

# Retroposed SNOfall—A mammalian-wide comparison of platypus snoRNAs

Jürgen Schmitz,<sup>1,4</sup> Anja Zemann,<sup>1</sup> Gennady Churakov,<sup>1</sup> Heiner Kuhl,<sup>2</sup> Frank Grützner,<sup>3</sup> Richard Reinhardt,<sup>2</sup> and Jürgen Brosius<sup>1</sup>

<sup>1</sup>Institute of Experimental Pathology (ZMBE), University of Münster, Münster 48149, Germany; <sup>2</sup>MPI Molecular Genetics, Berlin-Dahlem 14195, Germany; <sup>3</sup>School of Molecular & Biomedical Science, The University of Adelaide, Adelaide 5005, Australia

Diversification of mammalian species began more than 160 million years ago when the egg-laying monotremes diverged from live bearing mammals. The duck-billed platypus (*Ornithorhynchus anatinus*) and echidnas are the only potential contemporary witnesses of this period and, thereby, provide a unique insight into mammalian genome evolution. It has become clear that small RNAs are major regulatory agents in eukaryotic cells, and the significant role of non-protein-coding (npc) RNAs in transcription, processing, and translation is now well accepted. Here we show that the platypus genome contains more than 200 small nucleolar (sno) RNAs among hundreds of other diverse npcRNAs. Their comparison among key mammalian groups and other vertebrates enabled us to reconstruct a complete temporal pathway of acquisition and loss of these snoRNAs. In platypus we found *cis*- and *trans*-duplication distribution patterns for snoRNAs, which have not been described in any other vertebrates but are known to occur in nematodes. An exciting novelty in platypus is a snoRNA-derived retroposon (termed snoRTE) that facilitates a very effective dispersal of an H/ACA snoRNA via RTE-mediated retroposition. From more than 40,000 detected full-length and truncated genomic copies of this snoRTE, at least 21 are processed into mature snoRNAs. High-copy retroposition via multiple host gene-promoted transcription units is a novel pathway for combining housekeeping function and SINE-like dispersal and reveals a new dimension in the evolution of novel snoRNA function.

[Supplemental material is available online at [www.genome.org](http://www.genome.org). The sequence data from this study have been submitted to NCBI GenBank under accession nos. EU093990–EU094198.]

Small nucleolar RNAs (snoRNAs) are essential for RNA maturation, guiding 2'-O-ribose methylation (C/D-box snoRNAs) and pseudouridylation (H/ACA-box snoRNAs) of chiefly ribosomal RNAs, the central components of the protein synthesis machinery. Such an important housekeeping function implies high genomic conservation. However, in nematodes, we recently found an extraordinary degree of plasticity in the generation of snoRNA paralogs as well as in the selection of their modified targets (Zemann et al. 2006). In a process termed *cis*-duplication, we found that some snoRNAs duplicated from one intronic location to a neighboring intron of the same gene. Others distributed via *trans*-duplication to distant locations in other genes or chromosomes. One mechanism of *trans*-duplication is retrotransposition (Vitali et al. 2003). In mammals, nonautonomous retroposons such as short interspersed elements (SINEs) are predominantly dispersed by an autonomous long interspersed element 1 (LINE1; L1)-mediated mechanism. The autonomous, non-LTR (long terminal repeat), L1 retroposons provide the machinery needed for the reverse transcription and integration of the associated nonautonomous retroposons. In mammals the activity of L1 elements is restricted to therians (placentals and marsupials). To date, no L1 activity has been demonstrated for monotremes (Kordis et al. 2006); thus, L1-mediated co-retroposition of RNAs is also expected to be restricted to therian mammals. In monotremes, a different retropositional landscape, mediated by the L2

LINE machinery, dominates. Representative mobile elements retroposed in this way include the L2-mobilized mammalian-wide interspersed elements (MIRs) and the monotreme-specific (Mon1) SINEs. Another class of retroposon-like non-LTR transposable elements (RTEs) facilitating retroposition, originally discovered in *Caenorhabditis elegans* (Youngman et al. 1996), is occasionally represented in mammals (Malik and Eickbush 1998), comprising, for example, ~2.3% of the *Monodelphis domestica* (a small South American opossum) genome (Gentles et al. 2007). snoRNA distribution associated with nonautonomous retroposition usually occurs only in few copies (Weber 2006; Luo and Li 2007). However, this is remarkably different in platypus, where we have identified about 40,000 genomic copies of a platypus-specific snoRNA retroposon we have called a snoRTE, reflecting its hybrid nature consisting of a snoRNA and an RTE-like sequence. Here we provide evidence for a platypus-specific processing pathway that differentially leads to H/ACA snoRNAs or full-length snoRTE retroposons.

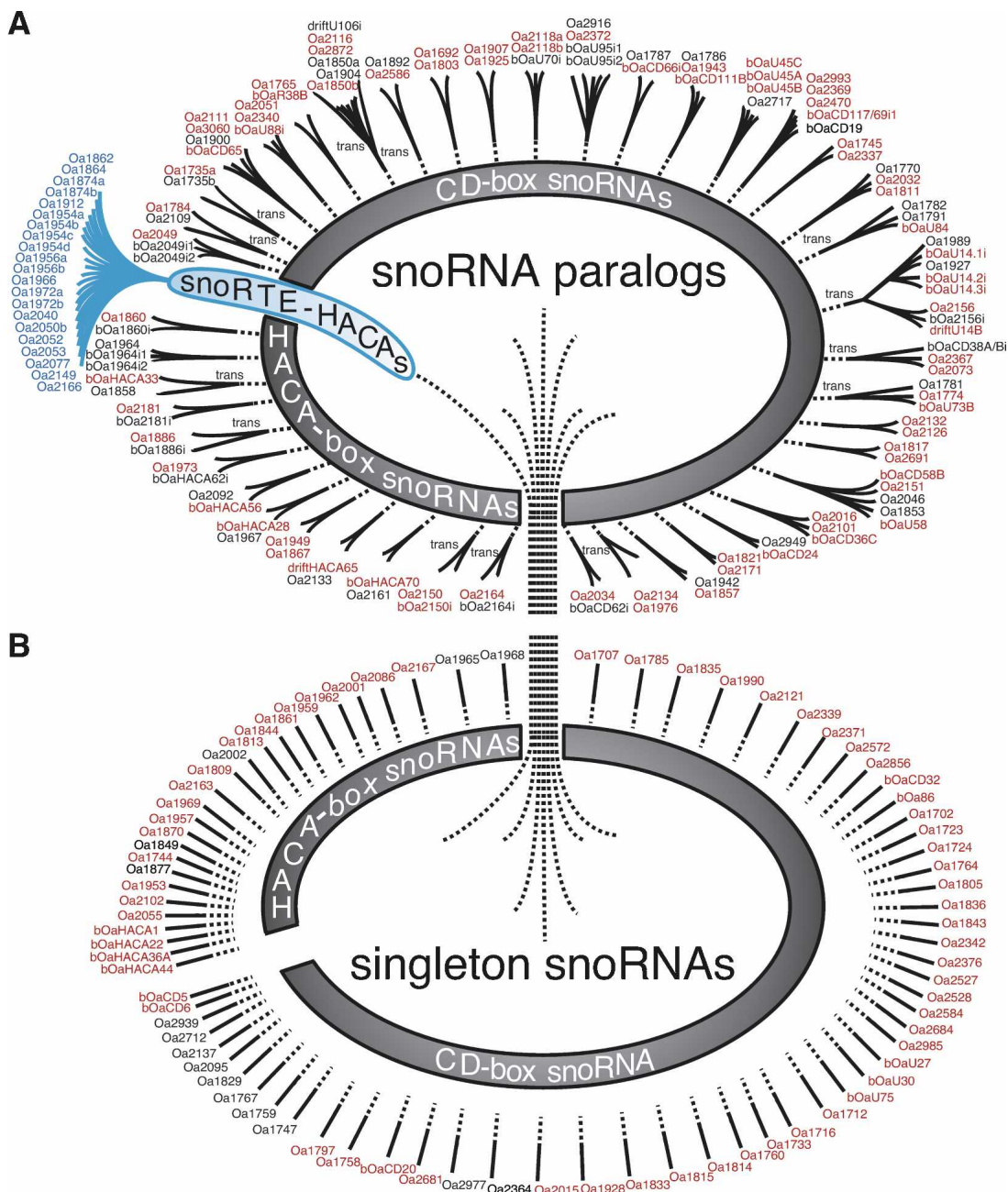
## Results and Discussion

In a platypus brain cDNA library generated from small non-protein-coding RNAs (npcRNAs), we identified 166 individual snoRNAs among more than 20,000 cloned sequences. Furthermore, BLAST searches of platypus genomic sequences revealed an additional 51 such snoRNAs (for a complete description, see Fig. 1; Supplemental Fig. S1; Supplemental material). Seventy-three of the snoRNAs are H/ACA snoRNAs, and 144 are C/D snoRNAs. A comparative analysis of these snoRNAs with the genomic information from other key vertebrate species revealed that 49 are

#### <sup>4</sup>Corresponding author.

E-mail [jueschm@uni-muenster.de](mailto:jueschm@uni-muenster.de); fax 49-251-8352134.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.7177908>.

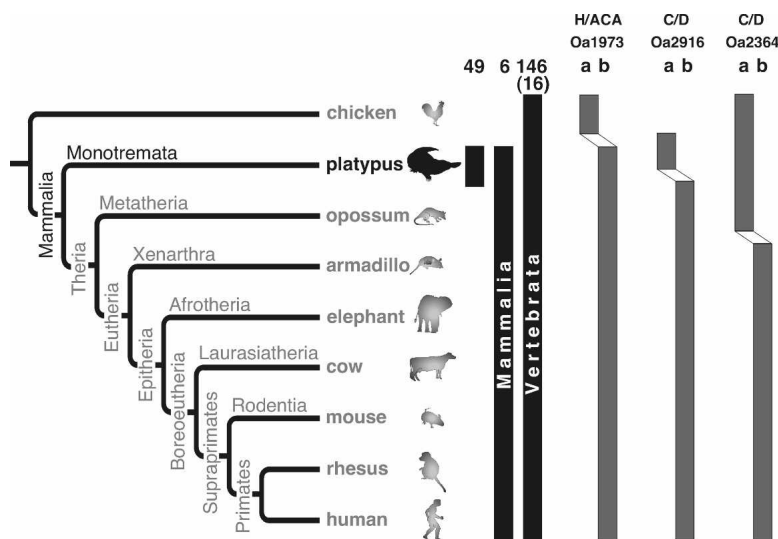


**Figure 1.** snoRNA representation in the genome of platypus. (A) snoRNAs with corresponding paralogs distributed by *cis*- or *trans*-duplication (trans). (Red lettering) snoRNAs conserved in human. snoRNAs are grouped as CD-box snoRNAs, H/ACA-box snoRNAs, and snORTE-H/ACAs, the last group was generated by retroposition. (B) Nonduplicated snoRNAs. (Oa) Experimentally identified snoRNAs; (bOa) snoRNAs identified computationally. Diverged snoRNAs (drift) show no canonical structural requirements necessary for function.

platypus/monotreme-specific (Supplemental Material 2). Of the remaining 168, we detected 146 orthologs in non-mammalian vertebrates, and six were present in all mammals but were clearly absent in birds and other non-mammalian species (Fig. 2). The remaining 16 snoRNAs may belong to either of the last two groups but could not be clearly assigned because of a lack of corresponding sequence information in non-mammalian species. The high degree of conservation among vertebrate snoRNAs is indicative of the purifying selection on their function in modifying housekeeping RNAs. At the same time, snoRNAs also show

a high degree of plasticity among their paralogs (Fig. 2; Supplemental material). This dualism of conservation levels is similar to that of certain vertebrate populations of miRNAs (Tanzer and Stadler 2006).

Forty-five of the snoRNAs had undergone one or more duplications in platypus, yielding a total of 138 paralogs (the number of paralogs refers to the total number of related forms with a common origin), 91 of which were C/D-box and 47 were H/ACA-box snoRNAs (Fig. 1). The majority of C/D and H/ACA snoRNAs in our study were generated via *cis*-duplication (119 paralogs).



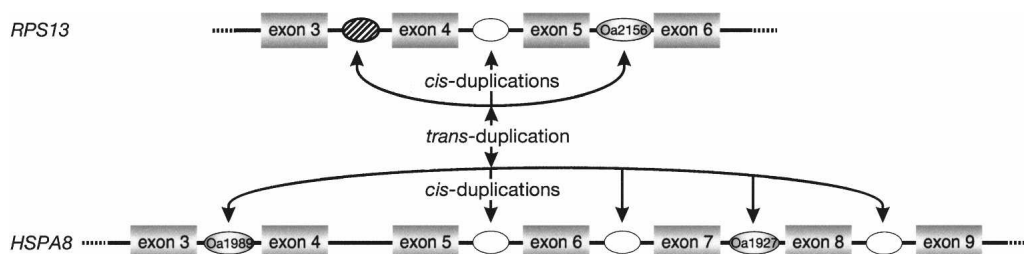
**Figure 2.** Conservation and plasticity of platypus snoRNAs. Phylogenetic tree of mammals modified after Kriegs et al. (2006). (Black vertical bars) The conservation (platypus-specific, mammalian-specific, and presence in at least mammals plus birds) of the platypus snoRNAs. The numbers above the black bars represent the number of experimentally identified snoRNA orthologs that show clear presence or absence in the corresponding groups. The numbers in parentheses denote mammalian conserved loci for which no non-mammalian sequence information was accessible. (Gray bars) Examples of shifts from one potentially active paralog (a) to another potentially active paralog (b) at different genomic locations (the details are presented as Supplemental material). The snoRNA designations in the cDNA library are Oa1973, Oa2916, and Oa2364.

This is somewhat different in nematodes—the only other species for which such analysis is available—in which only H/ACA snoRNAs were dispersed in this way (Zemann et al. 2006). If the immediate neighboring introns were already occupied by other snoRNAs, the next free intron was often targeted. In nine of 138 *cis*-duplications, we observed the birth of novel platypus-specific expressed C/D snoRNA paralogs into neighboring introns, evidenced by the fact that the orthologous loci in other mammals clearly lacked the snoRNA paralogs. Among the cDNA sequences, we found only a few cases of paralogs generated via *trans*-duplication (nine C/D and four H/ACA snoRNAs). A combination of *cis*- and *trans*-duplication was observed for the eight C/D snoRNA paralogs hosted by the ribosomal protein S13 (*RPS13*) and the heat-shock protein 8 (*HSPA8*) genes in platypus (Fig. 3). In this case, three paralogs (including one highly diverged form) reside in neighboring introns in the *RPS13* gene, and five closely

related paralogs are located in adjacent introns of the *HSPA8* gene. Based on sequence similarities, we propose that a single *trans*-duplication event initiated further *cis*-dispersal within one of the two genes. In nematodes, we found only single duplication events (Zemann et al. 2006), while in one particular H/ACA platypus snoRNA, at least 20 duplications occurred leading to 21 paralogs in our cDNA library (Supplemental material). Compared to known duplication patterns, this extraordinarily high number is rather unusual. Alignment of these sequences to platypus genomic data revealed a highly conserved 229-nt 3'-flank (Supplemental material). Parts of the conserved flank exhibit ~70% sequence similarity to the 3'-region of the RTE-mobilized marsupial MAR1b-SINES and ~90% sequence similarity to BovB\_Platt, a recently characterized LINE-related RTE retroposon (Arian Smit, unpubl.; Fig. 4). RTE retroposons described for ruminants (Okada and Hamada 1997), afrotherians (Gogolevsky et al. 2007), and marsupials (Gentles et al. 2007) show a mosaic distribution in

mammals and are proposed to be dispersed by horizontal transfer (Zupunski et al. 2001; Piskurek and Okada 2007). A genomic BLAST of the 372-nt consensus sequence in platypus, including the snoRNA plus the conserved flanking region, revealed 14,844 full-length (~27% with perfect direct repeats  $\geq 10$  nt) and 27,332 5'- and/or 3'-truncated copies. One thousand nine hundred and five of the full-length forms and 346 of the truncated forms (but including the entire snoRNA) were located in currently annotated introns (out of a total of about 142,000 introns), with a significant preference for the antisense orientation relative to the transcribed direction of the host gene (1286 of 2251;  $\chi^2$ -test,  $P < 0.0001$ ).

The high copy number, flanking direct repeats, and the high similarity and temporal overlap in activity to BovB\_Platt RTE-elements (Supplemental Fig. S2) are indications of a novel retroposon class (termed snoRTE) associated with RTE-autonomous



**Figure 3.** *Cis*- and *trans*-duplication of C/D snoRNA paralogs in platypus. *Cis*-duplications of three paralogs of a C/D snoRNA were detected in the gene for ribosomal protein S13 (*RPS13*). One of these was found in the cDNA library (snoRNA Oa2156); the second was detected by BLAST in the neighboring intron and is probably functional (oval between exons 4 and 5; intact box motifs, conserved between platypus and humans), and the third, also detected by BLAST, is diverged and most likely nonfunctional (based on structural requirements) in platypus but functional in human, mouse, and cow (oval between exons 3 and 4). *Trans*-duplication of one of these paralogs occurred to the 70-kDa heat-shock protein 8 gene (*HSPA8*) or vice versa. Within the *HSPA8* gene, we also found two paralogs in our cDNA library that evolved via *cis*-duplications (snoRNAs Oa1989 and Oa1927). Three additional, potentially functional paralogs were found by BLAST search and are conserved between platypus and humans. (Filled ovals) snoRNAs in our platypus cDNA library; (open ovals) snoRNAs located by BLAST search with the library sequences; (hatched oval) a nonfunctional snoRNA in platypus.



eral, the birth of new retropositionally active master genes is quite rare. New master genes require that the internal RNA polymerase III promoter elements in the corresponding SINE be fortuitously juxtaposed at the right distance to sequences at the locus of integration that act as external RNA polymerase III promoter elements (Ludwig et al. 2005; Khanam et al. 2007). Hence, most SINE copies are transcriptionally silent. The existence of a snoRTE that can hitch a ride on virtually any intron-containing RNA polymerase II transcript might pose a great challenge to the organism, as, unchecked, and owing to the extremely high number of potential active snoRTE “mastergenes” compared with the temporary activity of one or a few mastergenes (e.g., for SINE retroposition), it might develop into a dangerous runaway situation of retroposon amplification. It was shown in *Drosophila* that an increased genomic copy number of transposable elements had deleterious effects on organismal fitness (Pasyukova et al. 2004). In the case of the significant dominance of antisense-over sense-oriented intronic snoRTEs, which are cotranscribed but not capable of being retroposed, the platypus genome might already have developed countermeasures, whose mechanisms will be interesting to reveal.

## Methods

Brain tissue was obtained from an adult male duck-billed platypus (*Ornithorhynchus anatinus*) collected at the upper Barnard river, New South Wales (Animal ethics permit no. R.CG.07.03 to F.G.). Total RNA was prepared using the TRIzol method as previously described (Zemann et al. 2006) and fractionated on denaturing PAGE gels. The 10–60-nt (small) and 60–500-nt (large) RNA fractions were excised, passively eluted, ethanol-precipitated, and then C-tailed (DeChiara and Brosius 1987). Full-length cDNAs were generated using the adapter ligation protocol (Chen et al. 2007) with a SalI restriction site and T4 RNA ligase. First-strand cDNA was synthesized with a ThermoScript cDNA synthesis kit (Invitrogen) using a NotI primer. The RNA–cDNA hybrid strand was amplified using a SalI primer. After SalI and NotI double digestion, the cDNAs were cloned into the pSPORT1 vector (Invitrogen) and transformed. cDNAs were sequenced using M13 standard primers (detailed methodologies are included in the Supplemental Methods).

A total of 11,521 and 10,369 cDNA clones were sequenced from the small and large RNA fractions, respectively. Vectors were trimmed and assembled with a modified version of the DNASTAR Lasergene 7.1 package. For mammalian comparative analyses, we used NCBI-BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi>), the Ensembl Platypus-BLAST (<http://www.ensembl.org/Multi/blastview>), and the UCSC BLAT (<http://genome.ucsc.edu/cgi-bin/hgBlat>); for RNA structures, the RNAfold program, mfold (<http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi>); for scanning additional snoRNAs, the Ribosomal Protein Genes Database (<http://ribosome.med.miyazaki-u.ac.jp/>); and for analysis of transposed elements, RepeatMasker (<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>).

The 21 paralogs of the snoRTE were BLASTed against platypus genomic sequences and the 3'-conserved RTE-like parts extended. Full-length snoRTEs were aligned and preliminary consensus snoRTE sequences derived and used to create a RepeatMasker library. The *Ornithorhynchus anatinus*-5.0.1 chromosomes were then locally screened with this library. A novel C-script was used to extract about 40,000 sequences that were aligned by MAFFT (Katoh et al. 2002) and used to define snoRTE

consensus sequences. Intronic snoRTEs were identified from [http://www.ensembl.org/Ornithorhynchus\\_anatinus/index.html](http://www.ensembl.org/Ornithorhynchus_anatinus/index.html). All experimentally obtained snoRNA sequences were submitted to the NCBI database.

## Acknowledgments

We thank Russell Jones (Newcastle University) for help with the tissue collection. Facilities were provided by Macquarie Generation and Glenrock Station, New South Wales. Approval to collect animals was granted by the New South Wales National Parks and Wildlife Services, New South Wales Fisheries and the Animal Experimentation and Ethics Committee, Australian National University. We thank Jean Marie Smith for optimizing the DNASTAR package Version 7.1 to perform high-throughput data analysis of our cDNA library sequences, Marsha Bundman for editorial assistance, and Janina Thiel and Sven Klages for excellent technical assistance. This work was supported by the Nationales Genomforschungsnetz (NGFN, 0313358A) to J.B. and J.S., the Deutsche Forschungsgemeinschaft (DFG, SCHM1469) to J.S. and J.B., the Max-Planck Society to R.R., and the Australian Research Council to F.G.

## References

- Cavaillé, J. and Bachellerie, J.-P. 1996. Processing of fibrillar-associated snoRNAs from pre-mRNA introns: An exonucleolytic process exclusively directed by the common stem-box terminal structure. *Biochimie* **78**: 443–456.
- Chen, X.S., Rozhdestvensky, T.S., Collins, L.J., Schmitz, J., and Penny, D. 2007. Combined experimental and computational approach to identify non-protein-coding RNAs in the deep-branching eukaryote *Giardia intestinalis*. *Nucleic Acids Res.* **35**: 4619–4628.
- DeChiara, T.M. and Brosius, J. 1987. Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. *Proc. Natl. Acad. Sci.* **84**: 2624–2628.
- Gentles, A.J., Wakefield, M.J., Kohany, O., Gu, W., Batzer, M.A., Pollock, D.D., and Jurka, J. 2007. Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res.* **17**: 992–1004.
- Gogolevsky, K.P., Vassetzky, N.S., and Kramerov, D.A. 2007. Bov-B-mobilized SINEs in vertebrate genomes. *Gene* **407**: 75–85.
- Katoh, K., Misawa, K., Kuma, K.-I., and Miyata, T. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**: 3059–3066.
- Khanam, T., Rozhdestvensky, T.S., Bundman, M., Galiveti, C.R., Handel, S., Sukonina, V., Jordan, U., Brosius, J., and Skryabin, B.V. 2007. Two primate-specific small non-protein-coding RNAs in transgenic mice: Neuronal expression, subcellular localization and binding partners. *Nucleic Acids Res.* **35**: 529–539.
- Kordis, D., Lovsin, N., and Gubensek, F. 2006. Phylogenomic analysis of the L1 retroposons in Deuterostomia. *Syst. Biol.* **55**: 886–901.
- Kriegs, J.O., Churakov, G., Kiefmann, M., Jordan, U., Brosius, J., and Schmitz, J. 2006. Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol.* **4**: e91. doi: 10.1371/journal.pbio.0040091.
- Krull, M., Petrusma, M., Makalowski, W., Brosius, J., and Schmitz, J. 2007. Functional persistence of exonized mammalian-wide interspersed repeat elements (MIRs). *Genome Res.* **17**: 1139–1145.
- Long, M., Betran, E., Thornton, K., and Wang, W. 2003. The origin of new genes: Glimpses from the young and old. *Nat. Rev. Genet.* **4**: 865–875.
- Ludwig, A., Rozhdestvensky, T.S., Kuryshev, V.Y., Schmitz, J., and Brosius, J. 2005. An unusual primate locus that attracted two independent Alu insertions and facilitates their transcription. *J. Mol. Biol.* **350**: 200–214.
- Luo, Y. and Li, S. 2007. Genome-wide analyses of retrogenes derived from the human box H/ACA snoRNAs. *Nucleic Acids Res.* **35**: 559–571.
- Malik, H.S. and Eickbush, T.H. 1998. The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. *Mol. Biol. Evol.* **15**: 1123–1134.
- Okada, N. and Hamada, M. 1997. The 3' ends of tRNA-derived SINEs originated from the 3' ends of LINES: A new example from the

- bovine genome. *J. Mol. Evol.* **44**: S052–S056.
- Pasyukova, E.G., Nuzhdin, S.V., Morozova, T.V., and Mackay, T.F. 2004. Accumulation of transposable elements in the genome of *Drosophila melanogaster* is associated with a decrease in fitness. *J. Hered.* **95**: 284–290.
- Piskurek, O. and Okada, N. 2007. Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. *Proc. Natl. Acad. Sci.* **104**: 12046–12051.
- Schmitz, J., Churakov, G., Zischler, H., and Brosius, J. 2004. A novel class of mammalian-specific tailless retroseudogenes. *Genome Res.* **14**: 1911–1915.
- Tanzer, A. and Stadler, P.F. 2006. Evolution of microRNAs. *Methods Mol. Biol.* **342**: 335–350.
- Vitali, P., Royo, H., Seitz, H., Bachelier, J.P., Hüttenhofer, A., and Cavaille, J. 2003. Identification of 13 novel human modification guide RNAs. *Nucleic Acids Res.* **31**: 6543–6551.
- Weber, M.J. 2006. Mammalian small nucleolar RNAs are mobile genetic elements. *PLoS Genet.* **3**: e36. doi: 10.1371/journal.pgen.0020205.
- Xing, Y. and Lee, C. 2006. Alternative splicing and RNA selection pressure evolutionary consequences for eukaryotic genomes. *Nat. Rev. Genet.* **7**: 499–509.
- Youngman, S., van Luenen, H.G.A.M., and Plasterk, R.H.A. 1996. Rte-1, a retrotransposon-like element in *Caenorhabditis elegans*. *FEBS Lett.* **380**: 1–7.
- Zemann, A., op de Bekke, A., Kiefmann, M., Brosius, J., and Schmitz, J. 2006. Evolution of small nucleolar RNAs in nematodes. *Nucleic Acids Res.* **34**: 2676–2685.
- Zupunski, V., Gubensek, F., and Kordis, D. 2001. Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. *Mol. Biol. Evol.* **18**: 1849–1863.

Received September 21, 2007; accepted in revised form December 12, 2007.