Retroposed SNOfall—A mammalian-wide comparison of platypus snoRNAs

Jürgen Schmitz,1,4 Anja Zemann,1 Gennady Churakov,1 Heiner Kuhl,2 Frank Grützner,3 Richard Reinhardt,2 and Jürgen Brosius1

1Institute of Experimental Pathology (ZMBE), University of Münster, Münster 48149, Germany; 2MPI Molecular Genetics, Berlin-Dahlem 14195, Germany; 3School 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. 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 (Vitale 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 Eckbush 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 nonprotein-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
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).
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_Plat, a recently characterized LINE-related RTE retroposon (Arian Smit, unpubl.; Fig. 4). RETE 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 ≥ 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; χ²-test, P < 0.0001).

The high copy number, flanking direct repeats, and the high similarity and temporal overlap in activity to BovB_Plat RTE-elements (Supplemental Fig. S2) are indications of a novel retroposon class (termed snoRTE) associated with RTE-autonomous...
retroposition. snoRTEs are platypus/monotreme-specific and could not be detected in other vertebrates (Supplemental Material 2). A typical vertebrate snoRNA cotranscription embedded in an RNA polymerase II-transcribed host gene RNA with subsequent trimming of the flanking regions (Cavaillé and Bachellerie 1996) and an associated SINE-like retroposition via an RTE-mechanism represents a novel dual “symbiotic” evolutionary pathway of snoRNA and retroposition propagation manifested as a snoRTE. Splicing and differential trimming might lead to either snoRNAs or snoRTEs (Fig. 4). RNAs of both forms were identified by Northern blot analysis (Supplemental Fig. S3). We propose that in the absence of active snoRTE retroposition, the classical snoRNA trimming process (Cavaillé and Bachellerie 1996) removes the 3′-RTE part. The remaining processed sequence constitutes the mature H/ACA-snoRNA. Alternatively, in the presence of an active RTE retroposition machinery, the RTE part of the snoRTE is associated with proteins forming a ribonuclease complex. This association might protect the 3′-snoRNA part against further trimming. Consequently, the partially processed RNA constitutes the retropositionally active snoRTE. These elements are a source for new snoRNA isoforms and potentially new functions apart from originally modifying a specific 28S rRNA.

Figure 4. Structure of the novel class of platypus snoRTEs. snoRTEs are composed of a 5′-H/ACA-snoRNA with the H-box consensus sequence ANANNA and ACA-box sequence (structured part). The sequences that potentially interact with ribosomal or spliceosomal RNA by base complementarities guiding pseudouridylation are indicated (N-ψ). The (gray bar) 3′ part is similar to RTE-elements. Coordinates are given for the BovB_Plat RTE LINE corresponding 145-nt 5′ and 67-nt 3′ sequence regions. [(GACAA)₆] Terminal simple repeats; (A₄) a spacer sequence between the H/ACA snoRNA and BovB_Plat region. The flanks of the entire snoRTE exhibit (DR) direct repeats. The snoRTE is distributed via retroposition (retro-duplication), is dependent on the host gene RNA polymerase II promoter, and is probably subsequently trimmed in the presence of RTE proteins to a retropositionally active RNA, absence to an H/ACA-box snoRNA.

A BLAST search for this new snoRTE in all major vertebrate clades revealed the snoRNA's original intronic locus in the RPL32 gene, which is conserved from human to lizard. At that locus in platypus, however, a diverged, nonfunctional copy (lacking the ACA-box) was identified. In contrast, for most of all 21 retroposed paralogs identified in the cDNA library, the predicted secondary structures characteristic for snoRNAs are highly conserved, further indicating their possible functionality.

Conservation and diversity are two major opposing forces shaping the evolution of genomes. As an example, the interplay of these forces is clearly observed in the presence of both constitutive and alternative splicing (Xing and Lee 2006; Krull et al. 2007) or of both genes and pseudogenes (Long et al. 2003). The duplication processes of highly conserved npcRNAs, like snoRNAs, are important mechanisms for maintaining advantageous functions, and at the same time for evolving potentially novel ones. This fundamental mechanism of duplication accompanied by the functional change of one of the paralogs has been described in relatively simple multicellular nematodes (Zemann et al. 2006), and the analysis of platypus snoRNAs clearly demonstrates that a similar diversification occurs in vertebrates. It is interesting that, although jumping snoRNAs represent an unusual pathway to generate genomic and functional diversity, the high degree of amplification inherent in the “strategy” followed by sense-oriented intronic insertion appears to be limited by unknown regulatory mechanisms. This is clearly noticeable in the restricted amount of such snoRTEs compared to the number of antisense-oriented, retropositionally inactive elements. In gen-
eral, the birth of new retroactively actively 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 intrinsic 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 J.S. and J.B., 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. and Sven Klages for excellent technical assistance. This work was supported by the Nationales Genomforschungszentrum (NGFN, 0315585A) to J.B. and the Max-Planck Society to R.R., and the Australian Research Council to F.G.

**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 Genomforschungszentrum (NGFN, 0315585A) to J.B. and 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**


Okada, N. and Hamada, M. 1997. The 5’ ends of tRNA-derived SINEs originated from the 3’ ends of LINES. A new example from the...

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