



## The contribution of RNAs and retroposition to evolutionary novelties

Jürgen Brosius

*Institute of Experimental Pathology, ZMBE, University of Münster, Von-Esmarch-Str. 56, Münster, Germany  
(Phone: +49-251-835-8511; Fax: +49-251-835-8512; E-mail: RNA.world@uni-muenster.de)*

*Key words:* genetic novelties, genome evolution, genome expansion and shrinkage, retroposition, survival of reverse transcriptase

### Abstract

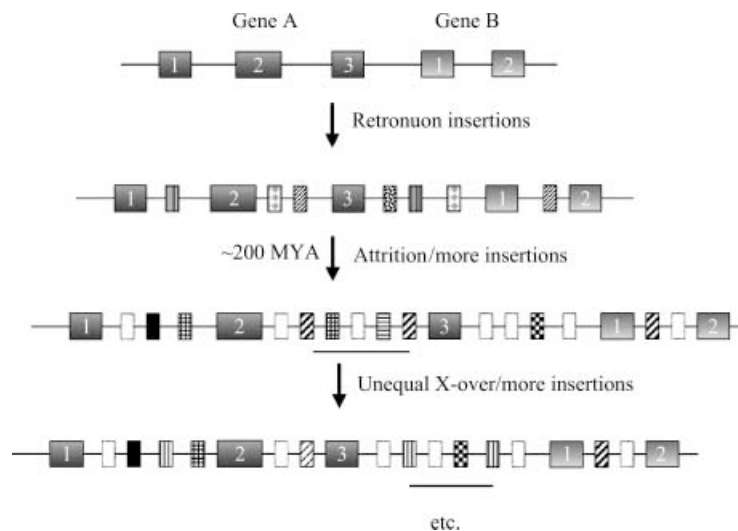
Retroposition is an ancient process dating back to the conversion of RNA to DNA genomes. Nevertheless, it continues to make tremendous structural and functional contributions to extant genomes. This process and the endurance, or even renaissance, of an RNA world in many lineages sheds a new light on the Central Dogma of Molecular Biology. The question of why reverse transcriptase has survived billions of years without an apparent cellular function is discussed. Retroposition constitutes one of the pervasive conflicts, in this case between host genome on one hand and mobile genetic elements on the other, that fuel the evolutionary process. It is obvious that retroposition has, thus far, contributed numerous useful novelties to genomes.

### Modern genomes contain archives of by-gone eras

One interesting outcome of the current genome projects is the realization that many lineages still bear molecular evidence of earlier evolutionary stages. For example, the combination of genome sequencing, computational analysis and experimental approaches has revealed that many Eucarya still exhibit a lively RNA world within their cells (Martignetti & Brosius, 1993b; Brosius & Tiedge, 1995b; Jeffares, Poole & Penny, 1998; Poole, Jeffares & Penny, 1998; Brosius, 1999d, 2002a; Herbert & Rich, 1999; Eddy, 2001, 2002; Erdmann et al., 2001; Hüttenhofer & Brosius, 2002; Pasquinelli, 2002; Storz, 2002). Likewise, many genomes bear witness to the process that was responsible for transforming an RNA genome to a DNA genome as well as the surprising fact that this process is by no means concluded (Brosius & Tiedge, 1995b; Jurka, 1998; Brosius, 1999a). Around 42% of the human genome contains sequences whose origins can still be traced to retroposition. Retroposition is a general genomic process that converts RNAs into DNA with the aid of reverse transcriptase (Weiner, Deininger & Efstratiadis, 1986). Any type of cellular RNA can serve as template for reverse transcription (Brosius,

1999c). The resulting cDNA copies are integrated on a more or less random basis into the genome leaving hallmarks such as an adenosine-rich region at their 3'-ends and direct repeats of some 5–15 bp flanking the inserted fragments. When transcripts with protein coding regions serve as templates it is almost exclusively the mature mRNA and not the intron-containing hnRNA form that is used. Hence, a third hallmark for a subclass of sequences derived by retroposition from intron-containing source genes is the lack of such introns (Weiner, Deininger & Efstratiadis, 1986).

The fact that almost seven sixteenths of the total genome are discernible retronuons [a nuon is any definable nucleic acid sequence (Brosius & Gould, 1992, 1993; Brosius, 2003b)] also means that they arose sometime in the last 200 million years (MYA). Furthermore, there is no reason to assume that in the intervening period, between the onset of generating DNA genomes a couple of billion years ago until ~200 MYA, there was no retroposition. However, we can no longer discern individual retronuons unless they were exapted into a function and their sequences were conserved via positive selection. Often, only short motifs have been conserved while the remainder of the retronuon has been mutated beyond recognition. This leads to another important realization, namely



*Figure 1.* Model for expansion and contraction of genomes. Two genes, A with three exons (dark gray boxes) and B with two exons (light gray boxes) are depicted. Control regions are not shown. The second line shows the same locus after retronuons including LINEs and SINEs (rectangles with different patterns reflecting different retronuon types, not drawn to scale) integrate, usually in intergenic regions and introns. With time, retronuon sequences deteriorate (white rectangles with interrupted outlines) while at the same time other retronuons insert (third line). Intergenic regions and introns are shortened by illegitimate recombination and unequal homologous recombination, for the latter case often between retronuons of the same class (lines 3 and 4). At the same time more retronuons insert (line 4).

that retroposition has not only made a sizable quantitative contribution to genomes but has also contributed to the evolution of functional elements and novel genes (Brosius, 1991, 1999c; Brosius & Gould, 1992; Brosius & Tiedge, 1995b).

If retroposition dates back a couple of billion years (at least 10–15 times longer than retronuons remain discernible), why are modern genomes not larger? The main reason is that there are various levels of tolerance in the domain of Eucarya (Petrov, 2001). While vertebrates are generally quite permissive for extra DNA, invertebrates can vary between being highly restrictive and highly permissive. For example, *Drosophila* has a 60-fold faster DNA loss than mammals, resulting in a much more compact genome (Petrov, 2001). In contrast, most organisms of the bacterial and archaeal domains lack or suppress retroposition quite efficiently. For the sake of completeness, it should be mentioned that, of course, segmental duplication also plays an important role in genome expansion. As with retroposition, periods of relative quiescence alternate with periods of enhanced activity. In the primate lineage leading to humans, there was a great spurt of segmental duplication less than 35 MYA (Eichler, 2001a, b; Johnson et al., 2001; Samonte & Eichler, 2002). As a result, more than 10% of human chromosome 22 and 5% of the genome in general consists of discernible

segmental duplications (Bailey et al., 2002; Eichler, 2001a). For a long time, it was thought that whole genome and segmental gene duplication was the only route for generating duplicate genes (Bridges, 1936; Muller, Prokofyeva-Belgovskaya & Kossikov, 1936; Ohno, 1970; Sturtevant, 1925).

### Genomes are subject to antagonistic forces of growth and reduction

Mammalian genomes are quite similar in size. From comparisons of the mouse chromosome 16 sequence with syntenic regions on human chromosomes, mouse and human are estimated to vary by about 10% (Mural et al., 2002). The major sources of this variation are SINE (short interspersed repetitive elements) and LINE (long interspersed repetitive elements) retronuons, which account for  $21.7 + 12.3 = 34\%$  of euchromatic regions of mouse chromosome 16 and for  $31.6 + 16.4 = 48\%$  of the total sequence in the corresponding regions in humans; genomewide, the figures were estimated as 36% and 46%, respectively (Mural et al., 2002).

Despite the massive onslaught of retronuon integration and a sizable level of segmental duplication, an apparently optimal genome size is maintained in mammals by an approximate equilibrium between

amplification and deletion of genetic material. A closer look at non-coding loci supports collateral sequence gain and loss (Kurychev et al., 2001). The mechanisms leading to deletions were, until recently, not considered as often as the ones leading to genome expansions (Petrov et al., 2000; Comeron, 2001). Interestingly, retroposition itself can set the stage for deletions between retronuons of similar sequences by unequal homologous recombination (Batzer & Deininger, 2002). However, illegitimate recombination appears to play a superior role in genome contractions (Devos, Brown & Bennetzen, 2002). A hypothetical scenario depicting continued expansion of retronuon families, their mutational randomization as well as their deletions is shown in Figure 1. The retropositional bombardment of a locus surrounding the gene encoding neuronal BC200 small non-messenger RNA (snmRNA), documents how intergenic regions are subject to change among members of a single mammalian order, in the relatively short time span of a few dozen million years (Kurychev et al., 2001).

Among vertebrates, the pufferfish genomes of *Fugu rubripes* and *Tetraodon nigroviridis* are unusually small; consisting of 400 million bp, about one quarter of the more typical zebrafish genome and one eighth of a mammalian genome. The pufferfish genome sequence will be valuable for establishing the underlying causes of their unusually compact genomes. Is it due to a lack of or reduced retroposition, or increased deletion or a combination of both?

### Young genes – one of the results of genomic upheaval

The generation of new genes is one of the consequences of genomes that are constantly in flux. In the future, recently created genes will certainly receive more attention. The availability of complete genomic sequences from mammals both more distantly and more closely related to man, such as mouse, rhesus monkey and chimpanzee, will be of great value in determining which protein- and RNA-encoding genes and their regulatory elements differ from species to species and are, thus, responsible for the phenotypic differences among species.

However, it will be no small feat to sort out the genes that are unique to one species. This is illustrated by the recent comparison between complete sequences of mouse chromosome 16 and regions of conserved synteny on human chromosomes 3, 8, 12, 16, 21

and 22 (Mural et al., 2002). In mouse, initially 1055 genes were predicted with high to medium confidence. Gene predictions with weak confidence were omitted. After closer examination of genes that may have been split or merged in the autoannotation process and after eliminating pseudogenes and viral-related sequences, 731 genes remained. Of those, 509 have orthologs on syntenic blocks in the human genome, 44 are likely paralogs to these genes and 164 have homologues elsewhere in the human genome. Hence, only 14 genes were identified that were present only in mouse and not in humans and 21 genes *vice versa*.

Extrapolating the length of euchromatin of mouse chromosome 16 to the entire genome, I arrive at about 630 genes that are present only in mice but not in humans and about 429 genes *vice versa* totaling >1000 genes that arose or disappeared (as a total) since the two lineages split. Alternatively, extrapolating from the 731 *bona fide* genes on mouse chromosome 16, to estimates of 30,000 or 35,000 or 40,000 total genes in mammals, the number of genes that arose or disappeared would be approximately 1436, 1675 or 1915, respectively. Copeland, Jenkins & O'Brien (2002) arrive at a much smaller number of about 469 genes when 'extrapolating across the entire genomes and presuming a 90-million year interval since mouse and human shared an ancestor [...], this means that one new gene arose or disappeared on average every 192,000 years'. With the uncertainty of information, the margin of error is already great. Alone among the 44 likely paralogs and 164 homologs that are not counted as potential differences between the two genomes (see above), some might actually correspond to non-orthologous genes that have been recruited or exapted into novel functions. It is also not unlikely that, in addition to the assumed 731 *bona fide* genes, one or more dozen genes with lower confidence will turn out to be true genes (see below for examples of true genes that initially were categorized as retropseudogenes). As a consequence, the number of 'true' or *bona fide* genes on mouse chromosome 16 may actually include many more that are young and specific to one or the other lineage, thus also boosting the role of these young genes in species differentiation over the entire genome.

Much more information and a thorough analysis will be necessary to arrive at more reliable figures. Nevertheless, the above estimates already fall within a reasonable range as they approach estimates on gene duplications in the primate lineage. With the caveat that this is an extrapolation from chromosome 22

to the entire genome, Bailey et al. estimate (Bailey et al., 2002) that about 500–1,100 transcripts may have been created or modified from duplicated sequences during the past 35 MYA. Humans and chimpanzees have been estimated to differ by 150–350 transcripts (Bailey et al., 2002). Assuming 1675 genes that arose or disappeared in the mouse and human lineages over a divergence time of about 90-MYA, the 6 MYA divergence time assumed between chimpanzee and human would predict about 110 gene differences.

### **What remains constant in genomes with continuous flux?**

Apart from exons and conserved regulatory elements in a sea of genomic change, are there other more stable sequences in genomes? A comparison of mouse chromosome 16 with regions of conserved synteny on human chromosomes 3, 8, 12, 16, 21 and 22, revealed close to 12,000 ‘syntenic anchors’, short stretches of DNA sequences conserved in both species (Mural et al., 2002). Such ‘syntenic anchors’ are not low copy number genomic repeats as no other matches in the genome were found. About 44% of the mouse chromosome 16 ‘syntenic anchors’ are outside the limits of known genes. Likewise, a comparison of human chromosome 20 with corresponding regions of mouse and pufferfish (*T. nigroviridis*) identified a large number of RSCs (regions of sequence conservation) in intergenic regions (Deloukas et al., 2001). It will be interesting to determine what proportion of these extra-genic anchors or RSCs will turn out to correspond to as yet unknown protein coding genes, conserved control elements or conserved non-messenger RNA species. Surprisingly though, analogous ‘syntenic anchors’ or RSCs could not be detected when comparing even closely related genomes in plants (J.L. Bennetzen, personal communication).

### **Retronuons encoding active reverse transcriptase: engines of retroposition**

LINE retronuons play a central role in retroposition. They lack LTRs (long terminal repeats) and a fraction of the full-length copies exhibit intact open reading frames featuring an endonuclease and reverse transcriptase domain (Ostertag & Kazazian, 2001). They thereby provide not only some of the enzymatic machinery for their own retroposition but also that for

short retronuons, SINEs and possibly mRNAs (Okada et al., 1997; Weiner, 2002). In addition to specialized functions such as X-chromosome inactivation (Lyon, 1998, 2000; Bailey et al., 2000), LINE retroposition has also recently been implicated in repair of double-stranded breaks (Morrish et al., 2002). Interestingly, LINEs can move exons that are located 3' to their original insertion points around the genome, perhaps to be recruited by the genes near their new loci of integration (Moran, DeBerardinis & Kazazian, 1999; Goodier, Ostertag & Kazazian, 2000).

Furthermore, transduction of a cellular RNA by a human endogenous retrovirus, HERV, has recently been shown for the first time. Although this alternative route, which involved template switching to a cellular mRNA encoding *FAM8AI* only lead to inactive genes, it does demonstrate the potential to virally transduce sequences within the same genome or between individuals of the same or different species (Jamain et al., 2001).

### **‘Junked-up’ genomes**

Less than 2% of mammalian genomes consist of protein coding regions. A further small percentage is assumed to be occupied by regulatory elements. Most of the remainder of the genomes has been contributed by retroposition (see first paragraph). Is all this junk? If a distinction between garbage and junk is made, whereby garbage is something you dispose of and junk is something you keep in the attic, garage or basement until you either dispose of it later or re-use it, perhaps for a different purpose than the original, it is quite acceptable to refer to the non-gene coding DNA, including retronuons, as junk (Brenner, 1998; Gould, 2002). Just like people, genomes differ widely in their predilections for tidiness or for keeping junk. The relative benefit of such preferences is hard to predict. The inhabitants of a house that save loads of old books in the attic may, unwillingly, contribute to the house burning down in case of an otherwise manageable fire or make an unpredictable great profit a few decades later should these books become valuable to collectors.

### **Retroposition does not always lead to inactive retronuons. Are these rare or frequent events?**

However, retroposition can provide more than just junk, like segmental duplications, it also leads to the

generation of novel genes. About 15–25% of all genes in the human genome lack introns in their coding regions (J.J. Emerson & M. Long, personal communication). In addition, many genomes exhibit large amounts of inactive intronless retrogenes, so-called retropseudogenes perhaps equivalent in number to 25–50% of the number of all active genes (Betran et al., 2002). Does that mean that as many as one in three retroposition events involving mature mRNAs lead to active genes? Certainly, a 30% success rate appears to be much too high because (i) it is not clear yet what proportion of genes never had introns and how many are ‘successful’ products of retroposition; and (ii) only those inactive retrogenes generated and/or temporarily functioning within the last 200 MYA, and not removed in the interim by recombination (Devos, Brown & Bennetzen, 2002), are still discernible. All these inactive retrogenes bear witness to the fact that, unlike segmental duplications, only a small proportion of retroposition events led to active genes.

The odds of successful retroposition appear to be even less favorable for snmRNA-derived retronuons. The one million copies of Alu elements in the human genome compared with the paucity of actively transcribed Alu elements, such as BC200 RNA, give an impression of how rare the haphazard juxtaposition of snmRNA-derived retronuons with functional promoter elements might be. Likewise, of the estimated 130,000 ID retronuons in rat, only a handful are source genes (Kim et al., 1994). For a more detailed discussion of the interrelationship of source genes and SINEs in case of BC1 and BC200 RNAs see below.

When from a hundred thousand to a million retronuons only about a handful are transcriptionally active, retroposition does not appear to be a very efficient route for generating novel snmRNA genes. The ratio of exaptations to non-exaptations (Gould & Vrba, 1982) may be significantly higher if one includes retronuon-derived regulatory elements. The poor yield derived from generating active snmRNA genes by retroposition contrasts with the observation that about a quarter of all human genes are intronless and that a significant fraction thereof have been generated by retroposition (see above). There are a number of explanations for this apparent discrepancy: (i) Most intronless genes are ancient genes that remained intronless after introns arose in an ‘intron late’ scenario (Patthy, 1991); (ii) The retrogene hallmark ‘intronless’ survives longer than any of the other hallmarks [A-tail at the 3'-end and short direct repeats flanking the insertion (Weiner, Deininger & Efstratiadis, 1986)]; hence, a retronuon-

derived snmRNA gene or regulatory element is hardly discernible after 150–200 MYA; (iii) There are large intronless families of genes that are clustered. They include members of the interferon and olfactory receptor families and arose by segmental duplication of intronless genes; (iv) Many more SINE retronuons can initially be transcribed, especially as they already carry part of the internal promoter elements, with them. Because of the lack of selective pressure to maintain further versions of the SINE master gene, the promoter soon becomes deactivated by mutation and transcription subsides.

In conclusion, not all of the estimated 5–10,000 intronless genes arose directly by retroposition. Furthermore, not every SINE retronuon was dead on arrival, but many were silenced after a few million years if not under selective pressure to be actively transcribed. On the other hand, along the ever growing list of recognized, active, intronless genes that were generated by retroposition over the last 100–200 MYA (e.g., see <http://exppc01.uni-muenster.de/expath/alltables.htm#table3>, <http://exppc01.uni-muenster.de/expath/alltables.htm#table4>, <http://exppc01.uni-muenster.de/expath/alltables.htm#table5>, <http://exppc01.uni-muenster.de/expath/alltables.htm#table6>, <http://exppc01.uni-muenster.de/expath/alltables.htm#table7>) supports the notion that the number of intronless genes that arose over a time span, at least an order of magnitude longer, could be sizable. Finally, a number of genes that initially were categorized as retropseudogenes based simply on the fact that they exhibit some or all of the hallmarks, were subsequently shown to be *bona fide* genes (Koller & Strehler, 1988; Yaswen et al., 1992; Noyce & Piper, 1994; Noyce, Conaty & Piper, 1997).

### **Retroposition – high risk, high return**

Retroposition keeps the evolutionary wheels of fortune turning. Although, as argued above, the odds for a win are apparently not very high, the potential gains can be considerable. Despite low odds for creating active genes instead of pseudogenes by retroposition, this route of gene amplification has decisive advantages over other forms of gene duplications. Many active retrogenes recombined at sites of integration with existing regulatory elements. This differs from segmental duplications that require subsequent changes in their regulatory elements so that the duplicated genes can be expressed

differentially from the parental genes. For example, point mutations and/or deletions/insertions (including retrons) near regulatory regions must occur so as to generate different control regions and thus alter temporal and/or spatial expression patterns. In contrast, a significant proportion of retrogenes may exhibit different expression patterns from those of their respective founder genes from the start of their active life. Examples have been reviewed previously (Brosius, 1991, 1999b,c,d; Brosius & Gould, 1992; Brosius & Tiedge, 1995a,b; Britten, 1996, 1997; Makalowski, 2000) and can be found on our www page <http://exppc01.uni-muenster.de/expath/alltables.htm#table3>.

Over the years, it has also become more and more apparent, that retroposition, especially when snmRNA templates are the substrate for retroposition, can contribute regulatory elements to existing genes (Brosius & Gould, 1992; Brosius, 1999a; Makalowski, 2000). Examples range from Alu retrons as splice sites (Nekrutenko & Li, 2001; Sorek, Ast & Graur, 2002) to B2 SINE retrons as promoters (Ferrigno et al., 2001). For further such examples see: <http://exppc01.uni-muenster.de/expath/alltables.htm#table2>.

In addition, parts of retrons that are related to viruses have contributed significantly to regulatory elements, for example, lone LTRs were often exapted as promoter elements (<http://exppc01.uni-muenster.de/expath/alltables.htm#table1>); additional cases can be found in the following publications (Kurdyukov et al., 2001; Medstrand, Landry & Mager, 2001; Schön et al., 2001; Costas, 2002; Paces, Pavlicek & Paces, 2002; Vinogradova et al., 2002).

### **Examples of relatively young, retropositionally-derived, snmRNA genes**

Retroposition of snmRNAs, or even large non-messenger RNAs, can lead to the generation of new genes encoding non-messenger RNAs. Two of the best-studied examples are BC1 RNA and BC200 RNA. See Table 1 for a summary of their characteristics. Previously, it had been assumed that many non-messenger RNAs of the cell were remnants from the RNA world, a time when both phenotype and genotype resided in nucleic acids (Woese, 1967; Brosius, 2003b), and evolution had not yet 'gotten around' to replacing them. However, both BC1 and BC200 RNAs possess very specific, specialized expression patterns in an organ system that is not very ancient.

Furthermore, the expression of BC1 RNA, which originated by retroposition of a tRNA<sup>Ala</sup>, is restricted to the mammalian order Rodentia, which may lead some to believe that it could not be important functionally. Nevertheless, its sequence is conserved when compared to flanking regions and it is the first known RNA polymerase III transcript expressed almost exclusively in nerve cells, where it is found even in distal parts of dendrites (DeChiara & Brosius, 1987; Tiedge et al., 1991; Martignetti & Brosius, 1993b, 1995). BC200 RNA has a very similar morphological, and subcellular distribution as BC1 RNA but is found only in all anthropoid primates examined, but not in prosimians (Martignetti & Brosius, 1993a; Tiedge, Chen & Brosius, 1993; Skryabin et al., 1998). The age we can infer is between 35–55 MYA (Kurychev et al., 2001). The founder of BC200 RNA is a monomeric Alu element, in turn derived from SRP RNA (Ullu & Tschudi, 1984; Quentin, 1994) and hence, BC1 and BC200 RNAs are not homologues but conceivably functional analogues. Their location suggests roles in either transport of cellular components into dendrites or translational regulation of dendritic mRNAs. Interestingly, their respective ribonucleoprotein particles both contain poly(A) binding protein (Muddashetty et al., 2002; West et al., 2002), favoring the latter.

Studies to delineate the function of very young gene products are challenging yet rewarding. They are challenging because it is likely that the functional contributions are subtle, as other lineages function perfectly well without the respective gene products. In gene deletion experiments using knockout mice, the mixed genetic background, alone, of the gene-depleted mice may conceal subtle differences. Nevertheless, functional studies involving young genes are of great importance as we may begin to understand the mechanisms that lead to exaptation of duplicated genes into novel functions.

For example, we generated mice that lack the gene encoding neuronal BC1 RNA (B.V. Skryabin et al., manuscript in preparation). These mice do not exhibit any obvious neuroanatomical or morphological changes. Apart from the fact that BC1 RNA is not expressed, the expression patterns and subcellular distributions of other dendritic RNAs examined are unchanged. Thus BC1 RNA does not appear to be involved in the transport of dendritic mRNAs.

In a large scale project involving several hundred BC1-negative mice from three lines derived from three ES stem cells featuring the same BC1 gene deletion,

Table 1. Characteristics of neural BC1 and BC200 RNAs

BC1 RNA	BC200 RNA
All rodents <sup>1</sup>	Anthropoidea (primates) <sup>2,3</sup>
152 nt non-mRNA, polIII <sup>4,5</sup>	200 nt non-mRNA polIII <sup>6,7</sup>
A single active gene 60–110 MYA old <sup>4,5</sup>	A single active gene 35–55 MYA old <sup>2,3,6–9</sup>
Ancestor: tRNA <sup>Ala</sup> <sup>4</sup>	Ancestor: monomeric FLAM-C Alu element <sup>10</sup>
Master gene of the ID <sub>1</sub> repetitive element sub-family <sup>11,12</sup>	Master gene of >200 BC200 retropseudogenes <sup>3</sup>
Secondary structure of 5' domain: stable stem <sup>13</sup>	Secondary structure of 5' domain as in SRP <sup>2,14</sup>
8.7S RNP <sup>15,16</sup>	11.4S RNP <sup>17</sup>
Proteins: PABP <sup>18,19</sup> , La (transient) <sup>20</sup>	SRP9/14 <sup>14</sup> , PAPBP <sup>18,19</sup> , La (transient) <sup>20</sup>
Neurons – not glia, distal dendrites <sup>21–25</sup>	Neurons – not glia, distal dendrites <sup>7</sup>
Testes (low level) <sup>26</sup>	Testes (low level)
Developmentally regulated <sup>23</sup>	?
Activity dependent expression <sup>27</sup>	?
Expression deregulated in immortalized cell lines and tumors <sup>28</sup>	Expression deregulated in immortalized cell lines and tumors (potential marker) <sup>29</sup>
KO mice: reduced exploratory behavior <sup>30</sup>	

<sup>1</sup> Martignetti and Brosius (1993b). <sup>2</sup> Skryabin et al. (1998). <sup>3</sup> Kurychev et al. (2001). <sup>4</sup> DeChiara and Brosius (1987). <sup>5</sup> Martignetti and Brosius (1993a). <sup>6</sup> Martignetti and Brosius (1995). <sup>7</sup> Tiedge, Chen and Brosius (1993). <sup>8</sup> Taylor et al. (1997). <sup>9</sup> Basile et al. (1998). <sup>10</sup> Brosius (1999c). <sup>11</sup> Kim et al. (1994). <sup>12</sup> Deininger et al. (1996). <sup>13</sup> Rozhdestvensky et al. (2001). <sup>14</sup> Kremerskothen et al. (1998b). <sup>15</sup> Kobayashi, Goto and Anzai (1991). <sup>16</sup> Cheng, Tiedge and Brosius (1996). <sup>17</sup> Cheng, Tiedge and Brosius (1997). <sup>18</sup> West et al. (2002). <sup>19</sup> Muddashetty et al. (2002). <sup>20</sup> Kremerskothen et al. (1998a). <sup>21</sup> Tiedge et al. (1991). <sup>22</sup> Tiedge, Chen and Brosius (1992). <sup>23</sup> Lin, Brosius and Tiedge (2001). <sup>24</sup> Brosius and Tiedge (2001). <sup>25</sup> Brosius and Tiedge (1995a). <sup>26</sup> Muslimov et al. (2002). <sup>27</sup> Muslimov et al. (1998). <sup>28</sup> Chen et al. (1997b). <sup>29</sup> Chen et al. (1997a). <sup>30</sup> Skryabin et al. (2003).

we carried out a regimen of health checks and behavioral tests at three different locations including one semi-naturalistic setting. When compared with control mice of the same hybrid genetic background, the BC1 knock-out mice tended to display a reduction in exploratory behavior in several tests. Currently, we are examining the underlying biochemical changes that are precipitated by the lack of BC1 RNA expression.

The lack of, or impeded function of, BC200 RNA in humans might lead to a more obvious phenotype, possibly exhibiting apparent behavioral or other neuropsychiatric abnormalities. However, in this case we are restricted to the 'experiments' of nature. We examined a Norwegian family in which linkage analyses identified a region on chromosome 2p15-p16, which cosegregated with dyslexia (Fagerheim et al., 1999). This *DYX3* locus maps closely to the one that has been mapped for BC200 RNA on human chromosome 2 (Basile et al., 1998). However, we did not find any abnormalities in any of the samples of affected family members within the 200 nt-long, BC200 RNA coding region, or in the putative upstream regulatory region (M. Kiefmann, T. Fagerheim & J. Brosius, unpublished results). Currently, we are collecting samples from patients with various neuropsychiatric diseases.

While we are convinced that deletion or mutation in the BC200 RNA gene will have a notable phenotype in humans, we may have to wait for a 'lucky' coincidence to find such a gene defect.

Another young snmRNA is 4.5S<sub>I</sub> RNA, a ubiquitously expressed RNA polymerase III transcript about 100 nt in length. It is related to rodent B2 ret-ronuons, which in turn were derived from a tRNA<sup>Lys</sup> (Okada, 1991). In comparison to BC1 RNA, 4.5S<sub>I</sub> RNA has a much narrower species distribution as it is detected in only four myomorph families (Muridae, Cricetidae, Spalacidae and Rhizomyidae) but not, for example, in Dipodidae, Zapodidae, Geomyidae, or Heteromyidae. Consequently, it cannot be much older than 25–30 MYA. Yet, as in BC1 RNA, the coding sequence is conserved while the flanking regions diverge in the expected manner. This indicates that 4.5S<sub>I</sub> RNA is also under positive selection (Gogolevskaya & Kramerov, 2002).

#### Out of a plethora of novel snmRNAs many might be evolutionarily young but of unknown biogenesis

The number of known snmRNAs is steadily increasing due to biocomputational as well as experimental ap-

proaches to RNomics (Storz, 2002). One of the fastest growing subfamilies comprises small nucleolar RNAs (snoRNAs). An outcome of the systematic search for additional snoRNA family members is the realization that targets of interaction (e.g., for modifications or other types of processing) are not merely ribosomal RNAs (rRNAs) but other cellular RNAs including perhaps mRNAs (Cavaillé et al., 2000; Filipowicz, 2000; Hüttenhofer et al., 2001). Several snoRNAs that are prevalently expressed in the brain and paternally imprinted map to human chromosome 15q11-13 in the Prader-Willi (PWS) region. Alterations in this labile region are causative of this neurodevelopmental disease (Cavaillé et al., 2000; Runte et al., 2001). While we have no indication of a retropositional of these particular genes, some of which are arranged in up to 50 repeat units, their origin might just predate some of the eutherian radiation as homologues have not, as yet, been found in non-mammalian vertebrates or even in marsupials (Cavaillé et al., 2001).

Another snoRNA, RBII-36 can be detected only in rat species *sensu lato* but not in mouse (Cavaillé et al., 2001). RBII-36 is located in an intron of exons that are apparently not translatable. As is the case with the snoRNAs in the PWS region, units are repeated several times in tandem. Apart from an internal A-rich sequence within the coding region of RBII-36, the sequenced repeat units do not feature hallmarks of retroposition. In a scenario, where the snoRNA was generated by retroposition and subsequent segmental duplication to create the 0.9 kb repeat units, the direct repeats should be discernible, in particular, when considering that the events took place rather recently. A number of examples have been reported, where snoRNAs could serve as templates for retroposition (Bachelierie, Cavaillé & Hüttenhofer (2002). However, even if snoRNAs are poor templates for retroposition, if they land within the introns of genes transcribed by RNA polymerase II and are subsequently processed into mature snoRNAs governed by their own internal structural elements, they stand a much better chance of exaptation than either the reverse transcribed snoRNAs that land in intergenic regions or retrons derived from other small RNAs. This is because the first still require external promoters and the second must use both the external processing *and* promoter signals from the locus of integration, which are rarely present or appropriately juxtaposed with respect to the inserted retrons. It will be interesting to further address the biogenesis of young snoRNA by phylogenetic studies of the loci involved.

Last year witnessed an explosive growth in the number of known micro RNAs (~22 nt long) in organisms as diverse as worms, flies, mammals and plants (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee & Ambros, 2001). These are regulatory RNAs involved in post-transcriptional processes in different cell types at various developmental stages (Lagos-Quintana et al., 2002; Lai, 2002; Pasquinelli, 2002). It would not be surprising if some of these RNAs turned out to be products of young genes. In fact, some of the miRNA precursors appear to be restricted to vertebrates (T. Tuschl, personal communication).

In addition to small non-messenger RNAs, genomes encode large RNAs that do not serve as mRNAs. A number of them may even contain features of mRNAs, such as caps, A-tails, and processing from hnRNA-like progenitors. The sole function of a sub-class of these RNAs is to serve as precursors harboring snoRNAs in introns. Among other steps, snoRNAs are liberated by the processing of these precursors into mRNA-like RNAs (Tycowski, Shu & Steitz, 1996; Pelczar & Filipowicz, 1998; Cavaillé et al., 2000, 2001, 2002; Tycowski & Steitz, 2001). Other RNAs might have more direct functions. H19 RNA, itself imprinted, is oncofetally expressed (Ariel, de Groot & Hochberg, 2000) and under stabilizing selection (Hurst & Smith, 1999). Although its precise function is still enigmatic, it has been implicated in oncogenesis or tumor suppression, and as a riboregulator of gene expression (Lottin et al., 2002). *Xist* and *Tsix* RNAs have been shown to function during inactivation of X chromosomes (Avner & Heard, 2001; Boumil & Lee, 2001). Likewise, in flies the non-messenger RNAs *roX1* and *roX2* are involved in dosage compensation (Amrein & Axel, 1997; Meller et al., 1997; Meller & Rattner, 2002). The list of longer, non-messenger RNAs is rapidly growing. They include transcripts from the Ultrabithorax domain of the bithorax complex (Lipshitz, Peattie & Hogness, 1987), a polyadenylated, synapse-associated RNA transcribed from the *7H4* gene (Velleca, Wallace & Merlie, 1994), a 3.7 kb testis-specific RNA near the *SOX9* gene (Ninomiya et al., 1996), a 17 kb polyadenylated transcript from the human *NTT* gene (Liu et al., 1997), the spliced and polyadenylated transcripts from the bone morphogenetic, protein-responsive gene, *BORG* (Takeda et al., 1998) and the prostate-specific *DD3* gene (Bussemakers et al., 1999). Many of the longer non-messenger RNAs appear to act as antisense RNAs, such as *DISC2* that is antisense to *DISC1*, both of which are disrupted

in schizophrenia (Millar et al., 2000); the Wilms tumor suppresser gene *WT1* antisense transcript (Malik et al., 2000), *LIT1*, an imprinted antisense RNA in the human *KvLQT1* locus (Mitsuya et al., 1999; Horike et al., 2000); the imprinted *Air* RNA at the *Igf2r* locus (Lyle et al., 2000) which is involved in autosomal gene silencing (Sleutels, Zwart & Barlow, 2002) and the *SCA8* transcript that is expanded by (CUG)<sub>n</sub> in spinocerebellar ataxia type 8 and is antisense to a brain-specific actin-binding protein, *KLHL1* (Nemes et al., 2000; Benzow & Koob, 2002). The fission yeast polyadenylated meiRNA appears to target the otherwise cytoplasmic Mei2 protein to the nucleus simply by binding the locus where it promotes the first meiotic division (Watanabe & Yamamoto, 1994; Yamamoto, 1996; Yamashita et al., 1998). Recently, a 17.5 kb transcript without apparent protein coding capacity, *Ks-1* has been detected in the nuclei of honey bee neurons (Sawata et al., 2002). It might be involved in dosage compensation. The fact that it had been identified in *Apis mellifera* and *A. cerana* but not in *Drosophila melanogaster* might be an indication of a more recent evolutionary origin (Sawata et al., 2002).

Recently, Wang et al. (2002) discovered a large RNA, sphinx (Spx), in *D. melanogaster*. An intronless, retroposed copy of the ATP synthase chain F gene contributed a large portion of this RNA. *Spx* is absent in sibling species of *D. melanogaster* and hence is maximally 2–3 MYA old. During that time, *spx* has maintained well-defined splice sites, which would be lost in a non-functioning pseudogene but features substitution rates higher than a non-functional sequence, both suggesting a positive selection for novel functions (Wang et al., 2002).

### How did non-protein coding mRNA-like transcripts arise?

Many long non-messenger RNAs still contain structural or processing features of mRNAs. For example, their genes may still contain introns and the mature RNAs are polyadenylated and spliced from precursors. It is important to note that mRNAs do not only encode information for an open reading frame of a protein but they encode various other functions, especially in their 5'- and 3'-untranslated regions (Rastinejad & Blau, 1993; Rastinejad et al., 1993; Davis & Watson, 1996) Hence the frequent designation of non-messenger RNAs as non-coding RNAs (ncRNAs) is misleading. Additional roles of mRNA domains include the provision of binding sites for proteins,

signals for subcellular localization as well as providing sequences that permit interaction with chromatin. Just as non-messenger RNAs can transport a protein to a cellular location it would otherwise not occupy, RNAs are a viable way to bring, solely via their RNA binding domains, two or more proteins of different functions together into the same subcellular domains or environments. Such proteins would otherwise not be able to interact without first evolving the appropriate protein/protein interactions and possibly not even be in the same subcellular domain in the first place. This alone would be 'justification' enough not to abandon the continuing evolution of large or small non-messenger RNA molecules. Presumably some of the large RNAs once had dual functions that included templating for protein biosynthesis *and* structural and processing functions. Removal of selective pressure for the protein coding function, however, did not necessarily make the RNA superfluous and hence the other function(s) encoded on the RNA still might be under positive selection. Even if the encoded protein is indispensable, segmental duplication or retroposition are prime sources for generating additional transcripts. These may lose the protein coding capacity by mutationally compromising the open reading frame, but the RNA retains other functions. Due to the elimination of one function, the other(s) may be enhanced or altered by subsequent positive selection. In conclusion, often it might be premature to entitle a transcribed gene that lost its protein coding capacity as a 'pseudogene'.

### How selfish are retrons?

Since the early eighties, the quality of selfishness (or, label of *selfish*) has been closely associated with retrons such as LINEs and SINEs (Doolittle, 1980; Orgel & Crick 1980). For LINEs the verdict is still out as to what degree they are selfish. LINE retrons are, barring horizontal transfer as stowaways in retroviral particles, autonomous for intracellular propagation only. In addition to an internal transcription promoter, a portion of full length LINEs encode at least one intact polypeptide with reverse transcriptase and DNA endonuclease activity. However, if the sole 'purpose' of a LINE would be to produce more viable LINE elements, why are more than 98–99% of all L1 retroposons truncated and thus non-viable? Furthermore, of the 3000–5000 full-length elements per diploid human genome, only about 100 may be retropositionally competent (Ostertag & Kazazian, 2001).

Is it a reduction of 'virulence' to secure survival of the host and consequently its own evolutionary success or do LINE elements serve some other function? LINES have also been implicated in the repair of either single- or double-strand DNA lesions (Morrish et al., 2002), however, in this case, it would not matter whether the resulting LINE is full length or truncated. Even shorter SINEs that use the enzymatic machinery of LINES for retroposition would fulfill the purpose of 'stuffers' in DNA repair.

With respect to the selfishness of SINE retronuons, it is still a prevailing myth that every retroposed SINE is transcriptionally active, simply because they contain intact internal RNA polymerase III promoter elements. The opposite is true: only a tiny fraction of SINEs escapes the fate of 'death on arrival'. The exceptional transcriptional activity of a few retroposed SINEs may be a result of (i) the fortuitous acquisition of upstream RNA polymerase III promoter elements essential for efficient transcription near the locus of integration; (ii) integration into a locus with a reduced methylation status that does not block transcription or (iii) integration near elements that neutralize the negative impact of methylation on transcription. Experimental evidence (Martignetti & Brosius, 1995) favors the former scenario. The best explanation for the persistence of SINE retronuons is that the source genes transcribe RNAs that are under positive selection because they have a function in the cell and that the RNAs happen to be efficient templates for reverse transcription. An alternative explanation is that certain transiently active RNA molecules have features that allow them, in comparison to other small RNAs, an extremely high degree of 'reverse transcribability'. As a consequence, the next generations of retronuons might have produced, among the many failed attempts, at least a few transcribed retrogenes. Some of them have the ability to serve as source genes, themselves, in parallel to the initial founder gene. With several source genes in action, the loss of one of them, including the initial founder gene, by silencing due to lack of function and selective pressure can be compensated. This highly inefficient approach would secure SINE survival until the reigning source gene(s) lost the ability to produce enough retronuons for a seamless succession of transcriptionally active and retropositionally prodigal source genes. Was this the fate of mammalian MIR retronuons (Smit & Riggs, 1995) or did the function of their respective source genes become dispensable?

The evolutionary history of neuronal BC1 RNA and BC200 RNA as source genes for ID repetitive elements and BC200-like monomeric Alu elements, respectively, may serve as examples for the scenarios discussed (DeChiara & Brosius, 1987; Martignetti & Brosius, 1993a; Kim et al., 1994; Deininger et al., 1996; Brosius, 1999c). tRNAs have the potential to serve as templates for dozens, even hundreds of retronuons. BC1 RNA arose by retroposition of a tRNA<sup>Ala</sup> (DeChiara & Brosius, 1987). In rodents, BC1 RNA can be an efficient template for retronuons, and in the rat genome is present as an ID<sub>1</sub> subtype in about 10,000 copies (Kim et al., 1994). However, in some rodent species, for example again in the rat, there are at least three additional subfamilies of this SINE, namely ID<sub>2-4</sub>. These retronuons are relatively young and did not arise from transcripts of BC1 RNA, whose gene is present as a single active copy. Some of the ID elements must have been (or still are) transcriptionally active and have served as very efficient founder genes to expand ID elements in the rat genome by about 120,000 members of the ID<sub>2-4</sub>-types. Whether the transcripts that gave rise to the three additional ID subfamilies stem from three separate source genes or whether the sequence of one source gene changed over time with changes reflected in the three ID subtypes remains to be determined. Likewise, it is not clear yet, whether or not any of the ID transcripts had been exapted into a function. If so, their expression may be under selective pressure; if not they may document the relay example involving a succession of a few active retronuons among many dead ones as described above (see also Brosius (1999a)).

The interrelationship of Alu elements with neuronal BC200 RNA is equally complex. Although monomeric and especially dimeric Alu elements were, numerically, extremely successful in the order of Primates, there is, as of yet, little or no indication for a functional role of the transcripts derived from their source genes. An exception is BC200 RNA that arose from a left Alu monomer (FLAM-C) after Anthrozoidea diverged from prosimians about 35–55 MYA ago (Kurychev et al., 2001). The primary and secondary structure, CpG content, neuron-specific expression pattern and dendritic sub-cellular localization of BC200 RNA is conserved in New World and Old World monkeys and in apes. Over time, it served as a template for about 200 retronuons in the lineage that lead to *H. sapiens*. These BC200 retrogenes serve as a molecular archive documenting a

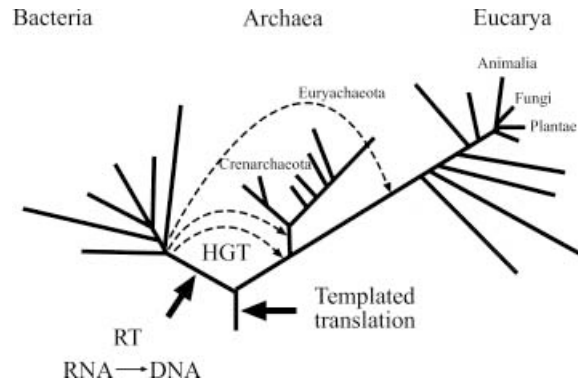
date line of the sequence changes the gene has experienced since its birth (Kurychev et al., 2001). None of these retronuons appears to be autonomously transcribed. As for any SINE, there always exists the possibility to be co-transcribed on a larger hnRNA unit. If BC200 RNA had not been exapted into a function, in all likelihood, it would have ceased to serve as a template for retronuons. Conversely, if BC200 RNA was a more efficient template for retronuons, it might not need a function to serve over a longer evolutionary period as a source gene. A few transcriptionally active members of its bountiful progeny would replace transcriptionally inactivated source genes.

### **An unsolved question in Biology**

One of the most interesting unsolved questions in Biology centers around the survival of reverse transcriptase and its gene(s) for possibly as long as three billion years, after having done its job of converting RNA genomes to DNA genomes during the transition from the RNP (RNA/protein) world to modern cells (Brosius, 1999d).

A major transition in evolution, the conversion from RNA to DNA genomes (Brosius, 1999d, 2003b), would not have been possible without retroposition. It is conceivable that subsequent major transitions (Szathmáry & Smith, 1995), such as the conversion from single cellular to multicellular organisms, evolution of the central nervous system, etc. might never have occurred (Brosius, 2003b). Reverse transcriptase has been described as a very ancient enzyme (Flavell, 1995). While it is conceivable that the initial role of reverse transcriptase, which arose fortuitously, was converting RNA to DNA, it is clear that its gene did not persist for the purpose of generating genetic novelties. Without a task in the cell, however, the persistence of the retroviral/retroposon-type reverse transcriptase is hard to imagine. While still being able to use any form of RNA as a template, one such task of the primordial form of the enzyme might have been the maintenance of telomeres. Later, perhaps, the enzyme specialized as telomerase and concurrently lost the ability to reverse transcribe other RNAs (Shippen-Lentz & Blackburn, 1990; Blackburn, 1991; Ware, Wang & Blackburn, 2000). By then a descendent of the primordial promiscuous reverse transcriptase might have begun its independent 'life' by being part of retroviruses, retrotransposons, retroelements

or similar elements without possessing a function in cellular metabolism. The strategy, of intragenic and intergenic transfer across species borders might have compensated the need for its continuous survival in the cell (Mouches, Bensaadi & Salvado, 1992; Kordis & Gubensek, 1998; Jordan, Matyunina & McDonald, 1999; Volff, Korting & Schartl, 2000; Zupunski, Gubensek & Kordis, 2001). An extreme Dawkinsian view (Dawkins, 1976) would be that cellular organisms are the survival machines of retroviruses, retrotransposons and autonomous retroelements. Leaving and re-entering the host coupled with its resistance to and evasion of purges is perhaps the secret of why reverse transcriptase survived billions of years without the selective pressure of an essential cellular function. However, the legacy of retroposition was not only the inflation of genomes in the numerous lineages where it acted. As a positive side effect, many genomes are, to a large degree (up to 90%, Brosius (2003a)), products of its activities. The increased plasticity of the affected genomes and recruitment of many retronuons as novel genes or their regulatory elements is an exaptation and, as it turns out, a valuable asset to their evolvability. Retronuons are often still considered to be solely parasites, at best nuisances. Nevertheless, retronuons have enabled a tremendous level of genomic flexibility. Perhaps, the action of retronuons and the ensuing defense mechanisms of genomes lead to strategies that were exapted as part of a genome's epigenetic instrumentation (J. McDonald, personal communication). The plasticity brought about by retronuons guaranteed long-term survival of species and lineages, while the lack of such plasticity would have meant stasis, possibly even extinction. It should be cautioned, however, that under different (environmental) conditions the burden of retronuons might have overridden their advantages and lineages that were permissive for retroposition might have vanished from the face of the earth. As remarkable as retronuons are in surviving perhaps billions of years, so phenomenal are, with hindsight, the opportunistic skills of hosts. Over evolutionary periods 'something' always happens to genomes. What makes the difference is whether 'something' is put to an advantageous use by a given lineage. Retroposition may be regarded as one of the compromises that abound in living systems (Williams, 1997) and should be judged with long-term hindsight only. The interaction of hosts with retroviruses, retrotransposons and retroelements is one of the eternal conflicts that drive the evolution of life (Brosius, 2003b).

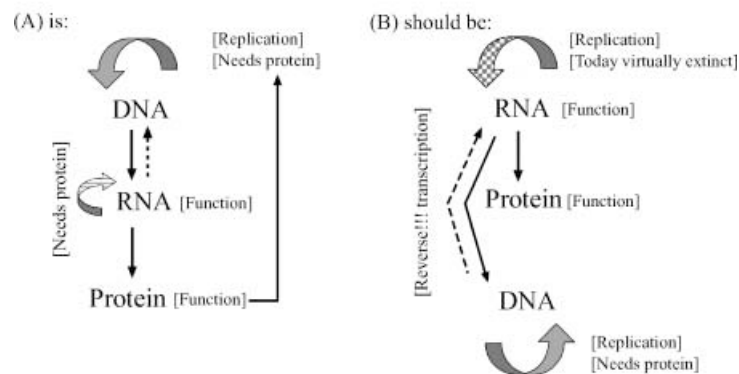


*Figure 2.* Onset of horizontal spread of reverse transcriptase-containing retronuons and resulting establishment of species barriers. Three domains of life, Bacteria, Archaea and Eucarya and some of the subdivisions such as archaeal kingdoms are indicated. Arrows denote the origin of templated translation and, tentatively, the origin of reverse transcriptase (RT). The initially primitive, templated translation generated more complex and interconnected systems. This transition was, according to Woese (2002), what triggered establishment of three separate lineages. An additional cementation of evolutionary barriers might have occurred by the sequential action of RT in the lineages that began to separate. The enzyme underwent horizontal gene transfer (HGT) with the aid of transposable elements (broken arrows). RT-containing retronuons remain, to this day, agents of HGT between more or less closely related branches on the universal phylogenetic tree (not shown).

Evolution abounds with, is even fueled by, such conflicts (Casti & Karlqvist, 1995; Brosius, 2003b). Examples include the evolution of multicellularity (Michod, 1996; Michod & Roze, 2001), host–parasite interactions (Renaud & de Meeus, 1991; Hurst, Atlan & Bengtsson, 1996), sex (Parker & Partridge, 1998; Partridge & Hurst, 1998), the debated case of parental gene imprinting (Hurst & McVean, 1998; Iwasa, 1998) as well as altruism and cooperation versus selfishness (Sober & Wilson, 1998). A remarkable case of cooperation has just been reported in the wood mouse, whereby sperm speed up by traveling in packs and sacrifice themselves for the benefit of the winner but against sperm from another male (Moore et al., 2002). One of the major eternal conflicts of evolution is preclusion versus integration. It appears as if nature wants to have the cake and eat it too: In 3–4 BYA of evolution there were always trends for biological systems to separate themselves from others. On the other hand, without extensive horizontal exchange of genetic information, life would probably not have been so successful (Brosius, 2003b; Woese, 2002). In order to aid preclusion, various barriers were established. We recognize the cell membrane, co-linear genomes, the Darwinian Threshold, Weismann’s Barrier and various epigenetic strategies, such as imprinting (Brosius, 2003b). Without these barriers, life as we know it would not have been possible. In addition to the cell membrane, a basic prerequisite for life, the concatenation of RNA nuons into a linear RNA genome might have established one of the first barriers against the free exchange of

genetic materials (Brosius, 2003b). The Darwinian Threshold, a hitherto unrecognized phase transition in the evolutionary process has been proposed by Carl Woese (Woese, 2002). It corresponds to the emergence of a new level of order in terms of cell organization. Cellular machineries became more and more complex and interconnected. As a consequence, the interchangeability of genetic material became gradually less feasible and thus curbed the acquisition of novelties via horizontal gene transfer. Consequently short-lived entities were replaced by more permanent ones – species. Hence, ‘the Darwinian Threshold truly represents the origin of speciation as we know it’ (Woese, 2002). In molecular terms, Woese sees this transition occurring when templated translation produced ever more sophisticated protein molecules including the ones that were involved in the translational machinery itself. In my opinion, additional barriers were important in establishing a molecular and cellular basis for the origin of species. It is conceivable that the linkage of individual RNA nuons as concatemers, precursors of chromosomes, already established a barrier against extensive horizontal RNA nuon transfer (Brosius, 2003b). After traversing the Darwinian Threshold (Woese, 2002), pre-existing barriers were further consolidated when RNA genomes were converted into DNA genomes. Typical for the volatile modes of evolution charged with conflict: the same mechanism did not completely shut the door on horizontal gene transfer, but even provided some alternative means through retroviral elements and other retronuons. The first occurrence of reverse

### Central Dogma of Molecular Biology Revisited



**Figure 3.** The Central Dogma of Molecular Biology revisited. The Central Dogma of Molecular Biology as it is interpreted today is shown on the left. It was amended several times to accommodate (i) the fact that RNA can be replicated (small curved arrow) without the involvement of DNA, (ii) reverse transcription (broken arrow) and (iii) the discovery that not only protein but also RNA can have functions other than encoding information. The restructured Central Dogma of Molecular Biology (right) is primarily based on the evolution of cells and their macromolecular components. In the past (RNA world), genomic RNA could be replicated without the aid of protein. RNA subsequently served as template for protein biosynthesis (RNP world) as it still does today and finally, DNA was generated from RNA templates (kinked arrow) with the aid of protein (reverse transcriptase). The latter process is still widespread, especially in the domain of Eucarya. Transcription of DNA into RNA (kinked broken arrow) again is achieved with the aid of protein. Today, direct replication of RNA with the aid of a replicase consisting of RNA is virtually extinct. However, with the inclusion of DNA  $\rightarrow$  RNA transcription one can state that replication of RNA occurs via a DNA state – just like most RNA viruses. In this restructured Central Dogma one would be tempted to term the process leading from RNA to DNA *transcription* and the one from DNA to RNA *reverse transcription*.

transcriptase might have been in the branch leading to Bacteria (Figure 2).

Although reverse transcriptase-containing retransposons have been described in Bacteria, their occurrence and actions are quite limited (Temin, 1989; Lampson, Inouye & Inouye, 1991; Inouye & Inouye, 1995). Generally, bacterial horizontal gene transfer is mediated via DNA, such as plasmids, phages and other episomes. It appears that Bacteria are the most ‘advanced’ of the three domains as they almost concluded the era of reverse transcription. That Bacteria are increasingly independent of RNA-based mechanisms, makes one wonder when species in the domain of Bacteria will abandon messenger RNA altogether. Apart from the potential difficulties of introducing multiple, simultaneous changes to adapt translation to DNA templates, there is little reason to exclude the possibility (perhaps in a couple hundred million years), that DNA might be translated directly. With coupled transcription/translation, ribosomes are already in ‘hot pursuit’ of the DNA template in Prokaryotes. Some remnants of the transcriptional apparatus, for example, for melting the ‘DNA template’, may be kept for translation.

In the remaining phyla, mobile elements harboring reverse transcriptase might have been transferred to the other budding branches (curved and dashed

arrows, Figure 2) and ‘locked in’ in some of the branching patterns leading to the extant domains. At least in the domain of Eucarya, there are indications for continued horizontal gene transfer activity involving elements carrying reverse transcriptase and occasionally other passengers between species within the various branches or even from branch to branch between more distantly related phyla.

#### Another amendment in the Central Dogma of Molecular Biology?

Considering the fact that in many branches of life the RNA world is not over yet, as evidenced by the pervasiveness of reverse transcription in modern genomes, the Central Dogma of Molecular Biology would, in my opinion, benefit from the inclusion of evolutionary considerations. The perceived sequence of events DNA  $\rightarrow$  RNA  $\rightarrow$  protein (Crick, 1958) merely reflects how cells operate to synthesize macromolecules. Since its inception, it required several amendments (Crick, 1970) (Figure 3(A)) and recently has been criticized for its narrow conception of informational flow in the cell, neglecting the complexity of regulatory networks (Thieffry & Sarkar, 1998). On an evolutionary level one might rather suggest a restate-

ment as RNA → protein → DNA, whereby in reality transcription involves the continuous conversion of RNA into DNA that is catalyzed by the protein reverse transcriptase (Figure 3(B)). Hence, the re-conversion of genomic DNA into messenger or other RNA would be more appropriately termed reverse transcription.

### Acknowledgements

This paper is a small tribute to the memory of Stephen Jay Gould (10.09.1941–20.05.2002). It may take decades to fully realize the magnitude of his loss. Thanks to Evan Eichler for comments on an earlier draft, Marsha Bundman for editing the manuscript and Martin Kiefman for help with the figures. I would also like to thank the participants of the Second Workshop of the European Science Foundation Network 'Retrotransposons: their impact on organisms, genomes and biodiversity' held in Helsinki, summer 2002, for lively discussions on some of the subjects presented here. The author is supported by the German Human Genome Project through the BMBF (#01KW9966).

### References

- Amrein, H. & R. Axel, 1997. Genes expressed in neurons of adult male *Drosophila*. *Cell* 88: 459–469.
- Ariel, I., N. de Groot & A. Hochberg, 2000. Imprinted H19 gene expression in embryogenesis and human cancer: the oncofetal connection. *Am. J. Med. Genet.* 91: 46–50.
- Avner, P. & E. Heard, 2001. X-chromosome inactivation: counting, choice and initiation. *Nat. Rev. Genet.* 2: 59–67.
- Bachellerie, J.-P., J. Cavallé & A. Hüttenhofer, 2002. The expanding snoRNA world. *Biochimie* 84: 775–790.
- Bailey, J.A., L. Carrel, A. Chakravarti & E.E. Eichler, 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *P. Natl. Acad. Sci. USA* 97: 6634–6639.
- Bailey, J.A., A.M. Yavor, L. Viggiano, D. Misceo, J.E. Horvath et al., 2002. Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. *Am. J. Hum. Genet.* 70: 83–100.
- Basile, V., A. Vicente, J.A. Martignetti, B.V. Skryabin, J. Brosius et al., 1998. Assignment of the human BC200 RNA gene (BCYRN1) to chromosome 2p16 by radiation hybrid mapping. *Cytogenet. Cell. Genet.* 82: 271–272.
- Batzler, M.A. & P.L. Deininger, 2002. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* 3: 370–379.
- Benzow, K.A. & M.D. Koob, 2002. The KLHL1-antisense transcript (KLHL1AS) is evolutionarily conserved. *Mamm. Genome* 13: 134–141.
- Betran, E., W. Wang, L. Jin & M. Long, 2002. Evolution of the phosphoglycerate mutase processed gene in human and chimpanzee revealing the origin of a new primate gene. *Mol. Biol. Evol.* 19: 654–663.
- Blackburn, E.H., 1991. Telomeres. *Trends Biochem. Sci.* 16: 378–381.
- Boumil, R.M. & J.T. Lee, 2001. Forty years of decoding the silence in X-chromosome inactivation. *Hum. Mol. Genet.* 10: 2225–2232.
- Brenner, S., 1998. Refuge of spandrels. *Curr. Biol.* 8: R669.
- Bridges, C., 1936. The Bar 'gene': a duplication. *Science*: 83: 210–211.
- Britten, R.J., 1996. DNA sequence insertion and evolutionary variation in gene regulation. *P. Natl. Acad. Sci. USA* 93: 9374–9377.
- Britten, R.J., 1997. Mobile elements inserted in the distant past have taken on important functions. *Gene* 205: 177–182.
- Brosius, J., 1991. Retroposons – seeds of evolution. *Science* 251: 753.
- Brosius, J., 1999a. Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica* 107: 209–238.
- Brosius, J., 1999b. Many G-protein-coupled receptors are encoded by retrogenes. *Trends Genet.* 15: 304–305.
- Brosius, J., 1999c. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* 238: 115–134.
- Brosius, J., 1999d. Transmutation of tRNA over time. *Nat. Genet.* 22: 8–9.
- Brosius, J., 2003a. Echoes from the past – are we still in an RNP world? (submitted).
- Brosius, J., 2003b. Gene duplication and other evolutionary strategies: from the RNA world to the future. *J. Struct. Funct. Genomics* 3: 1–17.
- Brosius, J. & S.J. Gould, 1992. On 'nomenclature': a comprehensive (and respectful) taxonomy for pseudogenes and other 'junk DNA'. *P. Natl. Acad. Sci. USA* 89: 10706–10710.
- Brosius, J. & S.J. Gould, 1993. Molecular constructivity. *Nature* 365: 102.
- Brosius, J. & H. Tiedge, 1995a. Neural BC1 RNA: dendritic localization and transport, pp. 289–330 in *Localized RNAs*, edited by H.D. Lipshitz & R.G. Landes, Austin, TX.
- Brosius, J. & H. Tiedge, 1995b. Reverse transcriptase: mediator of genomic plasticity. *Virus Genes* 11: 163–179.
- Brosius, J. & H. Tiedge, 2001. Dendritic BC1 RNA: intracellular transport and activity-dependent modulation, pp. 129–138 in *Cell Polarity and Subcellular RNA Localization*, edited by D. Richter. Springer Verlag, Berlin.
- Bussemakers, M.J., A. van Bokhoven, G.W. Verhaegh, F.P. Smit, H.F. Karthaus et al., 1999. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res.* 59: 5975–5979.
- Casti, J.L. & A. Karlqvist (Editors), 1995. *Cooperation and Conflict in General Evolutionary Processes*. John Wiley & Sons, Inc., New York.
- Cavallé, J., K. Buiting, M. Kiefmann, M. Lalonde, C.I. Brannan et al., 2000. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *P. Natl. Acad. Sci. USA* 97: 14311–14316.
- Cavallé, J., P. Vitali, E. Basyuk, A. Hüttenhofer & J.P. Bachellerie, 2001. A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a noncoding RNA gene in rats. *J. Biol. Chem.* 276: 26374–26383.
- Cavallé, J., H. Seitz, M. Paulsen, A.C. Ferguson-Smith & J.P. Bachellerie, 2002. Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum. Mol. Genet.* 11: 1527–1538.

- Chen, W., W. Böcker, J. Brosius & H. Tiedge, 1997a. Expression of neural BC200 RNA in human tumours. *J. Pathol.* 183: 345–351.
- Chen, W., J. Heierhorst, J. Brosius & H. Tiedge, 1997b. Expression of neural BC1 RNA: induction in murine tumours. *Eur. J. Cancer.* 33: 288–292.
- Cheng, J.G., H. Tiedge & J. Brosius, 1996. Identification and characterization of BC1 RNP particles. *DNA Cell Biol.* 15: 549–559.
- Cheng, J.G., H. Tiedge & J. Brosius, 1997. Expression of dendritic BC200 RNA, component of a 11.4S ribonucleoprotein particle, is conserved in humans and simians. *Neurosci. Lett.* 224: 206–210.
- Cameron, J.M., 2001. What controls the length of noncoding DNA? *Curr. Opin. Genet. Dev.* 11: 652–659.
- Copeland, N.G., N.A. Jenkins & S.J. O'Brien, 2002. Genomics. *Mmu* 16 – comparative genomic highlights. *Science* 296: 1617–1618.
- Costas, J., 2002. Characterization of the intragenomic spread of the human endogenous retrovirus family HERV-W. *Mol. Biol. Evol.* 19: 526–533.
- Crick, F.H.C., 1958. On protein synthesis. *Sym. Soc. Exp. Biol.* 12: 138–183.
- Crick, F.H.C., 1970. Central Dogma of Molecular Biology. *Nature* 227: 561–563.
- Davis, S. & J.C. Watson, 1996. *In vitro* activation of the interferon-induced, double-stranded RNA-dependent protein kinase PKR by RNA from the 3' untranslated regions of human alpha-tropomyosin. *P. Natl. Acad. Sci. USA* 93: 508–513.
- Dawkins, R., 1976. *The Selfish Gene*. Oxford University Press, Oxford, UK.
- DeChiara, T.M. & J. Brosius, 1987. Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. *P. Natl. Acad. Sci. USA* 84: 2624–2628.
- Deininger, P.L., H. Tiedge, J. Kim & J. Brosius, 1996. Evolution, expression, and possible function of a master gene for amplification of an interspersed repeated DNA family in rodents. *Prog. Nucleic Acid Res.* (edited by W.E. Cohn & K. Moldave) 52: 67–88.
- Deloukas, P., L.H. Matthews, J. Ashurst, J. Burton, J.G. Gilbert et al., 2001. The DNA sequence and comparative analysis of human chromosome 20. *Nature* 414: 865–871.
- Devos, K.M., J.K.M. Brown & J.L. Bennetzen, 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* 12: 1075–1079.
- Doolittle, W.F., 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284: 601–603.
- Eddy, S.R., 2001. Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* 2: 919–929.
- Eddy, S.R., 2002. Computational genomics of noncoding RNA genes. *Cell* 109: 137–140.
- Eichler, E.E., 2001a. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet.* 17: 661–669.
- Eichler, E.E., 2001b. Segmental duplications: what's missing, misassigned, and misassembled – and should we care? *Genome Res.* 11: 653–656.
- Erdmann, V.A., M.Z. Barciszewska, M. Szymanski, A. Hochberg, N. de Groot et al., 2001. The non-coding RNAs as riboregulators. *Nucleic Acids Res.* 29: 189–193.
- Fagerheim, T., P. Raeymaekers, F.E. Tonnessen, M. Pedersen, L. Tranebjaerg et al., 1999. A new gene (DYX3) for dyslexia is located on chromosome 2. *J. Med. Genet.* 36: 664–669.
- Ferrigno, O., T. Virolle, Z. Djabari, J.P. Ortonne, R.J. White et al., 2001. Transposable B2 SINE elements can provide mobile RNA polymerase II promoters. *Nat. Genet.* 28: 77–81.
- Filipowicz, W., 2000. Imprinted expression of small nucleolar RNAs in brain: time for RNomics. *P. Natl. Acad. Sci. USA* 97: 14035–14037.
- Flavell, A.J., 1995. Retroelements, reverse transcriptase and evolution. *Comp. Biochem. Phys. B* 110: 3–15.
- Gogolevskaya, I.K. & D.A. Kramerov, 2002. Evolutionary history of 4.5S RNA and indication that it is functional. *J. Mol. Evol.* 54: 354–364.
- Goodier, J.L., E.M. Ostertag & H.H. Kazazian, Jr., 2000. Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.* 9: 653–657.
- Gould, S.J., 2002. *The Structure of Evolutionary Theory*. Belknap, Harvard University Press, Cambridge, MA.
- Gould, S.J., & E.S. Vrba, 1982. Exaptation – a missing term in the science of form. *Paleobiology* 8: 4–15.
- Herbert, A. & A. Rich, 1999. RNA processing and the evolution of eukaryotes. *Nat. Genet.* 21: 265–269.
- Horiike, S., K. Mitsuya, M. Meguro, N. Kotobuki, A. Kashiwagi et al., 2000. Targeted disruption of the human LIT1 locus defines a putative imprinting control element playing an essential role in Beckwith-Wiedemann syndrome. *Hum. Mol. Genet.* 9: 2075–2083.
- Hurst, L.D. & G.T. McVean, 1998. Do we understand the evolution of genomic imprinting? *Curr. Opin. Genet. Dev.* 8: 701–708.
- Hurst, L.D. & N.G. Smith, 1999. Molecular evolutionary evidence that H19 mRNA is functional. *Trends Genet.* 15: 134–135.
- Hurst, L.D., A. Atlan & B.O. Bengtsson, 1996. Genetic conflicts. *Q. Rev. Biol.* 71: 317–364.
- Hüttenhofer, A. & J. Brosius, 2002. Experimental RNomics, pp. 217–240. In *Functional Genomics*, edited by M. Galperin & E.V. Koonin. Horizon Scientific Press, New York.
- Hüttenhofer, A., M. Kiefmann, S. Meier-Ewert, J. O'Brien, H. Lehrach et al., 2001. RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.* 20: 2943–2953.
- Inouye, S. & M. Inouye, 1995. Structure, function, and evolution of bacterial reverse transcriptase. *Virus Genes* 11: 81–94.
- Iwasa, Y., 1998. The conflict theory of genomic imprinting: how much can be explained? *Curr. Top. Dev. Biol.* 40: 255–293.
- Jamain, S., M. Giron dot, P. Leroy, M. Clergue, H. Quach et al., 2001. Transduction of the human gene FAM8A1 by endogenous retrovirus during primate evolution. *Genomics* 78: 38–45.
- Jeffares, D.C., A.M. Poole & D. Penny, 1998. Relics from the RNA world. *J. Mol. Evol.* 46: 18–36.
- Johnson, M.E., L. Viggiano, J.A. Bailey, M. Abdul-Rauf, G. Goodwin et al., 2001. Positive selection of a gene family during the emergence of humans and African apes. *Nature* 413: 514–519.
- Jordan, I.K., L.V. Matyunina & J.F. McDonald, 1999. Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. *P. Natl. Acad. Sci. USA* 96: 12621–12625.
- Jurka, J., 1998. Repeats in genomic DNA: mining and meaning. *Curr. Opin. Struc. Biol.* 8: 333–337.
- Kim, J., J.A. Martignetti, M.R. Shen, J. Brosius & P. Deininger, 1994. Rodent BC1 RNA gene as a master gene for ID element amplification. *P. Natl. Acad. Sci. USA* 91: 3607–3611.
- Kobayashi, S., S. Goto & K. Anzai, 1991. Brain-specific small RNA transcript of the identifier sequences is present

- as a 10S ribonucleoprotein particle. *J. Biol. Chem.* 266: 4726–4730.
- Koller, M. & E.E. Strehler, 1988. Characterization of an intronless human calmodulin-like pseudogene. *FEBS Lett.* 239: 121–128.
- Kordis, D. & F. Gubensek, 1998. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. *P. Natl. Acad. Sci. USA* 95: 10704–10709.
- Kremerskothen, J., M. Nettermann, A. op de Bekke, M. Bachmann & J. Brosius, 1998a. Identification of human autoantigen La/SS-B as BC1/BC200 RNA-binding protein. *DNA Cell Biol.* 17: 751–759.
- Kremerskothen, J., D. Zopf, P. Walter, J.G. Cheng, M. Nettermann et al., 1998b. Heterodimer SRP9/14 is an integral part of the neural BC200 RNP in primate brain. *Neurosci Lett.* 245: 123–126.
- Kurdyukov, S.G., Y.B. Lebedev, Artamonova, II, T.N. Gorodentseva, A.V. Batrak et al., 2001. Full-sized HERV-K (HML-2) human endogenous retroviral LTR sequences on human chromosome 21: map locations and evolutionary history. *Gene* 273: 51–61.
- Kurychev, V.Y., B.V. Skryabin, J. Kremerskothen, J. Jurka & J. Brosius, 2001. Birth of a gene: locus of neuronal BC200 sn-mRNA in three prosimians and human BC200 pseudogenes as archives of change in the Anthroidea lineage. *J. Mol. Biol.* 309: 1049–1066.
- Lagos-Quintana, M., R. Rauhut, W. Lendeckel & T. Tuschl, 2001. Identification of novel genes coding for small expressed RNAs. *Science* 294: 853–858.
- Lagos-Quintana, M., R. Rauhut, A. Yalcin, J. Meyer, W. Lendeckel et al., 2002. Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* 12: 735–739.
- Lai, E.C., 2002. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* 30: 363–364.
- Lampson, B.C., S. Inouye & M. Inouye, 1991. msDNA of bacteria. *Prog. Nucleic Acid Re.* 40: 1–24.
- Lau, N.C., L.P. Lim, E.G. Weinstein & D.P. Bartel, 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294: 858–862.
- Lee, R.C. & V. Ambros, 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 294: 862–864.
- Lin, Y., J. Brosius & H. Tiedge, 2001. Neuronal BC1 RNA: co-expression with growth-associated protein-43 messenger RNA. *Neuroscience* 103: 465–479.
- Lipshitz, H.D., D.A. Peattie & D.S. Hogness, 1987. Novel transcripts from the ultrabithorax domain of the bithorax complex. *Genes Dev.* 1: 307–322.
- Liu, A.Y., B.S. Torchia, B.R. Migeon & R.F. Siliciano, 1997. The human NTT gene: identification of a novel 17-kb noncoding nuclear RNA expressed in activated CD4+ T cells. *Genomics* 39: 171–184.
- Lottin, S., A.S. Vercoutter-Edouart, E. Adriaenssens, X. Czeszak, J. Lemoine et al., 2002. Thioredoxin post-transcriptional regulation by H19 provides a new function to mRNA-like non-coding RNA. *Oncogene* 21: 1625–1631.
- Lyle, R., D. Watanabe, D. te Vrugte, W. Lerchner, O.W. Smrzka et al., 2000. The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. *Nat. Genet.* 25: 19–21.
- Lyon, M.F., 1998. X-chromosome inactivation: a repeat hypothesis. *Cytogenet. Cell Genet.* 80: 133–137.
- Lyon, M.F., 2000. LINE-1 elements and X chromosome inactivation: a function for 'junk' DNA? *P. Natl. Acad. Sci. USA* 97: 6248–6249.
- Makalowski, W., 2000. Genomic scrap yard: how genomes utilize all that junk. *Gene* 259: 61–67.
- Malik, K., A. Salpekar, A. Hancock, K. Moorwood, S. Jackson et al., 2000. Identification of differential methylation of the WT1 antisense regulatory region and relaxation of imprinting in Wilms' tumor. *Cancer Res.* 60: 2356–2360.
- Martignetti, J.A. & J. Brosius, 1993a. BC200 RNA: a neural RNA polymerase III product encoded by a monomeric Alu element. *P. Natl. Acad. Sci. USA* 90: 11563–11567.
- Martignetti, J.A. & J. Brosius, 1993b. Neural BC1 RNA as an evolutionary marker: guinea pig remains a rodent. *P. Natl. Acad. Sci. USA* 90: 9698–9702.
- Martignetti, J.A. & J. Brosius, 1995. BC1 RNA: transcriptional analysis of a neural cell-specific RNA polymerase III transcript. *Mol. Cell. Biol.* 15: 1642–1650.
- Medstrand, P., J.R. Landry & D.L. Mager, 2001. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. *J. Biol. Chem.* 276: 1896–1903.
- Meller, V.H. & B.P. Rattner, 2002. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *EMBO J.* 21: 1084–1091.
- Meller, V.H., K.H. Wu, G. Roman, M.I. Kuroda & R.L. Davis, 1997. roX1 RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell* 88: 445–457.
- Michod, R.E., 1996. Cooperation and conflict in the evolution of individuality. II. Conflict mediation. *P. Roy. Soc. Lond. B Bio.* 263: 813–822.
- Michod, R.E. & D. Roze, 2001. Cooperation and conflict in the evolution of multicellularity. *Heredity* 86: 1–7.
- Millar, J.K., J.C. Wilson-Annan, S. Anderson, S. Christie, M.S. Taylor et al., 2000. Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum. Mol. Genet.* 9: 1415–1423.
- Mitsuya, K., M. Meguro, M.P. Lee, M. Katoh, T.C. Schulz et al., 1999. LIT1, an imprinted antisense RNA in the human KvLQT1 locus identified by screening for differentially expressed transcripts using monochromosomal hybrids. *Hum. Mol. Genet.* 8: 1209–1217.
- Moore, H., K. Dvorakova, N. Jenkins & W. Breed, 2002. Exceptional sperm cooperation in the wood mouse. *Nature* 418: 174–177.
- Moran, J.V., R.J. DeBerardinis & H.H. Kazazian, Jr., 1999. Exon shuffling by L1 retrotransposition. *Science* 283: 1530–1534.
- Morrish, T.A., N. Gilbert, J.S. Myers, B.J. Vincent, T.D. Stamato et al., 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* 31: 159–165.
- Mouches, C., N. Bensaadi & J.C. Salvado, 1992. Characterization of a LINE retroposon dispersed in the genome of three non-sibling *Aedes* mosquito species. *Gene* 120: 183–190.
- Muddashetty, R.S., T. Khanam, A. Kondrashov, M. Bundman, A. Iacoangeli et al., 2002. Poly(A) binding protein is associated with neuronal BC1 and BC200 ribonucleoprotein particles. *J. Mol. Biol.* 321: 433–445.
- Muller, H.J., A.A. Prokofyeva-Belgovskaya & K.V. Kossikov, 1936. Unequal crossing-over in the Bar mutant as a result of duplication of a minute chromosome section. *C. R. (Doklady) Acad. Sci. URSS* 1: 87–88.
- Mural, R.J., M.D. Adams, E.W. Myers, H.O. Smith, G.L. Miklos et al., 2002. A comparison of whole-genome shotgun-derived

- mouse chromosome 16 and the human genome. *Science* 296: 1661–1671.
- Muslimov, I.A., G. Banker, J. Brosius & H. Tiedge, 1998. Activity-dependent regulation of dendritic BC1 RNA in hippocampal neurons in culture. *J. Cell. Biol.* 141: 1601–1611.
- Muslimov, I.A., Y. Lin, M. Heller, J. Brosius, Z. Zakeri et al., 2002. A small RNA in testis and brain: implications for male germ cell development. *J. Cell. Sci.* 115: 1243–1250.
- Nekrutenko, A. & W.H. Li, 2001. Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* 17: 619–621.
- Nemes, J.P., K.A. Benzow, M.L. Moseley, L.P. Ranun & M.D. Koob, 2000. The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein (KLHL1). *Hum. Mol. Genet.* 9: 1543–1551.
- Ninomiya, S., M. Isomura, K. Narahara, Y. Seino & Y. Nakamura, 1996. Isolation of a testis-specific cDNA on chromosome 17q from a region adjacent to the breakpoint of t(12;17) observed in a patient with acampomelic campomelic dysplasia and sex reversal. *Hum. Mol. Genet.* 5: 69–72.
- Noyce, L. & A.A. Piper, 1994. Isolation of a potentially functional HPRT processed pseudogene from the hill kangaroo *Macropus robustus*. *Gene* 150: 361–365.
- Noyce, L., J. Conaty & A.A. Piper, 1997. Identification of a novel tissue-specific processed HPRT gene and comparison with X-linked gene transcription in the Australian marsupial *Macropus robustus*. *Gene* 186: 87–95.
- Ohno, S., 1970. *Evolution by Gene Duplication*. Springer, New York.
- Okada, N., 1991. SINEs. *Curr. Opin. Genet. Dev.* 1: 498–504.
- Okada, N., M. Hamada, I. Ogiwara, K. Ohshima, 1997. SINEs and LINEs share common 3' sequences: a review. *Gene* 205: 229–243.
- Orgel, L.E. & F.H.C. Crick, 1980. Selfish DNA: the ultimate parasite. *Nature* 284: 604–607.
- Ostertag, E.M. & H.H. Kazazian, Jr., 2001. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* 35: 501–538.
- Paces, J., A. Pavlicek & V. Paces, 2002. HERVd: database of human endogenous retroviruses. *Nucleic Acids Res.* 30: 205–206.
- Parker, G.A. & L. Partridge, 1998. Sexual conflict and speciation. *Philos. T. Roy. Soc. B* 353: 261–274.
- Partridge, L. & L.D. Hurst, 1998. Sex and conflict. *Science* 281: 2003–2008.
- Pasquinelli, A.E., 2002. MicroRNAs: deviants no longer. *Trends Genet.* 18: 171–173.
- Patthy, L., 1991 Exons – original building blocks of proteins? *Bioessays* 13: 187–192.
- Pelczar, P. & W. Filipowicz, 1998. The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential and is a member of the 5'-terminal oligopyrimidine gene family. *Mol. Cell. Biol.* 18: 4509–4518.
- Petrov, D.A., 2001. Evolution of genome size: new approaches to an old problem. *Trends Genet.* 17: 23–28.
- Petrov, D.A., T.A. Sangster, J.S. Johnston, D.L. Hartl & K.L. Shaw, 2000. Evidence for DNA loss as a determinant of genome size. *Science* 287: 1060–1062.
- Poole, A.M., D.C. Jeffares & D. Penny, 1998. The path from the RNA world. *J. Mol. Evol.* 46: 1–17.
- Quentin, Y., 1994. Emergence of master sequences in families of retroposons derived from 7sl RNA. *Genetica* 93: 203–215.
- Rastinejad, F. & H.M. Blau, 1993. Genetic complementation reveals a novel regulatory role for 3' untranslated regions in growth and differentiation. *Cell* 72: 903–917.
- Rastinejad, F., M.J. Conboy, T.A. Rando & H.M. Blau, 1993. Tumor suppression by RNA from the 3' untranslated region of alpha-tropomyosin. *Cell* 75: 1107–1117.
- Renaud, F. & T. de Meeus, 1991. A simple model of host–parasite evolutionary relationships. Parasitism: compromise or conflict? *J. Theor. Biol.* 152: 319–327.
- Rozhdestvensky, T.S., A.M. Kopylov, J. Brosius & A. Hüttenhofer, 2001. Neuronal BC1 RNA structure: evolutionary conversion of a tRNA-Ala domain into an extended stem-loop structure. *RNA* 7: 722–730.
- Runte, M., A. Huttenhofer, S. Gross, M. Kiefmann, B. Horsthemke et al., 2001. The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Hum. Mol. Genet.* 10: 2687–2700.
- Samonte, R.V. & E.E. Eichler, 2002. Segmental duplications and the evolution of the primate genome. *Nat. Rev. Genet.* 3: 65–72.
- Sawata, M., D. Yoshino, H. Takeuchi, A. Kamikouchi, K. Ohashi et al., 2002. Identification and punctuate nuclear localization of a novel noncoding RNA, Ks-1, from the honeybee brain. *RNA* 8: 772–785.
- Schön, U., W. Seifarth, C. Baust, C. Hohenadl, V. Erfle et al., 2001. Cell type-specific expression and promoter activity of human endogenous retroviral long terminal repeats. *Virology* 279: 280–291.
- Shippen-Lentz, D. & E.H. Blackburn, 1990. Functional evidence for an RNA template in telomerase. *Science* 247: 546–552.
- Skryabin, B.V., J. Kremerskothen, D. Vassilacopoulou, T.R. Disotell, V.V. Kapitonov et al., 1998. The BC200 RNA gene and its neural expression are conserved in Anthropoidea (Primates). *J. Mol. Evol.* 47: 677–685.
- Skryabin, B.V., V. Sukonina, U. Jordan, N. Sachser & L. Lewejohann, 2003. Role of a small non-messenger RNA in behavior: targeted deletion of neuronal BC1 RNA gene in mice. (submitted).
- Seutels, F., R. Zwart & D.P. Barlow, 2002. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 415: 810–813.
- Smit, A.F. & A.D. Riggs, 1995. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res.* 23: 98–102.
- Sober, E. & D.S. Wilson, 1998. *Unto Others. The Evolution and Psychology of Unselfish Behavior*. Harvard University Press, Cambridge.
- Sorek, R., G. Ast & D. Graur, 2002. Alu-containing exons are alternatively spliced. *Genome Res.* 12: 1060–1067.
- Storz, U., 2002. Counting all genes: how many other RNAs exist and what do they do? *Science* 296: 1260–1263.
- Sturtevant, A.H., 1925. The effects of unequal crossing over at the *Bar* locus in *Drosophila*. *Genetics* 10: 117–147.
- Szathmáry, E. & J.M. Smith, 1995. The major evolutionary transitions. *Nature* 374: 227–232.
- Takeda, K., H. Ichijo, M. Fujii, Y. Mochida, M. Saitoh et al., 1998. Identification of a novel bone morphogenetic protein-responsive gene that may function as a noncoding RNA. *J. Biol. Chem.* 273: 17079–17085.
- Taylor, B.A., A. Navin, B.V. Skryabin & J. Brosius, 1997. Localization of the mouse gene (Bc1) encoding neural BC1 RNA near the fibroblast growth factor 3 locus (Fgf3) on distal chromosome 7. *Genomics* 44: 153–154.
- Temin, H.M., 1989. Reverse transcriptases. Retrons in bacteria. *Nature* 339: 254–255.
- Thieffry, D. & S. Sarkar, 1998. Forty years under the central dogma. *Trends Biochem. Sci.* 23: 312–316.

- Tiedge, H., W. Chen & J. Brosius, 1993. Primary structure, neural-specific expression, and dendritic location of human BC200 RNA. *J. Neurosci.* 13: 2382–2390.
- Tiedge, H., U.C. Dräger & J. Brosius, 1992. Murine BC1 RNA in dendritic fields of the retinal inner plexiform layer. *Neurosci. Lett.* 141: 136–138.
- Tiedge, H., R.T. Freneau, P.H. Weinstock, O. Arancio & J. Brosius, 1991. Dendritic location of neural BC1 RNA. *P. Natl. Acad. Sci. USA* 88: 2093–2097.
- Tycowski, K.T. & J.A. Steitz, 2001. Non-coding snoRNA host genes in *Drosophila*: expression strategies for modification guide snoRNAs. *Eur. J. Cell. Biol.* 80: 119–125.
- Tycowski, K.T., M.D. Shu & J.A. Steitz, 1996. A mammalian gene with introns instead of exons generating stable RNA products. *Nature* 379: 464–466.
- Ullu, E. & C. Tschudi, 1984. Alu sequences are processed 7SL RNA genes. *Nature* 312: 171–172.
- Velleca, M.A., M.C. Wallace & J.P. Merlie, 1994. A novel synapse-associated noncoding RNA. *Mol. Cell. Biol.* 14: 7095–7104.
- Vinogradova, T., L. Leppik, E. Kalinina, P. Zhulidov, K.H. Grzeschik et al., 2002. Selective differential display of RNAs containing interspersed repeats: analysis of changes in the transcription of HERV-K LTRs in germ cell tumors. *Mol. Genet. Genomics* 266: 796–805.
- Volff, J.N., C. Korting & M. Schartl, 2000. Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes. *Mol. Biol. Evol.* 17: 1673–1684.
- Wang, W., F.G. Brunet, E. Nevo & M. Long, 2002. Origin of sphinx, a young chimeric RNA gene in *Drosophila melanogaster*. *P. Natl. Acad. Sci. USA* 99: 4448–4453.
- Ware, T.L., H. Wang & E.H. Blackburn, 2000. Three telomerases with completely non-telomeric template replacements are catalytically active. *EMBO J.* 19: 3119–3131.
- Watanabe, Y. & M. Yamamoto, 1994. *S. pombe* mei2+ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. *Cell* 78: 487–498.
- Weiner, A.M., 2002. SINES and LINES: the art of biting the hand that feeds you. *Curr. Opin. Cell. Biol.* 14: 343–350.
- Weiner, A.M., P.L. Deininger & A. Efstratiadis, 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* 55: 631–661.
- West, N.C., A.M. Roy-Engel, H. Imataka, N. Sonenberg & P.L. Deininger, 2002. Shared protein components of SINE RNPs. *J. Mol. Biol.* 321: 423–432.
- Williams, G.C., 1997. *The Pony Fish's Glow. And Other Clues to Plan and Purpose in Nature.* Basic Books, New York, N.Y.
- Woese, C.R., 1967. *The Genetic Code: The Molecular Basis for Genetic Expression.* Harper and Row, New York.
- Woese, C.R., 2002. On the evolution of cells. *P. Natl. Acad. Sci. USA* 99: 8742–8747.
- Yamamoto, M., 1996. Regulation of meiosis in fission yeast. *Cell Struct. Funct.* 21: 431–436.
- Yamashita, A., Y. Watanabe, N. Nukina & M. Yamamoto, 1998. RNA-assisted nuclear transport of the meiotic regulator Mei2p in fission yeast. *Cell* 95: 115–123.
- Yaswen, P., A. Smoll, J. Hosoda, G. Parry & M.R. Stampfer, 1992. Protein product of a human intronless calmodulin-like gene shows tissue-specific expression and reduced abundance in transformed cells. *Cell Growth Differ.* 3: 335–345.
- Zupunski, V., F. Gubensek & D. Kordis, 2001. Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. *Mol. Biol. Evol.* 18: 1849–1863.