Review

Non Coding RNAs and Viruses in the Framework of the Phylogeny of the Genes, Epigenesis and Heredity

Daniel Frías-Lasserre

Institute of Entomology, Metropolitan University of Educational Sciences,
Avenue J.P. Alessandri 774 Ñuñoa, Código Postal 7760197, Santiago, Chile;
E-Mail: daniel.frias@umce.cl; Tel.: +56-2-2412457; Fax: +56-2-2412699.

Abstract: The origin of genes is one of the most enigmatic events in the origin of life. It has been suggested that noncoding (nc) RNA was probably a precursor in the formation of the first polypeptide, and also at the origin of the first manifestation of life and genes. ncRNAs are also becoming central for understanding gene expression and silencing. Indeed, before the discovery of ncRNAs, proteins were viewed as the major molecules in the regulation of gene expression and gene silencing; however, recent findings suggest that ncRNA also plays an important role in gene expression. Reverse transcription of RNA viruses and their integration into the genome of eukaryotes and also their relationship with the ncRNA suggest that their origin is basal in genome evolution, and also probably constitute the first mechanism of gene regulation. I am to review the different roles of ncRNAs in the framework of gene evolution, as well as the importance of ncRNAs and viruses in the epigenesis and in the non-Mendelian model of heredity and evolution.

1. Introduction

The origin of genes, the genetic code and the genetic program, are puzzling events in the origin of life. At present, we know that in the genome of eukaryotic organisms there are different regions. Scattered throughout the genome are the regions of DNA coding for proteins, which in humans corresponds to approximately 2% of the genome. In the remaining 98%, the retroviral genes and also broad segments of DNA that code for non-coding (nc) RNAs are located.

Once it was established by Avery *et al.* in 1944 that the DNA corresponded to the hereditary material, the focus was mainly on the DNA as the central molecule in the hereditary models. This concept was consolidated as the Central Dogma of Molecular Biology [1]. Thus, a deterministic and reductionist inherited pattern emerged that not only had influence on molecular biology, but also on the population genetics and organic evolution. Under this scenario, both the genetic code and the genetics program were mainly associated with the coding DNAs.

At the dawn of genetics as an experimental science, the Mendelian gene was considered a symbol and a fixed and discrete heritable trait, indivisible by recombination, a concept that changed between the years 1955–1959 with Seymour Benzer's studies on T4 phages in *Escherichia coli* [2,3]. Benzer found that the gene could recombine internally and, unlike the Mendelian gene, the same segment of DNA could be defined based on different criteria such as mutation, recombination and function.

With the first DNA sequencing effort by Frederick Sanger, it was clearly demonstrated that the gene is a nucleotide sequence that encodes proteins, in which genes encode for the amino acid sequence of the primary structure of proteins [4]. Before the human genome project, the hypothesis was that the human species should have about 100,000 genes. However, surprisingly human genes are about 30,000. About 10,000 more than in *Drosophila melanogaster*, and in rice about 15,000 genes more than human have been described. There are discrepancies between the number of coding genes and the size of the genome, these paradoxical facts are explained because the more variable and abundant regions are located in the non-coding regions. Thus, the proteomes of the higher organisms are relatively stable; humans and mice share 99% of protein coding genes and between human and chimpanzee structural genes show 98% similarity [5,6]. Therefore, interspecific differences have to be located in the non-coding regions. Then, some questions arise: what is a gene? And also where is the regulatory program?

According to a holistic concept, a gene is a useful sequence of nucleotides. Repeated sections of non coding DNA involved in the formation of centromeres, telomeres and replication origins should be called also genes [7–10]. Also, it has been proposed that areas of the DNA coding for ncRNAs should also be called genes [11]. Under this framework, virtually the entire genome is useful to the organism. The genetic program important in the formation of a multicellular organism would be located in the non-coding area, through fine processes of genetic regulation via proteins, retroviruses, transposable genetic elements, introns and ncRNAs. Recently, a model has been proposed based on a hypothetical "master development program" for multi-cellular organisms where the DNA sequences are protected by the heterochromatin and transcribed into "Control Keys" that consist of nuclear messenger RNAs that regulate high-level transcription factor genes [12].

Genes were considered a continuous segment of nucleotides, until 1977 when Roberts and Sharp found that adenovirus transcribed RNA was fractionated into separate segments showing that genes are split into exons and introns. Later, in 1989, Thomas Cech and Sidney Altman discovered the catalytic properties of RNAs. They found that group I introns can be excised in the absence of any proteins acting as an enzyme (ribozyme).

The discovery of ribozymes not only changed the central position of the DNA in the dogma of molecular biology but also the paradigm that DNA was the precursor molecules of life, and spurred the idea of a primitive RNA world [13]. This hypothesis also contributed to the discovery of a huge variety of ncRNA with different functions in the three domains of the tree of life. Many classes of the non-

coding RNAs have been discovered in the human genome and constitute about 60% of the total transcriptional output [14–16].

These ncRNAs, together with DNA methylation and acetylation of histones, are epigenetic factors that have changed the understanding of genome structure, the concept of the gene, regulation of gene expression and the mechanisms that account for organic evolution. The questions that arise are: what is the origin of ncRNAs? What is the relationship between ncRNAs with transposable genetic elements and viruses? What impact have they had on the legacy and evolution of species? And, should viruses be considered as a fourth domain? This article aims to answer these questions.

2. Results and Discussion

2.1. The Discovery and the Origin of Non-Coding RNAs

The first ncRNAs were detected by Robert and Sharp in 1977 as micro (mi) RNAs during the discovering that structural genes were split into exons and introns. At first it was thought that introns have no function. However, since the discovery of messenger RNA splicing in adenoviruses [17], the alternative splicing of introns explain the production of several protein in basis to the same mRNA. Thus, it is estimated that the human genome has about 30,000 genes but about 75,000 different kinds of proteins.

The first ncRNAs with a function were described in 1993 in *Caenorhabditis elegans* when it was found that small miRNAs (constituted of about 22 nucleotides) were important for the appropriate timing of post-embryonic development [18,19]. Actually, it is known that around 98% of all transcriptional output in humans is ncRNAs [6]. Recently, the presence of ribozymes has been revealed in the ncRNAs of mammalian genomes [20–23]. One of these ribozymes, discontinuous hammerhead ribozyme has been initially described in viroids [24] and then also in eukaryotic genomes—plants and animals vertebrates and invertebrates. Some of these ribozymes have also been associated with retrotransposable elements [23]. The similarity showed between some those ribozymes suggests a viroid origin. It has also been reported that many of the ncRNA come from introns. Introns were inserted into preformed genes late in eukaryotic evolution [25].

Another hypothesis suggests that the first introns probably originated about 3500 million years ago in eubacteria and were restricted to tRNAs, and were mobile and self-splicing. From these spliced introns would have evolved spliceosomal from a common ancestor of eucaryotes and archaebacteria about 1700 million years ago with the origin of the nucleus and after the origin of mitochondria [26]. Surprising homology between viruses with very distantly related hosts by phylogenetic analysis suggest that genes might have flowed from viruses to eukaryotic chromosomes [27]. Viral oncogenes contain introns that are important in the expression of these oncogenes by alternative RNA splicing in papillomavirus genome [28], these early primary transcripts are bicistronic or polycistronic, and each contains exons and introns [29].

Similar sequences have been described between reverse transcriptase viruses, transposable elements, and mitochondrial introns [30]. Similarly, on the basis of genetic homology it has been discovered that a virophage represents a probably common origin between a DNA viruses and eukaryotic DNA transposons. The virophage parasitizes a giant virus and encodes 20 predicted proteins, these facts

suggest that transposons may have originated from ancient relatives of giant virus, and thereby influenced the evolution of eukaryotic genomes [31]. Also there are DNA virophages that parasitize species of bacteria, archeae and eukarya dependent in their reproduction of giant viruses [32].

The discovery of viroids by Diener [33] led to the modification of the paradigm that considered viruses as the smallest inciting agents of infectious diseases. Viroids, single-stranded circular RNAs of 246–375 nucleotides in length able to infect certain plants, are currently the lowest step of the biological scale. They are located in the nucleus of the host cell and probably interfere with the removal of introns and splicing exons, acting as ribozyme. The genomes of viroids are not translated, and are able to self-cleave through hammerhead ribozymes [33,34]. Viroids have nucleotide sequences similar to introns that are removed. They are parasites of plants and can be transmitted through seeds and parasite vectors. These peculiar features of viroids, along with the presence of ribozymes in some of them have been considered as molecular fossils that originated in a precellular environment of RNA, whose existence was probably before the emergence of life based on DNA and proteins [35].

In mammals, many miRNAs derive from repeat elements such as LINES and Alu [36–38]. Alu elements were originally thought to represent junk having no biological functions, nevertheless it has been postulated that ALU elements in human mRNAs are miRNAs targets, and also that other classes of ncRNAs such as short interfering (si)RNAs derive from other small RNAs [37].

The resemblance of these ribozymes in viruses and also in organism belonging to the three domains of life suggests horizontal gene transfer. For example, the similarity of some vertebrate ribozymes with those widespread within mobile genetic elements in trematodes suggests that recurrent genetic horizontal transfers could have taken place from parasites to hosts [23].

2.2. Functions of Non Coding RNAs

In recent years a wealth of information has accumulated about ncRNAs. They are characterized according to their function, location, length and also in relation with nearest structural genes. The non-coding RNAs have roles in a great variety of processes, including transcriptional regulation, chromosome replication, mRNA processing and modification, mRNA stability and translation, sex determination and even protein degradation and viral defense [39–41]. For example, the alteration or loss of non-coding RNAs results in modification in developmental processes and diseases [42].

When their function is known, ncRNAs can also be classified by whether they act in *cis* or *trans*. *Trans*-acting functions are associated with shorts ncRNAs (18–300 nt), such as siRNAs, micro (mi) RNAs, piwi-interacting(pi)RNAs and short nucleorar (sno)RNAs. Some miRNAs having both oncogenic and tumor-suppressive functions are dysregulated in many types of cancer. miRNAs also interfere with metastasis, apoptosis and invasiveness of cancer cells. [15,40,43,44]. In recent years, miRNAs have been described in mammals playing a role in the regulation of neural plasticity, synaptic plasticity regulation of learning and memory and cognitive capacity by regulating dendrite morphogenesis during early development [45,46]. Endo-siRNAs show a significant increase during an early stage of training and have been implicated in neuropsychiatric diseases [47].

Viruses can also generate viral miRNAs that disturb normal host cell functions. About 20–25% of human cancers have a known viral etiology. Expression of the viral oncogenes are regulated by alternative RNA splicing and several species of mRNAs can be derived from a primary transcript of a

single viral oncogene to encode different oncoprotein [28,48,43]. Also, viral ncRNAs have been shown to play an important role in virus-host interplay to facilitate virus replication [49].

In opposition, *cis*- acting functions have so far only been associated with macro or long nc (lnc) RNAs, which can be up to several hundred or thousand nucleotides long, about 200–2800 nt [50,51]. In the eukaryotic genome and, especially in mammals, there are thousands of lncRNAs that are expressed in different cell lines and tissues. These lncRNAs have different regulatory functions, principally X chromosome inactivation by hetrochromatinization (*Xist* gene in *cis*), [52], regulation of transcriptional and post transcriptional pathway programming, regulation of mRNA splicing, epigenetic gene activation in the regulation of *Hox* genes, in genome imprinting and as enhancers of gene expression [42,53–56]. LncRNAs have functions also in telomere-length of chromosomes [57–60].

2.3. The Non-Coding RNAs and Their Relationship with Epigenesis and Genomic Imprinting

The concept of epigenesis was born to oppose the theory of preformism. This term was coined in 1942 by Conrad H. Waddigton to explain how an adult can be formed from a zygote, through development, by cell differentiation and gene regulation. The epigenesis is defined as heritable changes in the expression of genes that do not involve a change in DNA sequence but only changes in the chromatin. These changes alter the ability of genes to respond to external signals [50]. Thus epigenetic control of gene expression allows heritable or transgenerational changes in gene expression without the need of mutations, not necessarily following Mendelian patterns of inheritance. Epigenesis can thus, for example, explain the hereditary basis of the classical concepts of norm of reaction and phenotypic plasticity. Classically, these concepts are related to a variation with an exclusively environmental base. However, now we know very well that epigenetic regulation of gene expression is accomplished by DNA cytosine methylation, histone acetylation, chromatin remodeling, and also involves the ncRNAs [61]. In mammals, DNA methylation is involved in normal cellular control of expression, and hypermethylation can lead to silencing of tumor-suppressor genes in carcinogenesis [62].

In recent years, growing evidence has emerged that non-coding RNAs play essential roles in the regulation of gene expression in plants and animals, vertebrates and invertebrates [50,63,64]. MiRNAs can negatively control their target gene expression posttranscriptionally. Recently, the expression of miRNAs has been linked to cancer development, and miRNA profiles can be used to classify human cancers. In mammals, as in *C. elegans*, miRNAs can function to prevent cell division and drive terminal differentiation. An implication of this hypothesis is that down regulation of some miRNAs might play a causal role in the generation or maintenance of tumors [65].

The concept of epigenesis also extends to DNA and RNA editing. RNA editing is an epigenetic regulatory mechanism that was discovered from the unicellular protozoa *Trypanosoma* mitochondria. A number of genes are expressed in an unconventional manner, the nucleotide sequence of primary transcripts is modified post-transcriptionally through the insertion or deletion of Uridine. These nucleotide alteration was coined as RNA editing [66,67]. RNA editing has been detected in unicellular and multicellular eukaryotes but not in prokaryotes. After this discovery, it was thought that this process affects only mRNAs, but now it is known that the editing also occurs in tRNAs, rRNAs and miRnas [68–72]. In humans RNA editing is a change of adenosine to inosine mediated by the enzyme adenosine deaminase, acting on double-stranded RNA, where the inosine acts as guanosine [70,71]. In

mammals another kind of RNA editing has also been described consisting in a change of cytosine to uridine [73].

RNA editing expands the possibilities for expression of the epigenome by the production of different proteins from a single structural gene [73]. The RNA editing involves not only post-transcriptional changes but also phenotypic changes and therefore greater phenotypic plasticity of the organism against its environment. Thus, RNA editing generates variation of the epigenome contributing to the adaptation of organisms to their environments. In plants, there are data that show that RNA editing mostly affects evolutionarily conserved RNA codon position. These findings support the hypothesis that natural selection has contributed to selective fixation of certain RNA editing sites [74]. In animals and in particular in mammals, RNA editing is especially active in the brain, altering codons in mRNAs. RNA editing and other functional ncRNAs could be involved in diseases and also in brain development, brain plasticity and brain evolution [75].

The epigenome is also very important for genomic imprinting consisting in an asymmetric expression of genes with different paternal origins. Thus, genes are silenced when inherited via sperm or via egg. The term is used in relation to genes that are maternal inherited but silenced when paternally inherited. Thus, genomic imprinting is an epigenetic mechanism that induces parental specific gene expression in diploid eukaryotic cells [76,77]. The genes belonging to both mother and father can be expressed in a cell. However, due to imprinting, the expression of some genes are restricted to only one of the two parental chromosomes causing sex-specific changes in gene expression or chromosome behavior. The gametic DNA methylation mark is then maintained on the maternal or paternal allele [78]. These allelic marks are transmitted through the generations and do not follow a Mendelian pattern of inheritance.

Genomic imprinting changes gene expression between parents without altering DNA sequence. However, these DNA sequences are crucial to score an imprinting domain that are composed of non-coding DNA sequences. The missing of these sequences implies the lost of genome imprinting [50,64]. NcRNA such as, snoRNAs, miRNAs, piRNAs, and siRNA are very important in genomic imprinting. SiRNA is a highly conserved post-transcriptional silencing mechanism in which double stranded RNAs are processed to form guides for the degradation of complementary RNA transcripts through an RNA silencing complex [79–81]. The production of non-coding RNA has been described at multiple imprinted regions in both mammals and plants [81,82].

Much evidence has accumulated showing that imprinted genes can influence animal behavior. A paradigm in this way in humans are the Angelman and Prader-Willi syndromes that included neuroendocrine problems. Thus, children who inherit a deletion on chromosome 15 in the same locus from their father show a behavior different from these children that inherit these same altered locus from their mother [83]. Also in Turner Syndrome, girls that inherit their single X chromosome from their mother present different social dysfunction that who inherit the X chromosome from their father [84]. In mice it was described that genomic imprinting is related with the maternal care of offspring [85].

2.4. The Non-Coding RNAs and Their Relationships with Viruses

RNA viruses have a simple organization. They carry short RNA sequences, which are surrounded only by a protein or lipoprotein capsule and do not have a metabolic system that allows them to reproduce itself autonomously. Indeed they do need the host for full metabolic activity. Viroids and

virusoids are even simpler because the RNA is devoid of lipids and proteins and has no information for encoding proteins, and their genomes are similar in a conserved sequence to group I of introns that are found in nuclear rRNA genes, mitochondrial mRNA, rRNA genes, and chloroplast tRNA genes. The hallmark of these elements is a 16 nucleotide phyllogenetically conserved RNA sequence [86].

According to their structure, viruses are more similar to a set of molecules than with an organism, placing them among the living and the nonliving. However, viruses have a hereditary material, an universal genetic code and structural genes similar to bacteria, archaea and eukarya, but they have not been considered in the origin of the first life-forms on our planet, nor as a domain in the tree of life. One reason is the dependence of the metabolic system of a host cell for their survival and reproduction. Additional evidence for not having a "virus domain" includes the idea that viruses are derived from the genome of eukaryotes, and have a very unstable genome [87]. All this lead viruses to be disregarded as important elements from the neo-Darwinian evolutionary theory and also in the endosymbiotic theory of Lyn Margulis. However, recently there are many convincing arguments on the influence of viruses not only in the root of the tree of life but also in all its ramifications of the different domains of bacteria, eukaria and archeae, calling for a need to include viruses in the tree of life [88–90].

Viruses can recombine their genetic material with each other as was classically demonstrated by Benzer [2,3] using mutants of bacteriophage T4. Genic recombination in RNA viruses has been known to occur in the poliovirus since about fifty years ago and has been used as a model to investigate molecular mechanisms of recombination in a single stranded genome of positive polarity [91]. Recombination plays an important role in the evolution of RNA viruses by generating genetic variation, by reducing mutational load and by producing new viruses by recombination between different strains [92].

Genetic recombination is a universal phenomenon in all living systems and is also one of the fundamental conditions of sexuality. In the most simplistic sense, sexuality is the transfer and recombination of genetic material between organisms of a species. From this point of view, sexuality occurs between the genetic material of different strains of viruses and also between viruses and the genetic material of their hosts.

2.5. Sexuality Between Viruses and Related Genetic Elements

Classically it has been demonstrated that viruses may be intermediaries in the sexuality of strains of *Salmonella typhimurium* [93]. The material transferred by viruses (transduction) is incorporated into the host bacterium and inherited through generations. It has been shown that transduction of genes may be restricted or generalized. Restricted transduction occurs in occasions when the virus vectors are inserted into specific places of host chromosome and can transduce genes or very specific pieces of DNA adjacent to the insertion site [94]. Such type of transduction is also described in bacteriophage λ and *E. coli* K12 strain. In this case, λ is inserted between the gal and biotin genes, where it is as a prophage, but when it disintegrates the host genome gal of K12 gene it is part of the DNA of the virus and can be transduced to a different strain of bacteria lacking this gene when infected by the virus. In generalized transduction viruses prophages are inserted anywhere along the genome and can transduce any gene of the host. An example of generalized transduction is what happens in lysogenic strains of *Salmonella typhimurium* attacked by phage P22 [94].

Horizontal gene transfer mediated by viruses and bacteria have been detected in eukaryotes. In *Saccharomyces cerevisiae*, recombination between retrotransposons is a source of chromosome rearrangements and a mechanism of genome evolution in this specie [95]. Recent studies show recombination between retroposons and exogenous retroviruses [96]. Extensive sequence similarity in various organisms showed that the capsid protein and RNA-dependent RNA polymerase genes from viruses have widespread homologies transversally in the genomes of eukaryotic organisms. Sequence comparison and phylogenetic analysis suggest that these genes were likely transferred horizontally from viruses to eukaryotic genomes and are also functional in the recipient genomes. Horizontal transfer of double-stranded RNA viral genes is widespread among eukaryotes and may give rise to functionally important new genes, thus entailing that RNA viruses may play significant roles in the evolution of eukaryotes [97].

In particular, the persistent endogenous retroviruses in a long evolutionary time scale on a specific host can explain the origin of numerous basic and highly complex functions of life such as the origin of DNA-based genomes and DNA replication, eukaryotic nucleus, adaptative immune system and interference RNAs and mammalian placenta and viviparous birth [88,98] and also all the different class of ncRNAs.

3. Conclusions

The affinity in terms of molecular structure, transmission and recombination of genetic material between viruses, transposons, and introns demonstrates the viral origin of the latter elements and strengthen the hypothesis of an initial world of RNAs. It is known that many retrotransposons and retroviruses differ only in the absence of a capsule. I concord with the idea that RNAs viruses should be seen as the first manifestations of life which culminated in the biochemical evolution in a naked RNA world that preceded the origin of the first cells. With the discovery of the ncRNAs and ribozymes, these were located in the center of the central dogma of molecular biology displacing the DNA. The most plausible hypothesis now in the emergence of the first molecules pioneers that life revolves around an RNA virus not dependent of DNA in their replication. Many of the functions within living cells such as replication, transcription and repair as well as their fine-tuned regulatory order are now known to also be of viral origin [80,99].

If the origin of introns and retrotransposable elements are viral, then ncRNAs, in base to its homology with these elements, have also a viral origin and are part of the genome in the different species of the tree of life. Thus, the ncRNAs, in the form of introns, by splicing of mRNA, regulates the genetic expression. Also, the other forms of small and ncRNAs, coded in the redundant DNA and located between structural genes, have many important roles in the organism.

The molecular evidence demonstrates that the ncRNAs are scattered in the species that constitute the three domains of the tree of life. These ncRNAs are endosymbiont coadapted molecules with the genome of hosts and are the product of molecular coevolution from the origins of the first cells. The importance of RNA and DNA viruses in the evolution of their host as a persistent evolutionary force has been analyzed extensively by Villarreal [88,98], this analysis revealing that not only the vertical Mendelian inheritance has been important in the organic evolution but also the horizontal virus-mediated gene transfers have been a fundamental evolutionary force in the diversification of

species and biodiversity. These horizontal gene transfers may even be the cause of disease. In fact, increasing evidence for a relationship between carcinogenesis and infections by helminths has accumulated during the last decades; trematode infection could affect the host genomic stability and promote activation of introns of host genes involved in tumor progression [23].

The natural viral transduction between the species that constitute the tree of life and genetic homologies between distant organisms suggest that the genome of living organisms, particularly multicellular is a "fluid mosaics of genetic information from different sources" [10]. This fact also suggests that all organisms are naturally transgenic.

The discovery of ncRNAs has represented a major shift in the way of conceiving genetic variation in natural populations. Before the advent of the ncRNAs, the norm of reaction and phenotypic plasticity were considered adaptive but it was thought they had no a hereditary basis. But, now we know that these phenomena have a strong epigenetic base and can be inherited, explaining the emergence of morphological adaptations such as camouflage and mimicry. The epigenesis may also explain the variation and heredity of complex traits and diseases in humans [100].

According to Lynn Margulis the most important force in organic evolution is the endosymbiosis where bacteria have played a fundamental role in the origin of mitochondria and the chloroplasts. But, the roles of ncRNAs, their relation to RNA viruses embedded in the genome of species belonging to the three domains of the tree of life show that the endosymbiotic model must be extended to viruses. The phylogenetic analysis of relationships between RNA virus sequences and their hosts reveal coespeciation, products of thousands of years of ecological interaction between them. For these reasons it is necessary that the study of genome evolution, including human evolution, must include viruses and other parasites that also evolved from the same environments [101].

Conflict of Interest

The author declares no conflict of interest.

Acknowledgments

References

1. Crick, F. Central dogma of molecular biology. *Nature* **1970**, *227*, 561–563.
2. Benzer, S. Fine structure of a genetic region in bacteriophage. *Science* **1955**, *41*, 344–354.
3. Benzer, S. On the topology of the genetic fine structure. *Proc. Natl. Acad. Sci. USA* **1959**, *45*, 1607–1620.
4. Yanofsky, C. Gene structure and protein structure. *Sci. Am.* **1967**, *216*, 80–94.
5. Taft, R.J.; Pheasant, M.; Mattick, J.S. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **2007**, *29*, 288–299.
6. Mattick, J.S. Non-coding RNAs: The architects of eukaryotic complexity. *EMBO Rep.* **2001**, *2*, 986–991.

7. Frías, L.D. *La hetrocromatina y su rol Functional: ¿Qué es un gen? : Desde el Dogma Central de la Biología Molecular hasta la Secuenciación del Genoma Humano (in Spanish)*; RIL Editores: Santiago, Chile, 2004; pp. 1–148.

8. Frías, D.L. The history of the mendelian gene. *Riv. Biol.* **2007**, *100*, 69–92.

9. Frías, L.D. The phylogeny of the gene, from "Lower" naked structural genes to "Higher" non-transcriptional genes. *Riv. Biol.* **2007**, *100*, 221–246.

10. Frías, L.D. Omissions in the synthetic theory of evolution. *Biol. Res.* **2010**, *43*, 299–306.

11. Pearson, H. What is a gene? *Nature* **2006**, *441*, 399–401.

12. Parris, G. A Hypothetical master development program for multi-cellular organisms: Ontogeny and phylogeny. *Bioscience Hypotheses* **2009**, *2*, 3–12.

13. Gilbert, W. The RNA world. *Nature* **1986**, *319*, 618–618.

14. Mercer, T.R.; Dinger, M.E.; Mattick, J.S. Long non-coding RNAs: Insights into functions. *Nat. Rev. Genet.* **2009**, *10*, 155–159.

15. Eddy, S.R. Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2001**, *12*, 919–929.

16. Mattick, J.S.; Makunin, I.V. Non-coding RNA. *Hum. Mol. Genet.* **2006**, *1*, 17–29.

17. Berget, S.M.; Moore, C.; Sharp, P.A. Spliced segments at the 5′ termine of adenoviruses 2 late mRNA. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 3171–3175.

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

19. Wightman, B.; Ha, I.; Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *Caenorhabditis elegans*. *Cell* **1993**, *75*, 855–862.

20. Teixeira, A.; Tahiri-Alaoui, A.; West, S.; Thomas, B.; Ramadass, A.; Martianov, I.; Dye, M.; James, W.; Proudfoot, N.J.; Akoulitchev, A. Autocatalytic RNAcleavage in the human beta-globin pre-mRNA promotes transcriptiontermination. *Nature* **2004**, *432*, 526–530.

21. Salehi-Ashtiani, K.; Lupták, A.; Litovchick, A.; Szostak, J.W. A genomewide search for ribozymes reveals an HDV-like sequence in the human *CPEB3* gene. *Science* **2006**, *313*, 1788–1792.

22. Martick, M.; Horan, L.H.; Noller, H.F.; Scott, W.G. A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature* **2008**, *454*, 899–902.

23. de la Peña, M.; García-Robles, I. Intronic hammerhead ribozymes are ultraconserved in the human genome. *EMBO Rep.* **2010**, *11*, 711–716.

24. Prody, G.A.; Bakos, J.T.; Buzayan, J.M.; Schneider, I.R.; Bruening, G. Autolytic processing of dimeric plant virus satellite RNA. *Science* **1986**, *231*, 1577–1580.

25. Palmer, J.D.; Logsdon, J.M., Jr. The recent origins of introns. *Curr. Opin. Genet. Dev.* **1991**, *4*, 470–477.

26. Cavalier-Smith, T. Intron phylogeny: A new hypothesis. *Trends Genet.* **1991**, *7*, 145–148.

27. Filée, J.; Forterre, P.; Lauren, J. The role played by viruses in the evolution of their hosts: A view based on informational protein phylogenies. *Res. Microbiol.* **2003**, *154*, 237–243.

28. Zheng, Z.M. Viral oncogenes, noncoding RNAs, and RNA splicing in human tumor viruses. *Int. J. Biol. Sci.* **2010**, *6*, 730–755.

29. Zheng, Z.M.; Baker, C.C. Papillomavirus genome structure, expression, and post-transcriptional regulation. *Front. Biosci.* **2006**, *11*, 2286–2302.

30. Xiong, Y.; Eickbush, T.H. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. *Mol. Biol. Evol.* **1988**, *5*, 675–690.

31. Fischer, M.G.; Suttle, C.A. A virophage at the origin of large DNA transposons. *Science* **2011**, *332*, 231–234.

32. la Scola, B.; Desnues, Ch.; Pagnier, I.; Robert, C.; Barrassi, L.; Fourmous, G.; Merchant, M.; Susan-Monti, M.; Forterre, P.; Koonin, E.; *et al*. The virophage as a unique parasite of the giant mimivirus. *Nature* **2008**, *455*, 100–104.

33. Diener, T.O. Potato spindle tuber virus. IV. A replicating, low molecular weight RNA. *Virology* **1971**, *45*, 411–428.

34. Navarro, B.; Flores, R. Chrysanthemum chlorotic mottle viroid: Unusual structural properties of a subgroup of self-cleaving viroids with hammerhead ribozymes. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 11262–11267.

35. Diener, T.O. Circular RNAs: Relics of precellular evolution. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 9370–9374.

36. Smalheiser, N.R.; Torvik, V.I. Alu elements within human mRNAs are probable micro RNA targets. *Trends Genet.* **2006**, *22*, 532–536.

37. Smalheiser, N.R.; Torvik, V.I. Mammalian microRNAs derived from genomic repeats. *Trends Genet.* **2005**, *21*, 322–326.

38. Berger, A.; Strub, K. Multiple roles of *Alu*-related non-coding RNAs. *Prog. Mol. Subcell. Biol.* **2011**, *51*, 119–146.

39. Stort, G. An expanding universe of noncoding RNAs. *Science* **2002**, *296*, 1260–1263.

40. Ryan, J.; Taft, R.J.; Pang, K.C.; Mercer, T.R.; Dinger, M.; Mattick, J.S. Non-coding RNAs: Regulators of disease. *J. Pathol.* **2002**, *220*, 126–139.

41. Sánchez, L. Sex determining mechanism in insects. *Int. J. Dev. Biol.* **2008**, *52*, 837–856.

42. Taft, R.J.; Pang, K.C.; Mercer, T.R.; Dinger, M.; Mattick, J.S. Non-coding RNAs: Regulators of disease. *J. Pathol.* **2010**, *220*, 126–139.

43. Munker, R.; Calin, G.A. MicroRNA profiling in cancer. *Clin. Sci.* **2011**, *121*, 141–158.

44. Backofen, R.; Bernhart, S.H.; Flamm, C.; Fried, C.; Fritzsch, G.; Hackermuller, J.; Hertel, J.; Hofacker, I.L.; Missal, K.; Mosig, A.; *et al*. RNAs everywhere: Genome-wide annotation of structured RNAs. *J. Exp. Zool. B* **2007**, *308*, 1–25.

45. Smalheiser, N.R.; Lugli, G. MicroRNA regulation of synaptic plasticity. *Neuromol. Med.* **2009**, *11*, 133–140.

46. Bredy, T.W.; Lin, Q.; Wei, W.; Baker-Andresen, D.; Mattick, J. MicroRNA regulation of neural plasticity and memory. *Neurobiol. Learn. Mem.* **2011**, *96*, 89–94.

47. Smalheiser, N.R.; Lugli, G.; Thimmapuram, J.; Cook, E.H.; Larson, J. Endogenous siRNAs and noncoding RNA-derived small RNAs are expressed in adult mouse hippocampus and are up-regulated in olfactory discrimination training. *RNA* **2011**, *17*, 166–181.

48. Cazalla, D.; Yario, T.; Steitz, J.A. Down-regulation of a host microRNA by a *Herpesvirus saimiri* noncoding RNA. *Science* **2010**, *328*, 1563–1566.

49. Pijlman, G.P.; Funk, A.; Kondratieva, N.; Leung, J.; Torres, S.; van der Aa, L.; Liu, W.J.; Palmenberg, A.C.; Shi, P.Y.; Hall, R.A.; *et al*. A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. *Cell Host Microbe* **2008**, *4*, 579–591.

50. Koerner, M.V.; Pauler, F.M.; Huang, R.; Barlow, D.P. The function of non-coding RNAs in genomic imprinting. *Development* **2009**, *136*, 1771–1783.

51. Satterlee, J.S.; Barbee, S.; Jin, P.; Krichevsky, A.; Salama, S.; Schratt, G.; Wu, D.Y. Noncoding RNAs in the brain. *J. Neurosci.* **2007**, *27*, 11856–11859.

52. Tian, D.; Sun, S.; Lee, J.T. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* **2010**, *143*, 390–403.

53. Dinger, M.E.; Amaral, P.P.; Mercer, T.R.; Pang, K.C.; Bruce, S.J.; Gardiner, B.B.; Askarian-Amiri, M.E.; Ru, K.; Soldà, G.; Simons, C.; *et al*. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res*. **2008**, *18*, 1433–1445.

54. Thakur, N.; Tiwari, V.K.; Thomassin, H.; Pandey, R.R.; Kanduri, M.; Gondor, A.; Grange, T.; Ohlsson, R.; Kanduri, C. An antisense RNA regulates the bidirectional silencing property of the Kcnq1 imprinting control region. *Mol. Cell. Biol.* **2004**, *24*, 7855–7862.

55. Kurakawa, R. Long noncoding RNAs as a regulator for transcription. *Prog. Mol. Subcell. Biol.* **2011**, *51*, 29–41.

56. Raman, R.P.; Kanduri, C. Transcriptional and post transcriptional programming by long noncoding RNAs. *Prog*. *Mol. Subcell. Biol*. **2011**, *51*, 1–27.

57. Azzalin, C.M.; Reichenbach, P.; Khoriauli, L.; Giulotto, E.; Lingner, J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **2007**, *318*, 798–801.

58. Schoeftner, S.; Blasco, M.A. Developmentally regulated transcription of mammalian telomeres by DNA dependent RNA polymerase II. *Nat. Cell Biol.* **2007**, *10*, 228–236.

59. Arora, R.; Brun, C.M.C.; Azzalin, C.M. TERRA: Long noncoding RNAs at eukaryotic telomeres. *Prog. Mol. Subcell. Biol.* **2011**, *51*, 65–94.

60. Orom, U.A.; Shiekhattar, R. Long non-coding RNAs and enhancers. *Curr. Opin Genet. Dev*. **2011**, *21*, 194–198.

61. Ahmad, A.; Zhang, Y.; Cao, X.F. Decoding the epigenetic language of plant development. *Mol. Plant* **2010**, *3*, 719–728.

62. Chuang, J.C.; Jones, P.A. Epigenetics and microRNAs. *Pediatr. Res.* **2007**, *61*, 24R–29R

63. Autran, D.; Huanca-Mamani, W.; Calzada, J.P.V. Genomic imprinting in plants: The epigenetic version of an Oedipus complex. *Curr. Opin. Plant Biol*. **2005**, *8*, 19–25.

64. Anaka, M.; Lynn, A.; McGinn, P.; Lloyd, V.K. Genomic imprinting in *Drosophila* has properties of both mammalian and insect imprinting. *Dev. Genes Evol.* **2009**, *219*, 59–66.

65. Lu, J.; Getz, G.; Miska, E.A.; Alvarez-Saavedra, E.; Lamb, J.; Peck, D.; Sweet-Cordero, A.; Ebert, B.E.; Mak, R.H.; Ferrando, A.A.; *et al*. MicroRNA expression profiles classify human cancers. *Nature* **2005**, *435*, 834–838.

66. Benne, R.; van den Burg, J.; Brakenhoff, J.P.J.; Sloof, P.; van Boom, J.H.; Tromp, M.C. Major transcript of the frameshifted *coxII* gene from Trypanosome mitochondria contains four nucleotides that are not incoded in the DNA. *Cell* **1986**, *46*, 819–826.

67. Benne, R. RNA editing in trypanosomes. *Mol. Biol. Rep.* **1992**, *16*, 217–227.

68. Lonergan, K.M.; Gray, M.W. Editing of transfer RNAs in *Acanthamoeba castellanii* mitochondria. *Science* **1993**, *259*, 812–816.

69. Rubio, M.A.T.; Pastar, I.; Gaston, K.W.; Ragone, F.L.; Janzen, C.J.; Cross, G.A.M.; Papavasiliou, F.N.; Alfonzo, J.D. An adenosine-to inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7821–7826.

70. Luciano, D.J.; Mirsky, H.; Vendetti, N.J.; Maas, S. RNA editing of a miRNA precursor. *RNA* **2004**, *10*, 1174–1177.

71. Liang, H.; Landweber, L.F. Hypothesis: RNA editing of microRNA target sites in human? *RNA* **2007**, *13*, 463–467.

72. Gott, J.M.; Emerson, R.B. Functions and mechanisms of RNA editing. *Annu. Rev. Genet.* **2000**, *34*, 499–531

73. Blanc, V.; Davidson, N.O. C-to-U RNA editing: Mechanisms leading to genetic diversity. *J. Biol. Chem.* **2003**, *278*, 1395–1398.

74. Jobson, R.W.; Qiu, Y.-L. Did RNA editing in plant organellar genomes originate under natural selection or through genetic drift? *Biol. Direct* **2008**, *3*, doi:10.1186/1745-6150-3-43.

75. Mehler, M.F.; Mattick, J.S. Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiol. Rev.* **2007**, *87*, 799–823.

76. Reik, W.; Walter, J. Genomic imprinting: Parental influence in the genome. *Nat. Rev. Genet*. **2001**, *2*, 21–32.

77. Úbeda, F.; Gardner, A. A model for genomic imprinting in the social brain: Juvenile. *Evolution* **2010**, *64*, 2587–2600.

78. Li, E.; Beard, C.; Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **1993**, *366*, 362–365.

79. Filipowicz, W. RNAi: The nuts and bolts of the RISC machine. *Cell* **2005**, *122*, 17–20.

80. Tang, G. siRNA and miRNA: An insight into RISCs. *Trends Biochem. Sci*. **2005**, *30*, 106–114.

81. Royo, H.; Cavaillé, J. Non-coding RNAs in imprinted gene clusters. *Biol. Cell* **2008**, *100*, 149–166

82. Zhang, Y.; Qu, L. Non-coding RNAs and the acquisition of genomic imprinting in mammals. *Sci. China C Life Sci*. **2009**, *52*, 195–204.

83. María, L.V.S.; Currotto, B.L.; Cortés, F.M.; Rojas, C.B.; Alliende, M.A.R. Methylation, cytogenetic and FISH tests in the molecular diagnosis of Prader-Willi and Angelman síndromes. *Rev. Med. Chile* **2001**, *129*, 1–9.

84. Skuse, D.H.; James, R.S.; Bishop, D.V.M.; Coppin, B.; Dalton, P.; Aamodt-Leeper, G.; Bacarese-Hamilton, M.; Creswell, C.; McGurk, R.; Jacob, P.A. Evidence from Turner's Syndrome of an imprinted X-linked locus affecting cognitive function. *Nature* **1997**, *387*, 705–708.

85. Wilkins, J.F.; Haig, D. Inbreeding, maternal care and genomic imprinting. *J. Theor. Biol.* **2003**, *221*, 559–564.

86. Dinter-Gottlieb, G. Viroids and virusoids are related to group I introns. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 6250–6254.

87. Moreira, D.; Lópes-García, P. Ten reasons to excluded viruses from the tree of life. *Nat. Rev. Microbiol.* **2009**, *7*, 306–311.

88. Villarreal, L.P. How viruses shape the tree of life. *Future Virol.* **2006**, *1*, 587–595.

89. Villarreal, L.P.; Witzany, G. Viruses are essential agents within the roots and stem of the tree of life. *J. Theor Biol.* **2010**, *262*, 698–710.

90. Mindell, D.P.; Rest, J.S.; Villarreal, L. Viruses the Tree of Life. In *Assembling the Tree of Life*; Cracraft, J., Donoghue, M.J., Eds.; Oxford University Press: New York, NY, USA, 2004; pp. 107–118.

91. Combelas, N.; Holmblat, B.; Joffret, M.L.; Colbére-Garapin, F.; Delpeyroux, F. Recombination between poliovirus and cocxackie A viruses of species C: A model of viral genetic plasticity and emergence. *Viruses* **2011**, *3*, 1460–1484.

92. Worobey, M.; Holmes, E.C. Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.* **1999**, *80*, 2535–2543.

93. Zinder, N.D.; Lederberg, J. Genetic exchange Salmonella. *J. Bacteriol.* **1952**, *64*, 679–699.

94. Levine, L. Bases físicas da Herença. In *Biologia do Gene (in Portuguese)*; Levine, L., Ed.; Edgard Blucher: São Paulo, Brazil, 1977; pp. 41–49.

95. Mieczkowski, P.; Francene, A.; Lemoine, J.; Thomas, D.; Pete, T.D. Recombination between retrotransposons as a source of chromosome rearrangements in the yeast *Saccharomyces cerevisiae*. *DNA Repair* **2006**, *5*, 1010–1020.

96. Geuking, M.B.; Weber, J.; Dewannieux, M.; Gorelik, E.; Heidmann, T.; Hengartner, H.; Zinkernagel, R.M.; Hangartner, L. Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. *Science* **2009**, *323*, 393–396.

97. Liu, H.; Fu, Y.; Jiang, D.; Li, G.; Xie, J.; Cheng, J.; Peng, Y.; Ghabrial, S.A.; Yi, X. Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. *J. Virol.* **2010**, *84*, 11876–11887.

98. Villarreal, L.P. DNA Virus Contribution to Host Evolution. In *Origin and Evolution of Viruses*; Domingo, E., Webster, R., Hollabd, J., Eds.; Academic Press: New York, NY, USA, 1999; pp. 391–420.

99. Villarreal, L.P. *Viruses and the Evolution of Life*; American Society for Microbiology Press: Washington, DC, USA, 2005; p. 472.

100. Petronis, A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* **2010**, *465*, 721–727.

101. van Blerkom, L.M. Role of viruses in human evolution. *Yearb. Phys. Anthropol.* **2003**, *46*, 14–46.