A Computational Screen for Methylation Guide snoRNAs in Yeast

Todd M. Lowe and Sean R. Eddy

Small nucleolar RNAs (snoRNAs) are required for ribose 2′-O-methylation of eukaryotic ribosomal RNA. Many of the genes for this snoRNA family have remained unidentified in *Saccharomyces cerevisiae*, despite the availability of a complete genome sequence. Probabilistic modeling methods akin to those used in speech recognition and computational linguistics were used to computationally screen the yeast genome and identify 22 methylation guide snoRNAs, snR50 to snR71. Gene disruptions and other experimental characterization confirmed their methylation guide function. In total, 51 of the 55 ribose methylated sites in yeast ribosomal RNA were assigned to 41 different guide snoRNAs.

The genome of the yeast *Saccharomyces cerevisiae* has been completely sequenced and is thought to contain about 6000 protein-coding genes (*1*). However, some of the largest eukaryotic gene families produce functional RNAs rather than protein products. For example, yeast contains about 140 tandemly repeated copies of ribosomal RNA (rRNA) genes (*1*) and 275 dispersed transfer RNA genes (*2*).

The small nucleolar RNAs (snoRNAs) (*3*) are involved at various stages of eukaryotic ribosome biogenesis within the nucleolus (*4*). Ribosomal RNA undergoes cleavages and dozens of nucleotide modifications before assembly with ribosomal proteins into the mature ribosome (*4*). The two major families of snoRNAs appear to be involved in the two most common types of rRNA modification: Box H/ACA snoRNAs are required for specific rRNA pseudouridylations (*5*), and most of the C/D box snoRNAs appear to be involved in rRNA ribose methylation (*6*, *7*). A small number of snoRNAs in each family are involved in other steps of rRNA processing (*3*).

Conserved among all C/D box snoRNAs, the C and D box sequence motifs are required for snoRNA nucleolar localization, accumulation, and association with the ribonucleoprotein particle complexes that carry out rRNA processing (*8*). C/D box snoRNAs involved in ribose methylation also contain an internal "guide" sequence that is able to base pair with a specific segment of rRNA (Fig. 1). In association with protein cofactors, a guide snoRNA specifies the precise location for a particular 2′-O-ribose modification through its guide sequence (*6*, *9*).

Although ribose 2′-*O*-methyls are numerous in all studied eukaryotic rRNAs (*10*) and have been known for decades (*11*), the precise function of these modifications remains unknown. The total number of rRNA 2′-O-methylations in *Saccharomyces carlsbergensis*, a close relative of *S. cerevisiae*, has been estimated at 55 (*12*), 42 of which have been placed to specific nucleotide positions in the rRNA (*10*, *13*, *14*). In *S. cerevisiae*, 19 C/D box snoRNAs have been predicted to be responsible for methylation at 20 sites (*3*, *6*, *7*, *15*), a little more than one-third of the total rRNA ribose methyl groups. Experimental evidence supporting these predictions is available only for U24 (*6*). If the hypothesis is correct that snoRNAs guide most or all ribose 2′-O-methylations in eukaryotes, the majority of this snoRNA gene family remains unidentified in *S. cerevisiae*.

Because the *S. cerevisiae* genome is completely sequenced (*1*), it is reasonable to consider identifying methylation guide snoRNAs computationally. However, sequence similarity of snoRNAs across phyla and within the gene family is generally weak, so commonly used computational methods such as BLAST and FASTA fail to identify new genes by similarity to known snoRNAs. Attempts have been made to identify snoRNAs by pattern searches based on the rRNA complementary guide sequence and other conserved features, but feature consensus is poor, so this approach has had only modest success (*7*, *16*).

Formal probabilistic models, based in part on methods used in speech recognition and computational linguistics, have been introduced for searching for complicated consensus features in biological sequences (*17*). Hidden Markov models (HMM) (*18*) are probably the best known of these approaches. Another class of model called stochastic context-free grammars (SCFGs) has been used to construct probabilistic profiles of RNA consensus, allowing sensitive searching for RNA secondary structure (*19*). Using these probabilistic modeling techniques, we can produce an integrated model of snoRNAs that is based on the sequence features specific to this RNA gene family (Fig. 2).

To rapidly scan for 2′-O-methylation guide snoRNA candidates in the genome sequence, we used a greedy search algorithm. The program sequentially identified six components characteristic of these genes: box D, box C, a region of sequence complementary to rRNA, box D′ if the rRNA complementary region is not directly adjacent to box D, the predicted methylation site within the rRNA based on the complementary region, and the terminal stem base pairings, if present. The program also takes into account the relative distance between identified features within the snoRNA, information we found critical to reducing the false positive identification rate.

Each candidate snoRNA alignment was scored against our probabilistic model (Table 1). SnoRNAs were ranked on the basis of a final logarithmic odds score (*20*) that incorporated information from each of the snoRNA features. The initial model was trained on 35 human C/D box snoRNAs proposed to function as methylation guides (*6*). Nine previously isolated yeast snoRNAs matched this snoRNA gene model with significant scores (25.91 to 43.55 bits). In a search of randomly generated sequences (*21*) equivalent in size to four complete yeast genomes, the maximum score for a false positive (29.65 bits) exceeded the score for only one of the nine known snoRNAs. Thus, we

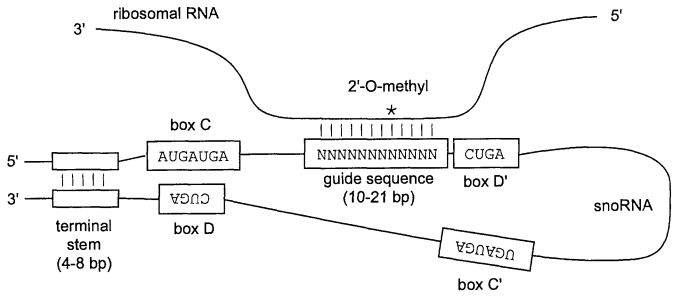

Fig. 1. C/D box methylation guide snoRNA consensus. The position of 2′-O-methylation of rRNA is within the helix formed by the complementary guide sequence of the snoRNA and precisely 5 nucleotides (nt) upstream of box D′ (or D, in snoRNAs that have their guide region adjacent to D instead of an internal D′ box). The C′ box feature was recognized (*29*) after the original snoRNA model was conceived; bp, base pairs.

Department of Genetics, Washington University School of Medicine, 4566 Scott Avenue, St. Louis, MO 63110, USA. E-mail: lowe@genetics.wustl.edu (T.M.L.) and eddy@genetics.wustl.edu (S.R.E.)

believed we had sufficient training data to search for unidentified snoRNAs in the yeast genome.

We began our search for previously unidentified snoRNAs by identifying family members that target 42 known 2′-*O*-methyl sites in *S. cerevisiae*, inferred from mapping data from *S. carlsbergensis* (*10*). Candidate snoRNAs were divided on the basis of target methyl site and sorted by score, producing 42 lists of best-to-worst snoRNA predictions, one for each methyl site. Depending on search parameter cutoffs and the specific target methyl site, the program found up to several dozen predictions for each methyl site. Candidates overlapping predicted protein-coding regions were noted and disfavored relative to other strong, nonoverlapping candidates. Seven previously published snoRNAs have been predicted to guide methylation at eight of the 42 sites (U14, U18, U24 at two sites, snR39, snR39b, snR40, and snR41) (*6*, *7*). Our searches did not show improved snoRNA predictions over the previously identified snoRNAs, so we did not pursue different assignments for these eight sites.

We tested the top scoring snoRNA gene predictions corresponding to the remaining sites by gene disruption (*22*). Each snoRNA-disrupted strain was tested for the ability to methylate at the predicted rRNA site by a deoxynucleoside triphosphate (dNTP) concentration-dependent primer extension assay (Fig. 3) (*23*). Out of 30 gene disruptions, 24 loci were verified as encoding methylation guide snoRNAs. Seven of these had been previously identified as C/D box snoRNAs, and 17 snoRNAs had not been previously identified (Table 2). Primer extension assays for two of the snoRNA disruption mutants, snR55 and snR70, showed a noticeable but minor change in the primer extension pattern at the expected sites (*24*); thus, we qualify these assignments as "inconclusive." None of these snoRNA gene disruptions were lethal, nor did we observe impaired growth on rich media.

Some of the 24 snoRNAs described above guide modification at more than one methyl site, as previously seen for U24 (*6*). The search program predicted and we experimentally verified one additional methylation target site for snR47, snR48, and snR51 (Table 2). We also found an additional target site for snR41 (Table 2), a snoRNA previously predicted to guide at a different methyl site (*6*). We verified snR41 methylation guide function for both the previously predicted site [small subunit (SSU) Gm1123] and the newly predicted site (SSU-Am541). With these additional site assignments, plus the eight previously assigned sites (*6*, *7*), 36 of 42 known methyl sites can be attributed to guide snoRNA.

Our searches gave no strong snoRNA candidates for four of the remaining six ribose methyl sites, all on large subunit (LSU) rRNA: Cm648, Gm1448, Am2279, and Gm2919. One common factor among these sites is that they are all one nucleotide 3′ to other ribose methyl sites that have confirmed or strong snoRNA assignments. Previous disruption of U24 results in unexpected loss of

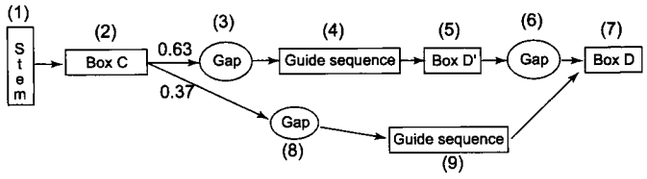

Fig. 2. Schematic of the probabilistic model. States (boxes and ovals) are connected by transitions (arrows). Each numbered state is a probabilistic model of a sequence feature (Table 1). Transition probabilities are 1.0, except those shown for transitions 2 → 3 and 2 → 8, which account for the proportion of snoRNAs with a guide sequence adjacent to box D′ and those with a guide sequence adjacent to box D, respectively.

Table 1. Summary of states within the snoRNA probabilistic model. State numbers correspond to Fig. 2. "Ungapped HMM" states represent fixed-length conserved sequence motifs. The state for the terminal stem is analogous but models base pairs rather than single positions [for example, an SCFG (*17*), instead of an HMM]. Duration models for gaps are estimated from binned length distributions (for example, the probability that a gap will be 11 to 20 nt, 21 to 30 nt, and so forth). The guide state is an HMM dependent on the rRNA target sequence; it includes terms for the probability of starting the complementarity at a given position relative to rRNA (this probability is high near known methylation positions), the length of the complementarity, and the probability of mismatches and noncanonical base pairs in the complementarity. For each state, the most common feature ("consensus") is shown to indicate the overall pattern we search for. The best, average, and worst feature scores are given for 41 methylation guide snoRNAs as an indication of the relative contribution of each state to the overall information in the model. For more details, see the program source code (*28*).

State number	Feature	Model	Consensus	Feature score (bits)		
				Best	Average	Worst
1	Terminal stem	SCFG, 4 to 8 bp	6 bp (when present)	7.60	3.09	0.35
2	Box C	7-bp ungapped HMM	AUGAUGA	12.73	11.63	5.84
3	Gap	Duration model	Length: 6 to 10 bp	−1.59	−2.09	−4.76
4	Guide sequence	HMM	12-bp duplex	15.67	11.11	2.54
5	Box D′	4-bp ungapped HMM	CUGA	7.34	4.85	−3.74
6	Gap	Duration model	Length: 36 to 45 bp	−1.59	−2.43	−5.36
7	Box D	4-bp ungapped HMM	CUGA	8.05	7.92	5.43
8	Gap	Duration model	Length: 56 to 75 bp	−1.50	−2.10	−4.17
9	Guide sequence	HMM	14-bp duplex	18.96	13.98	9.95

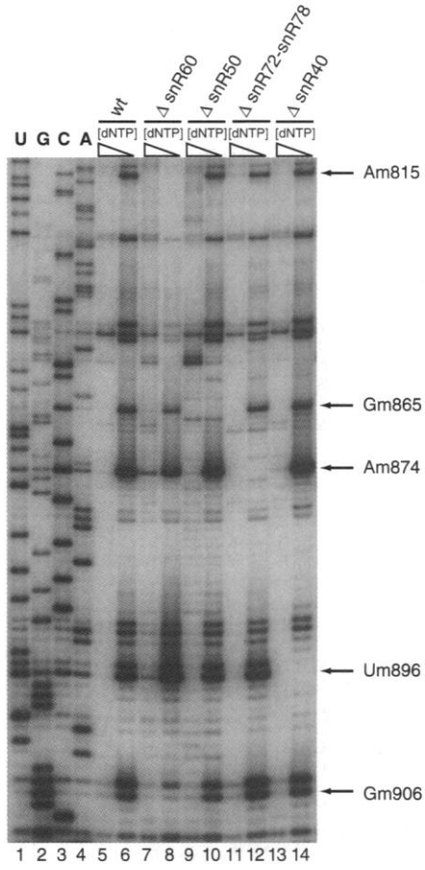

Fig. 3. Experimental confirmation of methylation guide function for yeast snoRNAs. 2′-O-methylation is detected as a reverse transcriptase primer extension stop on rRNA that occurs one nucleotide 3′ of the modified base in low but not high dNTP concentrations [for example, for wild-type (wt) rRNA, compare lane 5, high dNTP, to lane 6, low dNTP, for five methylated LSU rRNA positions indicated at the right]. Homologous disruption of a guide snoRNA (here, snR60, snR50, the snR72-snR78 array, and snR40 lanes) causes precise loss of corresponding methylation-dependent stops on the rRNA; for example, observe the loss of the Um896 band in the snR40 knockout in lane 14.

the methyl at Gm1448, adjacent to the predicted loss at Am1447 (*6*). Disruption of U18 results in loss at both Am647 and Cm648 (*25*). Our disruption of snR13 resulted in loss at Am2278 and Am2279. snR52 is strongly predicted, although not confirmed, to guide methylation at Um2918, one nucleotide adjacent to Gm2919. In each of these cases, we hypothesize that a change in the snoRNA-rRNA base pairing could allow a single snoRNA to guide modification at the observed 3′ adjacent ribose 2′-*O*-methyl sites. A previous alternative proposal suggests that an independent methyltransferase catalyzes addition of the 3′ adjacent methyl groups and that the reaction is dependent on the existence of the snoRNA-guided 5′ methyl sites (*6*).

We then turned to the 13 ribose methyl sites whose exact positions in the rRNA were not known. We used three lines of evidence to predict and then experimentally verify the position as well as the snoRNA assignment for 12 of the 13 unmapped sites. First, between one and five nucleotides of sequence context surrounding each of these methyl sites are known from ribonuclease T_1 fingerprints (*12*, *14*). Second, we checked the existing collection of C/D box snoRNAs for previously unrecognized rRNA complementary regions that could target sites not included in the list of known ribose 2′-*O*-methyls. Third, we went back to the *S. cerevisiae* genome search results from our program and extracted all high-scoring snoRNAs that could target previously unidentified rRNA methyl sites.

For each newly predicted methyl site, we experimentally checked for a rRNA primer extension pause typical of ribose 2′-*O*-methyls (*23*). For supported methyl sites, we then disrupted the corresponding guide snoRNA (all except snR38) to confirm anticipated loss of the methylation (Table 2; snoRNAs assigned to methyl sites in boldface). Six sites were assigned to known C/D box snoRNAs (Table 2; snR40, for example) and six to newly identified snoRNAs (Table 2; snR58, for example). Each of these 12 ribose methyl sites could be correlated with a T_1-RNase digest fragment for one of the 13 unmapped ribose methyls (*14*). We could not identify the location of the single unmapped methyl site in SSU rRNA (T_1 fragment GmU). snR190 has also been predicted to target a potential methyl site at LSU-Gm2393 (*6*). In our primer extension assay, this site did not give a visible band, nor did its sequence context correspond to an unassigned T_1 fragment. None of the verified guide snoRNAs were found to be essential, nor did gene disruption cause noticeably impaired growth.

Thus, in summary, we can attribute snoRNA-directed modification to 51 of the 55 ribose 2′-*O*-methyls in yeast rRNA (Table 2). This leaves four sites for which we could not assign a prediction (SSU-Am436), locate the methyl site (SSU-Gm?), or experimentally verify a prediction (LSU-Um2918 and LSU-Gm2919). Protein methyltransferases targeting these specific sites may account for our difficulty in finding or verifying guide snoRNAs in these cases.

From the perspective of the RNA gene family, we count 41 total guide snoRNAs assigned to 51 rRNA methylation sites (Table 2), 22 snoRNAs of which we identified in this work. We estimate that up to two methylation guide snoRNAs remain to be identified for the two unassigned methylation sites (SSU-Am436 and SSU-Gm?) and that two to four snoRNAs may be identified as being redundant with known snoRNAs for SSU-Um1265, SSU-Cm1637, LSU-Um2918, and LSU-Gm2919.

With nearly all 2′-O-methylation guide snoRNAs identified, we can assess the general genomic organization of the gene family (Table 2). Most are dispersed as independent singlets or within five small clusters of two to

Table 2. C/D box snoRNAs in *S. cerevisiae* that function as methylation guides. Previously unidentified snoRNAs or methylation sites are in boldface. Previously identified snoRNAs that have now been determined to be methylation guides are in italics. "Match/Mismatch" column refers to the number of base pairings (G-U included) and mismatches found within the snoRNA complementary region–rRNA duplex. "Len" refers to the known or predicted mature snoRNA length in nucleotides. "Position" and "Chr" refer to the 5′ end and chromosomal genomic locus according to the current version of the yeast genome available at the *S. cerevisiae* database (*30*). Strand designations: W, Watson, upper/forward strand; C, Crick, lower/complement strand. Ribosomal RNA positions are numbered as in (*10*). The last column gives the GenBank accession numbers.

snoRNA	Target 2′-O-Methyl Site	Predicted Target Site	Verified Target Site	Match/ Mismatch	Genomic Placement	Len	Chr	Position	Genbank Acc #
U14	SSU-Cm414	⊕ (6)	ND	13/1	cluster 5	123	10	139388 (C)	X96815
U18	LSU-Am647	⊕ (6)	⊕ (25)	15/0	intronic	102	1	142357 (W)	U12981
	LSU-Cm648		⊕ (25)						
U24	LSU-Cm1435	⊕ (6)	⊕ (6)	14/0	intronic	87	13	500071 (C)	Z48760
	LSU-Am1447	⊕ (6,7)	⊕ (6)	12/0					
	LSU-Gm1448		⊕ (6)						
snR13	LSU-Am2278	•	•	10/0	extragenic	107	4	1393187 (W)	U16692
	LSU-Am2279		•						
snR38	**LSU-Gm2812**	⊕ (6)	ND	13/0	intronic	95	11	282830 (W)	U26012
snR39	LSU-Am805	⊕ (6, 7)	ND	13/0	intronic	89	7	365249 (C)	U26011
snR39b	LSU-Gm803	⊕ (7)	•	14/0	extragenic	95	7	366466 (C)	X94605
snR40	SSU-Gm1267	⊕ (6,7)	•	11/1	extragenic	97	14	89208 (W)	U26015
	LSU-Um896	•	•	12/0					
snR41	SSU-Am541	•	•	11/1	cluster 3	105	16	719237 (C)	U26016
	SSU-Gm1123	⊕ (6,7)	•	20/1					
snR47	SSU-Am619	•	•	11/0	extragenic	99	4	541738 (C)	U56648
	LSU-Am2218	⊕ (25)	•	12/0					
snR48	LSU-Gm2788	•	•	17/1	extragenic	112	7	609578 (W)	AF06461
	LSU-Gm2790	•	•	15/0					
snR50 (Z14)	LSU-Gm865	•	•	12/0	extragenic	89	15	259489 (W)	AF06462
snR51	SSU-Am100	•	•	16/1	cluster 3	107	16	718803 (C)	AF06463
	LSU-Um2726	•	•	14/1					
snR52 (Z13)	SSU-Am420	•	•	13/0	extragenic	92	5	431217 (C)	AF06464
	LSU-Um2918	•	No	11/1					
	LSU-Gm2919	•	No						
snR53	SSU-Am796	•	•	11/0	cluster 4	91	5	61699 (W)	AF06465
snR54	SSU-Am973	•	•	13/0	intronic	86	13	163620 (C)	AF06466
snR55 (Z10)	SSU-Um1265	•	⊗	12/0	cluster 2	98	12	794793 (C)	AF06467
snR56	SSU-Gm1425	•	•	11/0	extragenic	86	2	88181 (W)	AF06468
snR57	SSU-Gm1570	•	•	15/0	cluster 2	88	12	795023 (C)	AF06469
snR58 (Z12)	**LSU-Cm661**	•	•	13/0	extragenic	96	15	136182 (C)	AF06470
snR59	LSU-Am805	•	ND	14/0	intronic	78	16	173826 (W)	AF06471
snR60 (Z15)	LSU-Am815	•	•	10/0	extragenic	104	10	348929 (C)	AF06472
	LSU-Gm906	•	•	19/0					
snR61 (Z11)	LSU-Am1131	•	•	11/0	cluster 2	90	12	794574 (C)	AF06473
snR62	LSU-Um1886	•	•	14/0	extragenic	100	15	409863 (C)	AF06474
snR63	LSU-Am2254	•	•	12/0	extragenic	255	4	323470 (C)	AF06475
snR64	**LSU-Cm2335**	•	•	11/0	extragenic	101	11	38812 (W)	AF06476
snR65	**LSU-Um2345**	•	•	11/0	extragenic	100	3	175909 (W)	AF06477
snR66 (Z16)	LSU-Um2415	•	•	12/0	extragenic	86	14	586088 (W)	AF06478
snR67	**LSU-Gm2616**	•	•	11/2	cluster 4	82	5	61352 (W)	AF06479
	LSU-Um2721	•	•	11/0					
snR68	LSU-Am2637	•	•	12/0	extragenic	136	9	97111 (W)	AF06480
snR69	LSU-Cm2945	•	•	18/3	extragenic	101	11	364418 (W)	AF06481
snR70	SSU-Cm1637	•	⊗	9/1	cluster 3	164	16	719047 (C)	AF06482
snR71	LSU-Am2943	•	•	9/1	extragenic	89	8	411228 (W)	AF06483
snR72 (Z2)	**LSU-Am874**	•	•	14/0	cluster 1	91	13	298554 (W)	Z69294
snR73 (Z3)	**LSU-Cm2956**	•	•	12/1	cluster 1	103	13	298306 (W)	Z69295
snR74 (Z4)	SSU-Am28	•	•	13/0	cluster 1	80	13	298138 (W)	Z69296
snR75 (Z5)	**LSU-Gm2286**	•	•	11/0	cluster 1	85	13	297915 (W)	Z69297
snR76 (Z6)	LSU-Cm2195	•	•	13/1	cluster 1	101	13	297727 (W)	Z69298
snR77 (Z7)	SSU-Um578	•	•	14/0	cluster 1	84	13	297506 (W)	Z69299
snR78 (Z8)	**LSU-Um2419**	•	•	12/0	cluster 1	82	13	297277 (W)	Z69300
snR79 (Z9)	SSU-Cm1006	•	•	16/1	extragenic	85	12	348511 (C)	Z70300

•, data presented in this work; ⊗, tentative assignment due to inconclusive assay for methyl loss; ND, not determined; ⊕, previously identified, references in parentheses.

seven tandemly arrayed guide snoRNAs. A total of 19 singlets occur outside of known protein-coding genes, presumably as independent transcription units. All tandemly arrayed snoRNAs within the same cluster are oriented on the same strand, and recent results indicate that these genes are polycistronic (*26*). Six yeast snoRNAs occur within the introns of host protein genes, all on the pre-mRNA coding strand. The mixture of snoRNAs in yeast occurring within introns and tandem arrays and as singlets is in contrast to vertebrates, where all currently known guide snoRNAs are within host gene introns. Polycistronic arrays of snoRNAs have also been reported in plants (*27*). Some plant polycistrons contain a mix of snoRNAs from both major families of guide snoRNAs (C/D box and H/ACA box snoRNAs), whereas none of the yeast tandem arrays contain members outside of the C/D box family.

It is possible that a large number of noncoding RNAs remain to be discovered. Both computational screens and experimental screens tend to be biased against RNAs. Many functional RNAs are not polyadenylated, so they are not well represented in oligo(dT) primed cDNA libraries or in expressed sequence tag sequencing projects. Often the genes for RNAs are small and may occur in multiple copies. RNAs are of course not affected by stop codons or frameshifts, so they are probably somewhat refractory to genetic screens. Most functional RNAs known today have been identified by biochemical means, but these approaches are best suited to abundant RNAs. Using probabilistic modeling methods, we are beginning to gather the tools necessary to computationally screen genome sequences for noncoding RNAs.

References and Notes

1. A. Goffeau *et al.*, *Science* **274**, 546 (1996).
2. T. M. Lowe and S. R. Eddy, *Nucleic Acids Res.* **25**, 955 (1997).
3. A. G. Balakin, L. Smith, M. J. Fournier, *Cell* **86**, 823 (1996); D. Tollervey and T. Kiss, *Curr. Opin. Cell Biol.* **9**, 337 (1997); J.-P. Bachellerie and J. Cavaille, *Trends Biochem. Sci.* **22**, 257 (1997); C. M. Smith and J. A. Steitz, *Cell* **89**, 669 (1997).
4. A. A. Hadjiolov, in *Cell Biology Monographs*, M. Alfert *et al.*, Eds. (Springer-Verlag, Vienna, 1985), pp. 1–268; J. L. Woolford Jr., *Adv. Genet.* **29**, 63 (1991).
5. J. Ni, A. L. Tien, M. J. Fournier, *Cell* **89**, 565 (1997); P. Gannot, M.-L. Bortolin, T. Kiss, *ibid.*, p. 799.
6. Z. Kiss-Laszlo, Y. Henry, J.-P. Bachellerie, M. Caizergues-Ferrer, T. Kiss, *ibid.* **85**, 1077 (1996).
7. M. Nicoloso, L.-H. Qu, B. Michot, J.-P. Bachellerie, *J. Mol. Biol.* **260**, 178 (1996).
8. A. Jarmolowski, J. Zagorski, H. V. Li, M. J. Fournier, *EMBO J.* **9**, 4503 (1990); N. J. Watkins, D. R. Newman, J. F. Kuhn, E. S. Maxwell, *RNA* **4**, 582 (1998); T. S. Lang, A. Borovjagin, E. S. Maxwell, S. A. Gerbi, *EMBO J.* **17**, 3176 (1998); D. A. Samarsky, M. J. Fournier, R. H. Singer, E. Bertrand, *ibid.*, p. 3747.
9. J. Cavaille, M. Nicoloso, J.-P. Bachellerie, *Nature* **383**, 732 (1996).
10. B. E. H. Maden, *Prog. Nucleic Acids Res. Mol. Biol.* **39**, 241 (1990).
11. H. Singh and B. G. Lane, *Can. J. Biochem.* **42**, 1011 (1964).
12. J. Klootwijk and R. J. Planta, *Eur. J. Biochem.* **39**, 325 (1973).
13. H. A. Raue, J. Klootwijk, W. Musters, *Prog. Biophys. Mol. Biol.* **51**, 77 (1988).
14. G. M. Veldman, J. Klootwijk, V. C. H. F. de Regt, R. J. Planta, *Nucleic Acids Res.* **9**, 6935 (1981).
15. J. Cavaille and J.-P. Bachellerie, *ibid.* **26**, 1576 (1998).
16. M. Nicoloso, M. Caizergues-Ferrer, B. Michot, M. C. Azum, J.-P. Bachellerie, *Mol. Cell. Biol.* **14**, 5766 (1994).
17. R. Durbin, S. R. Eddy, A. Krogh, G. Mitchison, *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids* (Cambridge Univ. Press, Cambridge, 1998).
18. S. R. Eddy, *Curr. Opin. Struct. Biol.* **6**, 361 (1996).
19. _____ and R. Durbin, *Nucleic Acids Res.* **22**, 2079 (1994); Y. Sakakibara *et al.*, *ibid.*, p. 5112.
20. C. Barrett, R. Hughey, K. Karplus, *Comput. Appl. Biosci.* **13**, 191 (1997).
21. Random sequences were generated by a fifth-order Markov chain based on 6-mer frequencies within the yeast genome.
22. snoRNA disruptions were generated by homologous gene replacement in *S. cerevisiae* haploid strain yM4585 (*Mat a his3Δ200 lys2-801 leu2-3,2-112 trp1-901 tyr1-501 URA3+ ADE2+ CAN^S*) and diploid strain yM4587 (*Mat a/Mat α his3Δ200/his3Δ200 lys2-801/ lys2-801 leu2-3,2-112/ leu2-3,2-112 trp1-901/ trp1-901 tyr1-501/ tyr1-501 URA3+/ URA3+ ADE2+/ ADE2+ CAN^S/ can^r*). A polymerase chain reaction (PCR)–based HIS3 insertion scheme was used as described by A. Baudin, O. Ozier-Kalogeropoulos, A. Denouel, F. Lacroute, and C. Cullin [*Nucleic Acids Res.* **21**, 3329 (1993)], with a protocol provided by L. Riles. Transformants growing on yeast extract, peptone, and dextrose His plates were picked and assayed by PCR for correct integration of the HIS3 marker gene replacing the target snoRNA.
23. Mapping of rRNA 2′-*O*-methyl sites was based on previously described methods from B. E. H. Maden, M. E. Corbett, P. A. Heeney, K. Pugh, and P. M. Ajuh [*Biochemie* **77**, 22 (1995)] and (*6*). The sequences of all mapping primers are available on the World Wide Web (WWW) (*28*). Total yeast RNA (0.4 μg/μl) was annealed with end-labeled mapping primers (0.15 pmol/μl) at 60°C for 4 min. Primer extensions were carried out in 5-μl reactions containing 0.8 μg of RNA and 0.3 of pmol ^{32}P–end-labeled primer in the presence of 50 mM tris-Cl (pH 8.6), 60 mM NaCl, 9 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM concentrations of each dNTP, and avian myeloblastosis virus reverse transcriptase (0.2 U/μl) for 30 min at 37°C. Low dNTP concentration reactions were carried out in the same manner except with 0.004 mM concentrations of each dNTP and 5 mM $MgCl_2$. Each reaction was analyzed by electrophoresis next to an RNA sequencing ladder on an 8% polyacrylamide gel.
24. T. M. Lowe and S. R. Eddy, data not shown. Primer extension gels are available on the WWW (*28*).
25. J. Ni, A. Balakin, M. J. Fournier, personal communication.
26. E. Petfalski *et al.*, *Mol. Cell. Biol.* **18**, 1181 (1998); G. Chanfreau, G. Rotondo, P. Legrain, A. Jacquier, *EMBO J.* **17**, 3726 (1998); G. Chanfreau, P. Legrain, A. Jacquier, *J. Mol. Biol.* **284**, 975 (1998); L. H. Qu *et al.*, *Mol. Cell. Biol.* **19**, 1144 (1999).
27. D. J. Leader *et al.*, *EMBO J.* **16**, 5742 (1997); P. J. Shaw, A. F. Beven, D. J. Leader, J. W. Brown, *J. Cell Sci.* **111**, 2121 (1998).
28. All computer codes, snoRNA search results, primer sequences, gel images, and other referenced data can be accessed on the WWW at rna.wustl.edu/snoRNAdb/
29. Z. Kiss-Laszlo, Y. Henry, T. Kiss, *EMBO J.* **17**, 797 (1998).
30. The *S. cerevisiae* database is available at: genome-www.stanford.edu/Saccharomyces/

31. We thank L. Lutfiyya and L. Riles from the M. Johnston lab for protocols and guidance in all aspects of yeast handling, gene disruptions, and colony PCR and RNA preparations; M. Fournier, J. Ni, D. Samarsky, and B. E. H. Maden for helpful discussions and sharing of unpublished observations; S. Johnson and L. Lutfiyya for careful reading of the manuscript; and M. Johnston for providing the haploid strain yM4585 and diploid strain yM4587. Supported by NIH grant R01-HG01363 and by a gift from Eli Lilly.

27 May 1998; accepted 19 January 1999

Novel Endotheliotropic Herpesviruses Fatal for Asian and African Elephants

Laura K. Richman,[1,2]* Richard J. Montali,[1] Richard L. Garber,[3] Melissa A. Kennedy,[4] John Lehnhardt,[1] Thomas Hildebrandt,[5] Dennis Schmitt,[6] Douglas Hardy,[7] Donald J. Alcendor,[2] Gary S. Hayward[2]

A highly fatal hemorrhagic disease has been identified in 10 young Asian and African elephants at North American zoos. In the affected animals there was ultrastructural evidence for herpesvirus-like particles in endothelial cells of the heart, liver, and tongue. Consensus primer polymerase chain reaction combined with sequencing yielded molecular evidence that confirmed the presence of two novel but related herpesviruses associated with the disease, one in Asian elephants and another in African elephants. Otherwise healthy African elephants with external herpetic lesions yielded herpesvirus sequences identical to that found in Asian elephants with endothelial disease. This finding suggests that the Asian elephant deaths were caused by cross-species infection with a herpesvirus that is naturally latent in, but normally not lethal to, African elephants. A reciprocal relationship may exist for the African elephant disease.

In 1997, there were an estimated 291 Asian elephants and 193 African elephants in North America. About 115 elephant births have occurred, mostly since 1960, but perinatal deaths and stillbirths have been and remain exceedingly high. Because elephants from the wild are no longer accessible for import, captive breeding populations could be repro-