Noncoding RNAs are essential to many biological processes. They serve as structural and catalytic components of cellular machines, such as the ribosome and spliceosome, and exert exquisite gene regulation through processes such as RNA silencing. Recent work has revealed that the transcribed portions of higher eukaryotic genomes are larger and more complex than previously thought, with the majority of transcripts being noncoding (Kapranov et al. 2007; Wilhelm et al. 2008). In one view, the number of such transcripts scales with evolutionary complexity as the need for sophisticated control circuitry expands (Mattick and Makunin 2005). Undoubtedly, large numbers of noncoding RNAs remain to be discovered; of those that are known, many are understudied and functions have been ascribed to only a minority (Mattick and Makunin 2005; Prasanth and Spector 2007; Wilusz et al. 2009). Here, we describe the properties and ancestry of the snaRs, a rapidly evolved family of small noncoding RNAs, and discuss their possible functions. We begin with the historical context in which the RNAs were discovered and then trace their ancestry to the Alu SINE (short interspersed nucleotide element) family, via two hitherto unreported sets of short genetic elements termed ASR (Alu/snaR-related) and CAS (Catarhine ancestor of snaR). This derivation entails a series of internal deletions followed by expansions. The evolution of these genes coincides with major primate speciation events: ASR elements are found in all monkeys and apes, whereas CAS elements are limited to Old World monkeys and apes. In contrast to ASR and CAS elements, which are retrotransposons, human snaR genes are predominantly located in three clusters on chromosome 19 and have been duplicated as part of a larger genetic element. Insertion of the element containing snaR-G into a gene encoding a chorionic gonadotropin β subunit generated new hormone genes in African Great Apes.

NF90: A MULTIFUNCTIONAL PROTEIN

Human adenoviruses transcribe two short (~160 nucleotides) noncoding RNAs, VA RNA, and VA RNA$, that accumulate to high cytoplasmic concentration in the late phase of viral infection (Mathews and Shenk 1991). Although VA RNA, is well characterized as a kinase inhibitor, counteracting the interferon-induced shut off of host cell protein synthesis, the function of VA RNA$ remains obscure. To approach the role of VA RNA$, our laboratory set out to characterize its cellular binding partners. Two RNA-binding proteins were identified: RNA helicase A and nuclear factor 90 (NF90), together with nuclear factor 45 (NF45) (Liao et al. 1998).

NF90 is the founding member of a family of RNA-binding proteins transcribed from the vertebrate-specific interleukin enhancer binding factor 3 (ILF3) gene (Duchange et al. 2000). Alternative splicing of ILF3 transcripts gives rise to two major protein isoforms: NF90 and NF110 (Fig. 1). These proteins carry two canonical double-stranded RNA-binding motifs (dsRBMs), a single-stranded nucleic-acid-binding RGG motif, and a DZF (dsRBMs and zinc-finger-associated) motif that is important in fly and mouse development (Meagher et al. 1999). NF90 and NF110 have distinct carboxyl termini that, in the case of NF110, are extended and contain a GQSY RNA-binding domain (Fig. 1). Both NF90 and NF110 have homology with the NF45 protein (Fig. 1), with which they exist in mutually stabilizing 1:1 heterodimeric complexes (Guan et al. 2008). As their names suggest, both NF90 and NF110 are present in the nucleus, but they carry a nuclear export signal (NES) as well as a nuclear localization signal (NLS), and a significant proportion (~20%) of NF90 is present in the cytoplasm (Parrott et al. 2005).

NF90 and NF110 bind numerous viral and cellular RNA and protein partners, and they have been associated with several biological processes. First recognized by binding to a promoter element (Corthesy and Kao 1994) and to synthetic duplex RNA (Bass et al. 1994), they have been assigned a number of functions, many of which await full investigation. Perhaps the best-characterized role is in the rapid induction of the interleukin-2 (IL-2) gene upon T-cell

snaR Genes: Recent Descendants of Alu Involved in the Evolution of Chorionic Gonadotropins

A.M. PARROTT AND M.B. MATHEWS
Department of Biochemistry and Molecular Biology, New Jersey Medical School, UMDNJ, Newark, New Jersey 07101-1709
Correspondence: mathews@umdnj.edu

We identified a novel family of human noncoding RNAs by in vivo cross-linking to the nuclear factor 90 (NF90) protein. These small NF90-associated RNAs (snaRs) are transcribed by RNA polymerase III and display restricted tissue distribution, with high expression in testis and discrete areas of the brain. The most abundant human transcript, snaR-A, interacts with the cell’s transcription and translation systems. snaR genes have evolved in African Great Apes (human, chimpanzee, and gorilla) and some are unique to humans. We traced their ancestry to the Alu SINE (short interspersed nucleotide element) family, via two hitherto unreported sets of short genetic elements termed ASR (Alu/snaR-related) and CAS (Catarhine ancestor of snaR). This derivation entails a series of internal deletions followed by expansions. The evolution of these genes coincides with major primate speciation events: ASR elements are found in all monkeys and apes, whereas CAS elements are limited to Old World monkeys and apes. In contrast to ASR and CAS elements, which are retrotransposons, human snaR genes are predominantly located in three clusters on chromosome 19 and have been duplicated as part of a larger genetic element. Insertion of the element containing snaR-G into a gene encoding a chorionic gonadotropin β subunit generated new hormone genes in African Great Apes.
NF90 facilitates IL-2 transcription (Shi et al. 2007), is specifically phosphorylated, and translocates to the cytoplasm where it stabilizes IL-2 mRNA (Shim et al. 2002; Pei et al. 2008). Increasingly, NF90 has been implicated in aspects of gene regulation, for example, through control of nuclear export (Pfeifer et al. 2008), mRNA stability (Pullmann et al. 2007), and microRNA (miRNA) processing (Sakamoto et al. 2009), making it a likely target for viral manipulation. Interestingly, NF90 binds VA RNAII considerably more efficiently than it binds VA RNAI, despite similarities between these two highly structured single-stranded viral RNAs (Liao et al. 1998). Its high affinity for NF90 suggests that VA RNAII may have evolved to displace cellular RNA from NF90, bringing some as yet undefined benefit to the virus.

**NF90 BINDS TO A NOVEL CELLULAR RNA FAMILY**

The dsRBMs of NF90 are largely occupied by cellular RNA throughout the cell cycle (Parrott et al. 2005). To verify known NF90-binding partners and to identify new ones, we isolated ribonucleoprotein (RNP) complexes after in vivo cross-linking. This technique stabilizes complexes in live cells and obviates partner reassortment that can occur during cell rupture and fractionation. As an additional advantage, it is not focused on detecting a particular RNA of interest. We used a human embryonic kidney 293 (HEK-293) stable cell line that expresses a tagged form of NF90 at moderate levels, comparable to endogenous NF90. RNP complexes containing NF90 were isolated by immunoprecipitation, and RNA ligands were extracted after reversal of the cross-links (Parrott et al. 2007). A heterogeneous mix of transcripts generated by RNA polymerase (pol) I, II, and III were found cross-linked to NF90, and all proved to be previously unidentified partners. The most intense band, visualized by autoradiography of 3′-end-labeled RNA and confirmed by oligonucleotide-directed RNase H digestion, was 5.8S rRNA. Other NF90-associated RNAs were cloned using 5′-rapid amplification of cDNA ends (5′-RACE) reverse transcriptase–polymerase chain reaction (RT-PCR) schemes (Parrott et al. 2007). The most abundant species identified by this means were members of a novel family of short RNAs that we termed snaR (small NF90 associated). Dimeric Alu transcripts, which are relatively rare in cells (see below), comprised the second most abundant group.

**THE snaR FAMILY OF NONCODING RNAs**

Two subsets of the snaR family of noncoding RNAs, snaR-A and snaR-B, were initially isolated with NF90. Search of the human genome identified genes encoding a further subset, snaR-C, and several outliers, mainly on chromosome 19 (Table 1 and Fig. 2) (Parrott and Mathews 2007). As shown experimentally for snaR-A (Parrott and Mathews 2007), snaR genes are predicted to be transcribed by pol III into highly structured RNAs of ~117 nucleotides that bind to the dsRBMs of NF90. The dominant snaR species, in terms of gene multiplicity (14 of 30 genes) as well as transcript abundance (~70,000 copies per HEK-293 cell), is snaR-A. Of the remaining genes, about half (seven genes) encode snaR-B and snaR-C, which are closely related to one another (Fig. 3A). snaR-A is found in a wide array of immortal cells, but it exhibits a sharply restricted distribution in human tissue, being expressed at a high level in testis (Parrott and Mathews 2007) and to a lesser extent in discrete parts of the brain (AM Parrott and MB Mathews, in prep.). Expression in testis and differential expression in the brain have also been observed for other snaRs. The snaR-A genes, together with those of snaR-B, -C and -D, are arrayed in two large inverted regions of tandem repeats on the q-arm of chromosome 19 (Fig. 2). The 2-Mb region separating the two clusters hosts more than 100 protein-coding genes including the luteinizing hormone (LH)/chorionic gonadotropin (CG) β subunit (LHβ/CGβ) gene cluster and a number of genes essential to sperm biology. In this region are two additional snaR genes, encoding snaR-G1 and snaR-G2. They are situated within the proximal promoters of the two most recent additions to the CGβ family, CGβ1 and CGβ2, respectively (Fig. 2), as discussed below.
STEPWISE EVOLUTION OF snaR COINCIDES WITH PRIMATE SPECIATION

Examination of available genome sequences revealed snaR genes in the chimpanzee (Table 1), but not in other mammals such as rhesus macaque and rodents. To trace the origin of this recently evolved gene family, we exploited a notable feature of the snaR genes, namely, that their flanking sequences are often more highly conserved than the genes themselves (Parrott and Mathews 2007). Analysis of great ape genomic DNA by PCR amplification using primers designed against these snaR flanking sequences, supplemented by database searches, revealed snaR genes in all of the African Great Apes (human, chimpanzee, bonobo, and gorilla) (Table 1) (AM Parrott et al., in prep.). In the Sumatran orangutan (an Asian Great Ape), however, the sequence spanned by the flanking homology exhibited only limited identity with snaR genes. Searches of the orangutan draft assembly failed to disclose authentic snaR sequences. We conclude that authentic snaR genes evolved in the African Great Apes.

The amplified orangutan sequence originates from chromosome 19 and is orthologous to a chromosome 19 region in macaque (Fig. 3B). In the macaque, the orthologous region is embedded in a triplet of adjacent 1.9-kb repeats that are demarcated by Alu sequences (Fig. 3B). This triplet repeat pattern is conserved in humans but fragmented in the orangutan, possibly due to draft assembly errors. snaR-F is present in the first repeat in human; elements resembling truncated snaR genes are present at the same position in both the first (Mm19) and the second (Mm19i) repeat in the macaque and in the first partial repeat in the orangutan (Pa19). Further searches disclosed that this class of truncated elements is restricted to Old World monkeys and apes (Catarrhines), and they were therefore termed CAS (Catarrhine ancestor of snaR) elements.

**Table 1. Subsets of African Great Ape snaR Genes**

<table>
<thead>
<tr>
<th>snaR</th>
<th>NCBI gene ID</th>
<th>Chromosome</th>
<th>Transcribed</th>
<th>Human</th>
<th>Pan</th>
<th>Gorilla</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1-14</td>
<td>100126798-99, 100169951-59, 100170216, 100191063</td>
<td>19q13 (53.102-53.140, 55.287-55.324)</td>
<td>+</td>
<td>14</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>B1-2</td>
<td>100170217, 100170224</td>
<td>19q13 (55.328-55.334)</td>
<td>+</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1-5</td>
<td>100170218-19, 100170223, 100170225-26</td>
<td>19q13 (53.107-53.152)</td>
<td>+</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>100170227</td>
<td>19q13 (55.335)</td>
<td>n.d.</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>100170220</td>
<td>19q13 (55.326)</td>
<td>n.d.</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>100126781</td>
<td>19q13 (55.800)</td>
<td>n.d.</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>100126780</td>
<td>19q13 (54.232)</td>
<td>–</td>
<td>1</td>
<td>2, +</td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>100170228</td>
<td>19q13 (54.227)</td>
<td>+</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H (2)</td>
<td>100170221</td>
<td>2p12 (78.036)</td>
<td>n.d.</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>I (3)</td>
<td>100170222</td>
<td>3q28 (192.078)</td>
<td>n.d.</td>
<td>1</td>
<td>7, +</td>
<td></td>
</tr>
<tr>
<td>J2</td>
<td>12p12 (21.158)</td>
<td>n.d.</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>21q21 (25.032)</td>
<td>n.d.</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* snaR nomenclature as defined by the HUGO Gene Nomenclature Committee. Former name is in parentheses.

*b* Human chromosome location (in Mb).

*c* snaR expression as confirmed by northern blot and/or RT-PCR of human tissue. n.d. indicates not determined.

*d* Number of gene copies in human and chimpanzee genomes.

*e* + denotes snaR subset presence in bonobo and gorilla as confirmed by genomic PCR.

*f* snaR-D and -E are pseudogenes of snaR-A and -B, respectively.

**Figure 2. snaR gene clusters.** Shown is the region of human chromosome 19q13.32-33 containing snaR genes. The locations and direction of transcription of snaR-A genes (black) and snaR-B, -C, -D, -E, -F, -G1, and -G2 genes (gray) are denoted by arrowheads. CGβ1 and CGβ2 exons are shown as open boxes and their direction of transcription by open arrows. Distances are in megabases (M).
The synteny evident in Figure 3B suggested that snaR genes evolved from CAS elements. Alignment of all snaR genes with CAS elements shows how this could have taken place at the molecular level (Fig. 4A). Compared to CAS elements, snaR genes have two internal 8-nucleotide expansions between pol III A- and B-box transcription elements. The first expansion (ε1), but not the second (ε2), is preserved in snaR-12, a putative "fossil" species that hints at the sequential nature of this evolutionary transition. A phylogram generated from this alignment emphasizes the distinction between the snaR family and CAS element sequences (Fig. 3A).

Searches for the triplet syntenic repeat in New World monkeys (Platyrrhines) disclosed a slightly longer element occupying the position equivalent to that of CAS in Catarrhine species. This longer element reflects the order of this evolutionary (Fig. 4A). Nucleotide substitutions also separate snaR from its ancestors. Most of the substitutions (~11 of 14) between CAS and snaR sequences are present in snaR-21, another probable "fossil" species that lacks both internal expansions ε1 and ε2, indicating that the majority of the substitutions preceded the internal expansions (Fig. 4A).

**Alu SINE: Presumptive Ancestor of snaR**

Dimeric Alu transcripts are ~300 nucleotides long and typically have a poly(A) tail and an internal AU-rich region separating two bulged hairpin domains, known as the scAlu (left) and Alu-RA (right) monomers (Fig. 5A). The left monomer region carries the pol III basal transcription promoter and enhancer. Alu’s dimeric form arose in primates through the fusion of a FLAM-C and a free right Alu monomer (FRAM) (Quentin 1992). Both monomers are independently descended from the highly conserved 7SL RNA that forms the scaffold of the signal recognition particle (SRP), which has a key role in the synthesis of secreted proteins (Kriegs et al. 2007). Alu genes underwent a series of three distinct expansion events through the self-replication mechanism of retrotransposition, with the most extensive uptake by the Alu-S subfamily ~35–40 million years ago (mya) (Bailey et al. 2003). Alu sequences now constitute more than 10% of the human genome, but most copies are transcriptionally silent because of mutations during retrotransposition and through neutral drift (Lander et al. 2001; Bennett et al. 2008). Consequently, pol-III-transcribed Alu RNAs are relatively rare, although they are rapidly induced up to 20-fold by cellular insults such as adenovirus infection and heat shock (Liu et al. 1995; Chang et al. 1996).
Alu retrotransposition is dependent on the reverse transcriptase and integrase/endonuclease genes encoded by long interspersed nucleotide elements (LINEs). As such, Alu has been regarded as a parasite’s parasite and a burden on the host genome. However, it is now realized that Alu has had a major role in shaping primate genomes, and it is a source of novel noncoding RNA species including snaR.

Alu insertion into protein-coding genes and their promoters can lead to alternative splice site selection or transcriptional regulation (Jurka 2004; Hasler and Strub 2006a). Alu sequences also promote genomic remodeling through nonallelic homologous recombination, resulting in localized gene disruption, duplication, deletion, or larger segmental duplications (Bailey et al. 2003). Segmental duplications are now considered to be critical in the evolution of the primate lineage; as a source of novel genes, they can facilitate both phenotypic variation and susceptibility to disease (Bailey and Eichler 2006).

snaR GENES DISSEMINATED BY SEGMENTAL DUPLICATION

Like Alu, ASR and CAS elements are scattered throughout primate genomes and the majority are flanked by short direct repeats. These are hallmarks of retrotransposon activity (Grindley 1978; Maraia et al. 1992). Thus, it appears that ASR and CAS elements are novel SINEs and are (or were) genes. With a single exception in chimpanzee, all human and chimpanzee (Tribe Hominini) CAS elements occupy syntenic positions. We conclude that almost all CAS elements have become transcriptionally silent in Hominini since divergence from the common ancestor. Thus, only five discrete CAS clones have been reported in expressed sequence tag (EST) databases (BU567151, BG526971, DW458066, DW458333, and BU56539), and they all originate from the same human locus. In contrast, at least half of the CAS elements in macaque and orangutan occupy unique loci, suggesting that CAS was an active retrotrans-

**Figure 4.** snaR evolution. (A) Clustal-X alignment of all snaR genes and selected CAS elements with 3′-oligo(A/T) tracts omitted. Full-length snaRs are marked with arrows: (green arrows) snaRs predicted to be transcriptionally active, (yellow arrows) snaRs with mutations in their pol III B box (solid box) or a truncated 3′-oligo(T) tract. Expansions e1 and e2 and the putative pol III A box are indicated. Asterisks show conserved nucleotides. This alignment generated the phylogram of Fig. 3A. (B) Schematic of the major deletions (red) and expansions (green) relating FLAM-C, ASR, CAS, and snaR genes. These molecular events apparently occurred contemporaneously with major events in primate evolution (dashed gray lines), allowing their timing to be estimated. pol III A and B boxes are represented by gray and black boxes, respectively. Two unique overlapping expansions in the Sumatran orangutan Pa19 CAS element are shown by gray and orange hatching. snaR-12 appears to be an intermediate species between CAS and snaR.
poson until quite recently in Catarrhines. These observations reflect a general decline in retrotransposon activity along the Hominoid lineage (apes) (Liu et al. 2009).

Unlike their ancestors, most snaR genes seem to have disseminated through segmental duplication of an encompassing locus, possibly as a result of nonallelic homologous recombination of flanking Alu sequences. Exceptionally, snaR-D and -E appear to represent cases of snaR retrotransposition (see Table 1) (Parrott and Mathews 2007). Strikingly, snaR gene radiation has coincided with a genomewide burst of segmental duplication within the African Great Ape ancestor (Marques-Bonet et al. 2009).

snaR’s mode of dissemination, which can result in gene duplication as well as deletion, has seemingly resulted in dramatic copy-number variation and snaR subset diversity within the African Great Apes (Table 1). For example, gorillas and humans have snaR-A genes, whereas chimpanzees do not; a single snaR-I gene is present in humans, whereas the chimpanzee has seven copies, and perhaps most interestingly, snaR gene duplication has contributed to the expansion of the well-studied LHβ/CGβ gene cluster (see below).

PREDICTED snaR STRUCTURES

The radical sequence changes, deletions, and expansions that took place during snaR evolution from a left monomeric Alu element (Fig. 4B) undoubtedly impacted the secondary structures of the RNA species in profound ways. If the successive deletions leading to CAS are mapped onto the experimentally determined model of Alu secondary structure (Sinnett et al. 1991), they constitute

Figure 5. snaRs appear to have structural plasticity. (A) Positions of deletions δ1 and δ2 superimposed on the Alu secondary structure model. (B) Predicted secondary structure of CAS RNA. A consensus sequence, derived from a multiple alignment of 120 primate CAS elements (after omitting 26 elements with single insertions/deletions) using WebLogo (Crooks et al. 2004), was folded using MFOLD (Zuker 2003). (C) Alternative secondary structures predicted for 23 of 29 snaRs. Six snaRs cannot adopt these structures due to internal deletions. The conserved apical stem-loop is boxed. (D) Clustal-X alignment of transcriptionally active (as defined in Fig. 4A) snaR genes with 3′-oligo(A/T) tracts omitted. (Cons) Consensus sequence, (#Var N) number of variable nucleotides at each position, (Multi) multiplicity of variable nucleotides at each position. The most recent snaR expansion, ε2, is indicated. (Asterisks) Conserved nucleotides. (E) Northern blots of cytoplasmic RNA from 293 cells resolved in 6% and 10% acrylamide/7 M urea gels and probed for snaR-A. 3′-pCp–labeled RNA served as a marker (M). (Asterisk) Band of unknown origin. (A, Adapted, with permission, from Mariner et al. 2008 [© Elsevier].)
the sequential loss of a middle stem of the left monomer of \textit{Alu} (sc\textit{Alu}; Fig. 5A). The second deletion could be a consequence of the first deletion, perhaps resulting in stabilization of the RNA structure. However, considering that numerous examples of the intermediate deletion (ASR) exist, some structural reorganization is likely. The predicted CAS RNA secondary structure (Fig. 5B) is stable, with a $\Delta G$ of folding of $-31.8$ kcal/mol, and distinct from that of \textit{Alu}.

Four predicted secondary structures can describe the lowest energy structure of most \textit{snaR}s (Fig. 5C). All four structures share a common apical stem-loop. Primer-extension analysis of \textit{snaR}-A uncovered a “strong-stop” at the predicted start of its apical stem-loop, supporting the existence of this conserved motif (data not shown). The apical stem-loop should make an ideal NF90 dsRBM-binding site (Parrott and Mathews 2007), but further work is needed to establish whether this motif is the point of contact between the two molecules because dsRBMs bind to imperfect as well as canonical dsRNAs (Bevilacqua et al. 1998; Tian et al. 2004). Three of the predicted secondary structures (ii–iv, Fig. 5C) have similar minimal folding energies (within 6% of the optimal folding energy). We attempted to deduce the preferred secondary structure by folding a \textit{snaR} consensus sequence. This was derived from an alignment of 18 \textit{snaR} sequences with an intact B box and $3'$-oligo (T) tract (Fig. 5D), considering species that lack these transcriptional requirements and/or harbor internal deletions (Fig. 4A) to be potential pseudogenes. The most variable region of the alignment, which is encompassed by the most recent expansion ($\epsilon_2$), dictates the length of the conserved apical stem-loop (Fig. 5D). No satisfactory unifying secondary structure model could be derived for the consensus sequence, however, because all predictions placed variant nucleotides in stabilizing G:C base pairs. One interpretation of this finding is that the \textit{snaR} subsets adopt different secondary structures while retaining a common apical stem-loop.

**RAPIDLY EVOLVED \textit{snaR}s MAY BE PARALOGS**

That \textit{snaR} genes are generally more divergent in sequence than their flanking regions hints at adaptation for distinct functions or maybe refinement of one function. Human accelerated region 1 (HAR1), a small non-coding transcript that is considered to be one of the most rapidly evolved RNAs in humans, exhibits 18 base changes in 118 nucleotides (15.3% difference) between its human and chimpanzee forms; consequently, it adopts a cloverleaf structure in humans but a hairpin structure in chimpanzees (Beniaminov et al. 2008). The function of HAR1 is not known, although a role in human neurodevelopment has been posited (Pollard et al. 2006), and it is surmised that the cloverleaf structure of the human ortholog adapts it for this role (Beniaminov et al. 2008).

The most abundant human \textit{snaR}s form two distinct subsets, \textit{snaR}-A and \textit{snaR}-B/C, that have undergone rapid divergence from other \textit{snaR} genes and from each other (see Fig. 3A). \textit{snaR}-A is present in humans and gorillas, arguing for an origin in the common ancestor of African Great Apes followed by loss in \textit{Pan}, whereas \textit{snaR}-B and -C are unique to humans (Fig. 3A and Table 1) and therefore probably evolved rapidly after the Homo-\textit{Pan} species divergence. \textit{snaR}-A underwent copy-number expansion, leading to 14 highly homologous alleles in humans and an unknown number in gorilla. Intriguingly, \textit{snaR}-B and -C genes are interspersed with \textit{snaR}-A genes in the two \textit{snaR} clusters in a tandem repeat pattern, suggesting that \textit{snaR}-B/C evolved from redundant \textit{snaR}-A alleles (see Fig. 2).

Conservatively, there are 12 base changes in 103 nucleotides (excluding the oligo[A/T] tract: 11.6% difference) between \textit{snaR}-A and \textit{snaR}-C species (Fig. 5D), and these two subsets are predicted to fold into distinctly different structures (Parrott and Mathews 2007). \textit{snaR}-B and -C are predicted to form a unique and very stable ($-58.3$ kcal/mol) secondary structure (structure i), reminiscent of that of \textit{scAlu}, whereas \textit{snaR}-A adopts an equally robust linear structure (structure iv, Fig. 5C). This species exhibits mobility differences in urea gels depending on the concentration of acrylamide, consistent with the existence of a structured region that persists under denaturing conditions (Fig. 5E). \textit{snaR}-A is expressed in the pituitary gland, whereas \textit{snaR}-B/C is not (AM Parrott and MB Mathews, in prep.), so we speculate that nucleotide substitutions have allowed these two subsets to adopt different secondary structures for different functions—in other words, they are paralogs.

**\textit{snaR} BIOLOGY AND POSSIBLE FUNCTIONS**

Despite its abundance, \textit{snaR}-A has a short half-life (only $\sim 20$ min) in HeLa cells, implying a rapid rate of transcription (Parrott and Mathews 2007) and suggestive of a regulatory role. \textit{snaR}-A is distributed between the nucleus and cytoplasm of immortal cells and appears to undergo processing to a slightly shortened form in the cytoplasm (Fig. 6A). NF90 is the only protein yet identified to bind \textit{snaR}, but its role in \textit{snaR} biology is undefined. Knockdown of NF90 by RNA interference (RNAi) resulted in reduction of \textit{snaR}-A levels in both the nucleus and cytoplasm (Fig. 6B). Hence, NF90 protein is necessary for the synthesis or stability of \textit{snaR}-A within the cell but possibly not for its maturation or nucleocytoplasmic shuttling. Knockdown of NF45 by RNAi did not impact \textit{snaR}-A levels, however, suggesting that this NF90 function may not depend on formation of a heterodimer with NF45.

The presence of \textit{snaR} RNA in the nucleus and cytoplasm is consistent with functions in either or both cellular compartments. \textit{Alu} RNAs offer a precedent: They have a role in translation, as discussed in the next section, and in transcription. As elucidated by Goodrich and colleagues, dimeric \textit{Alu} can bind tightly to pol II and inhibits its function (Mariner et al. 2008). The left and right monomers bind independently to pol II, but only \textit{Alu}-RA (Fig. 5A) is able to inhibit polymerase function in vitro. When tested in the same way, \textit{snaR}-A was found to mimic its \textit{scAlu} ancestor; it bound tightly to pol II but did not inhibit its function (J Goodrich, pers. comm.). These findings are compatible with the evolutionary origin of \textit{snaR}. 
genes discussed above, but they shed little light on a putative nuclear function at present because the significance of scAlu binding to pol II is still obscure.

Sucrose gradient sedimentation analysis revealed that a substantial fraction of cytoplasmic snar-A is associated with ribosomes, both with monosomes and polysomes (Fig. 6C). This suggests that snar, like 7SL RNA and some of its relatives, has a role in translation. Alu RNA has retained sufficient similarity to its 7SL ancestor to form a complex with the SRP9/14 heterodimer (Chang et al. 1996; Hasler and Strub 2006b). The SRP9/14 heterodimer positions the Alu domain of 7SL in the elongation-factor-binding site of the large 60S ribosomal subunit, such that the SRP can arrest translation (Halic et al. 2004). Interestingly, the Alu-SRP9/14 RNP can also inhibit translation in vitro (Hasler and Strub 2006b). BC200, a 200-nucleotide-long noncoding RNA derived from the left monomer of Alu/FLAM, also associates with the SRP9/14 heterodimer (Martignetti and Broius 1993; Kremerskothen et al. 1998). BC200 is specific to Simiformes and is expressed in the cytoplasm of dendritic cells, especially at the synapse (Tiedge et al. 1993; Skryabin et al. 1998). Like its murine functional analog BC1 (which is unrelated to Alu), BC200 can form complexes with the eukaryotic initiation factor 4A (eIF4A) helicase and uncouples ATPase hydrolysis from the factor’s RNA unwinding ability, thereby repressing the translation of mRNAs dependent on eIF4A for unwinding of their structured 5’UTRs (untranslated regions) (Lin et al. 2008). Thus, two RNAs descended from 7SL retain the ability to bind protein components of the SRP, associate with ribosomes, and perform translational control functions.

It is not known whether SRP proteins can bind snar RNA, but NF90 binds directly to dimeric Alu and 5.8S rRNA (Parrott et al. 2007). The latter is a structural component of the 60S ribosomal subunit, present at its surface and critical to ribosome translocation (Abou Elela and Nazar 1997; Graifer et al. 2005). NF90 is also present in the ribosomal salt wash fraction (Langland et al. 1999), possibly as a result of its association with 5.8S rRNA, but is dissociated from ribosomes in a sucrose gradient (Fig. 6C). Hence, snar’s defined protein partner may transiently associate with ribosomes and perhaps transport snar to them. The restricted tissue distribution of snarS raises the possibility that they function as tissue-specific gene regulators acting at the level of translation.

**NEW CHORIONIC GONADOTROPIN GENES**

Chorionic gonadotropin (CG) belongs to the same glycoprotein hormone family as luteinizing hormone, follicle-stimulating hormone, and thyroid-stimulating hormone, each composed of a common α subunit and a distinct β subunit. The CGβ subunit genes are located in a cluster on chromosome 19q13.33 together with LHβ, from which they evolved (Fig. 7, bottom). CGβ subunit genes have undergone several duplications during primate speciation (Maston and Ruvolo 2002; Hallast et al. 2008). Humans possess six CG genes: hCGβ3, 5, 7, and 8, which share 96% identity with LHβ, and hCGβ1 and hCGβ2, which have a unique 5’ region (Hallast et al. 2007). It is likely that CGβ1 and CGβ2 evolved stepwise in the African Great Apes (Fig. 7). In the first step, the 5’ region of an ancestral CGβ gene was substituted by a segmental duplication encompassing a snar-G gene, giving rise to CGβ1 (AM Parrott et al., in prep.). The gene sequence that was replaced contained the proximal promoter, 5’UTR, and first exon of the ancestral CGβ gene (Bo and Boime 1992). In the second step, CGβ1 was duplicated together with flanking sequences that included CGβ5, to generate CGβ2 and CGβ3 (Hallast et al. 2008). These events introduced snar-G1 and snar-G2 into the proximal promoters of CGβ1 and CGβ2, respectively.

Both gorillas and humans have snar-G1 and snar-G2 genes, whereas the genus Pan has two copies of snar-G1 but no snar-G2 (Table 1). This observation is consistent with recent sequencing of Hominini genomes, which found chimpanzees to possess two copies of CGβ1 but to lack CGβ2 and CGβ3, indicating that Pan-specific duplication of CGβ1 occurred after the Homo-Pan divergence (Hallast et al. 2008). Considering that the gorilla possesses...
have detected the snaR-G1 including in its B box and a 3′ (see Figs. 3A and 4A). Correspondingly, we snaR-G1 that of hCG (see Jameson and Hollenberg 1993). In humans, hCG provides binding sites for transcription factors, whereas snaR-G2, with multiple substitutions in human snaR-G1 structure (Hallast et al. 2007). Conceivably, snaR-G2 orthology appears to have diverged more than that of snaR-G2, with multiple substitutions in human snaR-G1 including in its B box and a 3′ deletion in gorilla snaR-G1 (see Figs. 3A and 4A). Correspondingly, we have detected the snaR-G2, but not snaR-G1, transcript in human tissues (AM Parrott and MB Mathews, in prep.). Moreover, snaR-G1 is predicted to contain a number of proximal transcription-factor-binding sites that are not in snaR-G2 (Hallast et al. 2007). Conceivably, snaR-G1 provides binding sites for transcription factors, whereas snaR-G2 acts in cis through its continued transcription or in trans via its expressed transcript.

FUNCTION AND REGULATION OF hCG

hCG is essential for the successful onset and progression of pregnancy in primates. Its classical function is to maintain the corpus luteum by sustaining progesterone production (Jameson and Hollenberg 1993). In humans, hCG has additional roles in angiogenesis, blastocyst implantation, and nourishment of the growing fetus (Lei et al. 1992; Zygmunt et al. 2002; Cole et al. 2006; Handschuh et al. 2007), as well as development of tolerance in the maternal immune system (Wan et al. 2008; Schumacher et al. 2009). Whether CGB1 and CGB2 contribute to these roles is an open question, and even their coding capacity has been held suspect.

Substitution of the first CG exon by the snaR-containing element introduced two new potential start codons in CGB1/2, presumptively leading to alternate reading frame usage and the synthesis of novel proteins lacking recognizable functional motifs (Bo and Boime 1992; Hallast et al. 2007). However, an experimental analysis of start codon selection demonstrated that these genes indeed give rise to authentic hCG (AM Parrott et al., in prep.). On the other hand, they display a distinct pattern of tissue expression, probably by virtue of their altered proximal promoter. Unlike other hCG genes, hCGβ1/2 are expressed weakly in placenta but at relatively high levels in testis, to the extent that they account for approximately one-third of the total hCGB mRNA in this tissue (Rull et al. 2008). Substantial concentrations of hCGB peptide have been detected in human semen and in fetal testis, where it is thought to act as the primary stimulus of fetal Leydig cells, resulting in early testosterone secretion and masculine differentiation of the male fetus (Clements et al. 1976; Saito et al. 1988; Brotherton 1989). Thus, hCGβ1/2 may have evolved to generate testicular hCGB, functioning as a paracrine alternative to LHβ endocrine stimulation of Leydig cells.

CONCLUSION

Noncoding RNAs are the most common transcripts in higher eukaryotes and a likely source of genes that propel speciation events. We have traced the ancestry of the snaR family of small noncoding RNA to Alu elements, the most populous SINE family in primates. The snaR genes evolved recently in the line leading to the African Great Apes, via two retrotransposon intermediates (ASR and CAS). Members of the snaR family display rapid sequence divergence, suggesting their accelerated evolution in the
African Great Apes toward new functions. snaR-A stably associates with pol II and ribosomes, possibly linking this descendant of the Alu family to control of gene expression. snaR genes display species-specific amplification and loss by recombination. Such events led to the generation of new hormone genes in the African Great Apes, with expression by recombination. Such events led to the generation of new genes display species-specific amplification and loss.

ACKNOWLEDGMENTS

We thank our collaborators Dr. James Goodrich and Linda Drullinger for their work on pol II binding, and Dr. Tsafi Pe’ery for invaluable discussion. Funding of this work is provided by National Institutes of Health (grant RO1 A1034552).

REFERENCES


Cell Biol 12: 1500–1506.
Mattick JS, Makunin IV. 2005. Small regulatory RNAs in mam-
Sinnett D, Richer C, Deragon JM, Labuda D. 1991. Alu RNA sec-
snaR Genes: Recent Descendants of Alu Involved in the Evolution of Chorionic Gonadotropins

A.M. Parrott and M.B. Mathews

Cold Spring Harb Symp Quant Biol 2009 74: 363-373 originally published online December 22, 2009
Access the most recent version at doi:10.1101/sqb.2009.74.038

References
This article cites 68 articles, 24 of which can be accessed free at:
http://symposium.cshlp.org/content/74/363.full.html#ref-list-1

License

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

Copyright © 2009, Cold Spring Harbor Laboratory Press