk K, Cerutti H. Diversification of the core RNA interference machinery in *Chlamydomonas reinhardtii* and the role of DCL1 in transposon silencing. Genetics 2008; 179:69-81.
25. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. Cell 2009; 136:642-55.
26. Faller M, Guo F. MicroRNA biogenesis: there's more than one way to skin a cat. Biochim Biophys Acta 2008; 1779:663-7.

27. Zhang R, Su B. Small but influential: the role of microRNAs on gene regulatory network and 3'UTR evolution. J Genet Genomics 2009; 36:1-6.
28. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 2005; 433:769-73.
29. Gu X, Su Z, Huang Y. Simultaneous expansions of microRNAs and protein-coding genes by gene/genome duplications in early vertebrates. J Exp Zoolog B Mol Dev Evol 2009; 312:164-70.
30. Heimberg AM, Sempere LF, Moy VN, Donoghue PC, Peterson KJ. MicroRNAs and the advent of vertebrate morphological complexity. Proc Natl Acad Sci USA 2008; 105:2946-50.
31. Taft RJ, Pheasant M, Mattick JS. The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays 2007; 29:288-99.
32. Lu YD, Gan QH, Chi XY, Qin S. Roles of microRNA in plant defense and virus offense interaction. Plant Cell Rep 2008; 27:1571-9.
33. Yu J, Ryan DG, Getsios S, Oliveira-Fernandes M, Fatima A, Lavker RM. MicroRNA-184 antagonizes microRNA-205 to maintain SHIP2 levels in epithelia. Proc Natl Acad Sci USA 2008; 105:19300-5.
34. Zhao C, Sun G, Li S, Shi Y. A feedback regulatory loop involving microRNA-9 and nuclear receptor TLX in neural stem cell fate determination. Nat Struct Mol Biol 2009; 16:365-71.
35. Pantaleo V, Szittya G, Burgyan J. Molecular bases of viral RNA targeting by viral small interfering RNA-programmed RISC. J Virol 2007; 81:3797-806.
36. Ghosh Z, Mallick B, Chakrabarti J. Cellular versus viral microRNAs in host-virus interaction. Nucleic Acids Res 2009; 37:1035-48.
37. Obbard DJ, Gordon KH, Buck AH, Jiggins FM. The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc Lond B Biol Sci 2009; 364:99-115.
38. Saleh MC, Tassetto M, van Rij RP, Goic B, Gausson V, Berry B, et al. Antiviral immunity in Drosophila requires systemic RNA interference spread. Nature 2009; 458:346-50.
39. Golden DE, Gerbasi VR, Sontheimer EJ. An inside job for siRNAs. Mol Cell 2008; 31:309-12.
40. Ramachandran V, Chen X. Small RNA metabolism in Arabidopsis. Trends Plant Sci 2008; 13:368-74.
41. Shukla LI, Chinnusamy V, Sunkar R. The role of microRNAs and other endogenous small RNAs in plant stress responses. Biochim Biophys Acta 2008; 1779:743-8.
42. Xie Z, Qi X. Diverse small RNA-directed silencing pathways in plants. Biochim Biophys Acta 2008; 1779:720-4.
43. Bellofatto V, Palenchar JB. RNA interference as a genetic tool in trypanosomes. Methods Mol Biol 2008; 442:83-94.
44. Best A, Handoko L, Schluter E, Goringer HU. In vitro synthesized small interfering RNAs elicit RNA interference in african trypanosomes: an in vitro and in vivo analysis. J Biol Chem 2005; 280:20573-9.
45. Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, et al. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature 2008; 455:1193-7.
46. Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell 2009; 137:509-21.
47. Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell 2009; 137:522-35.
48. Kurland CG, Collins LJ, Penny D. Genomics and the irreducible nature of eukaryote cells. Science 2006; 312:1011-4.
49. Macrae IJ, Li F, Zhou K, Cande WZ, Doudna JA. Structure of Dicer and mechanistic implications for RNAi. Cold Spring Harb Symp Quant Biol 2006; 71:73-80.
50. Prucca CG, Slavin I, Quiroga R, Elias EV, Rivero FD, Saura A, et al. Antigenic variation in *Giardia lamblia* is regulated by RNA interference. Nature 2008; 456:750-4.
51. Saraiya AA, Wang CC. snoRNA, a novel precursor of microRNA in *Giardia lamblia.* PLoS Pathog 2008; 4:1000224.
52. Brengues M, Parker R. Accumulation of Polyadenylated mRNA, Pab1p, eIF4E and eIF4G with P-Bodies in *Saccharomyces cerevisiae.* Mol Biol Cell 2007; 18:2592-602.
53. Kedersha N, Stoecklin G, Ayodele M, Yacono P, Lykke-Andersen J, Fritzler MJ, et al. Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. J Cell Biol 2005; 169:871-84.
54. Eulalio A, Behm-Ansmant I, Schweizer D, Izaurralde E. P-body formation is a consequence, not the cause of RNA-mediated gene silencing. Mol Cell Biol 2007; 27:3970-81.
55. Pontes O, Pikaard CS. siRNA and miRNA processing: new functions for Cajal bodies. Curr Opin Genet Dev 2008; 18:197-203.
56. Elias EV, Quiroga R, Gottig N, Nakanishi H, Nash TE, Neiman A, Lujan HD. Characterization of SNAREs determines the absence of a typical Golgi apparatus in the ancient eukaryote *Giardia lamblia.* J Biol Chem 2008; 283:35996-6010.
57. Kouzarides T. Chromatin modifications and their function. Cell 2007; 128:693-705.
58. Martin C, Zhang Y. Mechanisms of epigenetic inheritance. Curr Opin Cell Biol 2007; 19:266-72.
59. Probst AV, Dunleavy E, Almouzni G. Epigenetic inheritance during the cell cycle. Nat Rev Mol Cell Biol 2009; 10:192-206.
60. Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. Nat Rev Genet 2007; 8:413-23.
61. Royo H, Cavaille J. Non-coding RNAs in imprinted gene clusters. Biol Cell 2008; 100:149-66.
62. Mattick JS. The genetic signatures of noncoding RNAs. PLoS Genet 2009; 5:1000459.
63. Latos PA, Barlow DP. Regulation of imprinted expression by macro non-coding RNAs. RNA Biol 2009; 6: 100-6.
64. Franke A, Baker BS. The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila. Mol Cell 1999; 4:117-22.
65. Oh H, Park Y, Kuroda MI. Local spreading of MSL complexes from roX genes on the Drosophila X chromosome. Genes Dev 2003; 17:1334-9.
66. Park Y, Kelley RL, Oh H, Kuroda MI, Meller VH. Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins. Science 2002; 298:1620-3.
67. Heard E, Rougeulle C, Arnaud D, Avner P, Allis CD, Spector DL. Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. Cell 2001; 107:727-38.
68. Stavropoulos N, Lu N, Lee JT. A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. Proc Natl Acad Sci USA 2001; 98:10232-7.
69. Navarro P, Pichard S, Ciaudo C, Avner P, Rougeulle C. Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation. Genes Dev 2005; 19:1474-84.
70. Ogawa Y, Sun BK, Lee JT. Intersection of the RNA interference and X-inactivation pathways. Science 2008; 320:1336-41.
71. Kanellopoulou C, Muljo SA, Dimitrov SD, Chen X, Colin C, Plath K, Livingston DM. X chromosome inactivation in the absence of Dicer. Proc Natl Acad Sci USA 2009; 106:1122-7.
72. Nesterova TB, Popova BC, Cobb BS, Norton S, Senner CE, Tang YA, et al. Dicer regulates Xist promoter methylation in ES cells indirectly through transcriptional control of Dnmt3a. Epigenetics Chromatin 2008; 1:2.
73. Sinkkonen L, Hugenschmidt T, Berninger P, Gaidatzis D, Mohn F, Artus-Revel CG, et al. MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. Nat Struct Mol Biol 2008; 15:259-67.
74. Varambally S, Cao Q, Mani RS, Shankar S, Wang X, Ateeq B, et al. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. Science 2008; 322:1695-9.
75. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 2008; 322:750-6.
76. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 2007; 129:1311-23.
77. Brock HW, Hodgson JW, Petruk S, Mazo A. Regulatory noncoding RNAs at Hox loci. Biochem Cell Biol 2009; 87:27-34.
78. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 2009; 458:223-7.
79. Iyer LM, Anantharaman V, Wolf MY, Aravind L. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. Int J Parasitol 2008; 38:1-31.
80. Kulakova L, Singer SM, Conrad J, Nash TE. Epigenetic mechanisms are involved in the control of *Giardia lamblia* antigenic variation. Mol Microbiol 2006; 61:1533-42.
81. Chinnusamy V, Zhu JK. RNA-directed DNA methylation and demethylation in plants. Sci China C Life Sci 2009; 52:331-43.
82. Verdel A, Vavasseur A, Le Gorrec M, Touat-Todeschini L. Common themes in siRNA-mediated epigenetic silencing pathways. Int J Dev Biol 2009; 53:245-57.
83. Cam HP, Chen ES, Grewal SI. Transcriptional scaffolds for heterochromatin assembly. Cell 2009; 136:610-4.
84. Hawkins PG, Santoso S, Adams C, Anest V, Morris KV. Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells. Nucleic Acids Res 2009.
85. Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, et al. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell 2008; 31:785-99.
86. Epp C, Li F, Howitt CA, Chookajorn T, Deitsch KW. Chromatin associated sense and antisense noncoding RNAs are transcribed from the var gene family of virulence genes of the malaria parasite *Plasmodium falciparum.* Rna 2009; 15:116-27.
87. Waters LS, Storz G. Regulatory RNAs in bacteria. Cell 2009; 136:615-28.

©2009 Landes Bioscience. Do not distribute