

Birth of a chimeric primate gene by capture of the transposase gene from a mobile element

Richard Cordaux*, Swalpa Udit†, Mark A. Batzer*, and Cédric Feschotte†*

*Department of Biological Sciences, Biological Computation and Visualization Center, Center for BioModular Multi-Scale Systems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803; and †Department of Biology, University of Texas, Arlington, TX 76019

Edited by Susan R. Wessler, University of Georgia, Athens, GA, and approved March 27, 2006 (received for review February 10, 2006)

The emergence of new genes and functions is of central importance to the evolution of species. The contribution of various types of duplications to genetic innovation has been extensively investigated. Less understood is the creation of new genes by recycling of coding material from selfish mobile genetic elements. To investigate this process, we reconstructed the evolutionary history of *SETMAR*, a new primate chimeric gene resulting from fusion of a *SET* histone methyltransferase gene to the transposase gene of a mobile element. We show that the transposase gene was recruited as part of *SETMAR* 40–58 million years ago, after the insertion of an *Hsmar1* transposon downstream of a preexisting *SET* gene, followed by the *de novo* exonization of previously noncoding sequence and the creation of a new intron. The original structure of the fusion gene is conserved in all anthropoid lineages, but only the N-terminal half of the transposase is evolving under strong purifying selection. *In vitro* assays show that this region contains a DNA-binding domain that has preserved its ancestral binding specificity for a 19-bp motif located within the terminal-inverted repeats of *Hsmar1* transposons and their derivatives. The presence of these transposons in the human genome constitutes a potential reservoir of $\approx 1,500$ perfect or nearly perfect *SETMAR*-binding sites. Our results not only provide insight into the conditions required for a successful gene fusion, but they also suggest a mechanism by which the circuitry underlying complex regulatory networks may be rapidly established.

transposable elements | gene fusion | molecular domestication | DNA binding | regulatory network

As well characterized pathway for the emergence of new genes is through duplication of preexisting genes (1), resulting, for example, from segmental duplication (2) or retrotransposition (3). A much less understood, yet ostensibly recurrent, source of genetic innovation is the recycling of coding material from selfish mobile genetic elements (4–9). Mobile or transposable elements are “jumping genes,” pieces of DNA that can move and replicate within the genomes of virtually all living organisms (10). These elements are often considered “selfish” because they encode proteins devoted to their own propagation (or those of related elements), and they normally do not provide a selective advantage to the host organism carrying them (11). Despite (or because of) their selfish nature, transposable elements have had a considerable impact on the evolution of their host genomes; for example, as a result of insertional mutagenesis or by promoting genomic rearrangements (12, 13). In this study, we investigated another process by which mobile elements can impact their host genomes, often referred to as “molecular domestication,” whereby their coding sequences are recruited to serve a cellular function in their host genomes (4–9). Specifically, we reconstructed the evolutionary steps leading to the birth of *SETMAR*, a new chimeric primate gene located on human chromosome 3p26 (14).

SETMAR first was identified as a chimeric mRNA transcript consisting of two exons encoding a SET domain protein fused in-frame with the entire transposase-coding region of a *mariner*-like *Hsmar1* transposon (the MAR region) (14). The major

SETMAR transcript, which consists of these three exons, is predicted to encode a protein of 671 amino acids and is supported by 48 human cDNA clones from 18 different normal and/or cancerous tissues (Table 1, which is published as supporting information on the PNAS web site; refs. 14 and 15). These data suggest that the *SETMAR* protein is broadly expressed and has an important, yet unknown, function in human. Recently, it was shown that the SET domain of the *SETMAR* protein exhibits histone methyltransferase activity (15), as do all known SET domains (16, 17). By contrast, the function contributed by the MAR region to *SETMAR* (if any) is unclear. The MAR transposase region has evolved significantly slower than that of other human *Hsmar1* copies, being only 2.4% divergent from the ancestral *Hsmar1* gene (vs. $\approx 8\%$ average divergence for other *Hsmar1* transposases; ref. 14). Here, we used a combination of evolutionary and functional approaches to determine how the fusion between the SET and MAR regions occurred and to assess the contribution of the transposase to the *SETMAR* function.

Results and Discussion

Transposon Insertions at the Origin of *SETMAR*. Reciprocal BLASTP database searches and sequence comparisons of flanking regions by using the University of California, Santa Cruz, genome browser (<http://genome.cse.ucsc.edu>) revealed putative orthologs of the *SET* region of *SETMAR* in mouse, rat, dog, cow, opossum, and zebrafish. For example, the putative ortholog in zebrafish shows 45% identity with the human *SET* region of *SETMAR*, and the alignment of the zebrafish transcript (GenBank accession no. AL919348) with its corresponding genomic locus shows that this gene is interrupted by a single intron at the same position as in the human *SET* region of *SETMAR* and in the other mammalian *SET* homologs. None of these *SET* genes is followed by an *Hsmar1* transposon and all are apparently expressed as a transcript containing two exons and encoding a ≈ 300 -aa protein containing only the SET domain (Fig. 1). This result indicates that the *SET* region of *SETMAR* preexisted in the ancestral primate genome and that the MAR region was subsequently added downstream of the *SET* region during primate evolution.

To pinpoint the birth of *SETMAR* within the primate lineage, we cloned and sequenced the orthologous MAR region in eight primate species. Sequence analysis revealed the presence of the *Hsmar1* transposon in all anthropoid lineages (humans, apes, and Old and New World monkeys) but not in tarsier (Fig. 1; see also Fig. 4, which is published as supporting information on the PNAS

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: MBP, maltose-binding protein; TIR, terminal inverted repeat.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. DQ341316–DQ341331).

See Commentary on page 7941.

†To whom correspondence should be addressed. E-mail: cedric@uta.edu.

© 2006 by The National Academy of Sciences of the USA

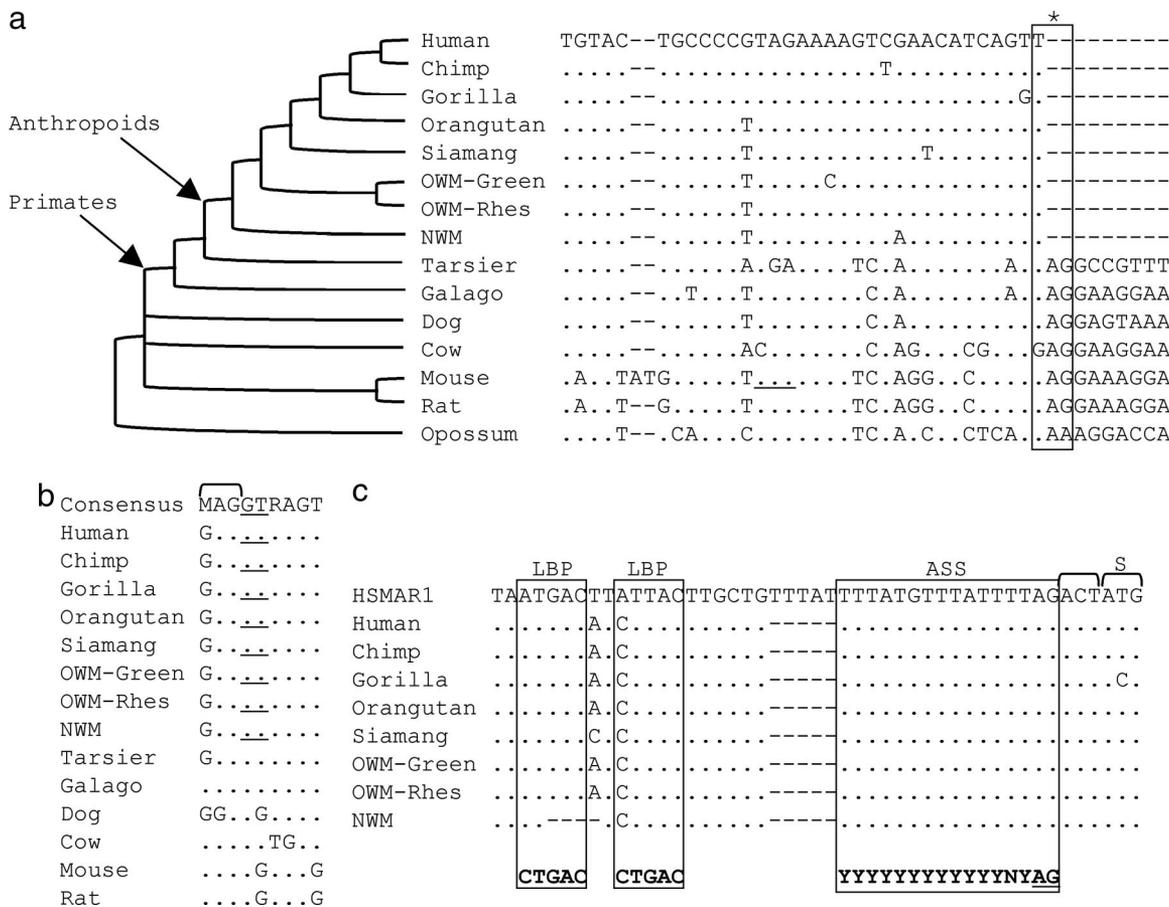


Fig. 2. Molecular events leading to the birth of *SETMAR*. (a) Schematic phylogeny (Left) and multiple alignment of the 3' end of *SET* exon 2 (Right) in 10 primates and 5 nonprimate mammals [OWM, Old World monkeys (Green, African green monkey; Rhes, Rhesus macaque); NWM, New World monkey (owl monkey)]. Dots indicate the identity with the top sequence, and hyphens denote sequence gaps. The asterisk and the box indicate the position of the ancestral *SET* stop codon (TAG) that is conserved in all mammals (except cow) but was removed by a deletion in anthropoid primates. In cow, the current stop codon is located five codons downstream of the original stop codon (data not shown), and in mouse, a 2-bp insertion resulted in a premature stop codon (underlined). (b) Multiple alignment of the 5' donor splice site of *SETMAR* intron 2. The human consensus splice motif (19) is used as a reference (top line). The GT dinucleotide after the last *SET* exon 2 codon (GAG) in anthropoids and delimiting the start of *SETMAR* intron 2 (in phase 0) is underlined. (c) Multiple alignment of the 5' end of the *MAR* coding region in anthropoids. The *Hsmar1* transposon family consensus (14) is used as a reference (top line). The two putative lariat branch points (LBP) and the 3' acceptor splice site (ASS) are boxed. The human consensus LBP and ASS motifs (19) are shown in bold below the boxes. The AG dinucleotide delimiting the end of *SETMAR* intron 2 is underlined. The *MAR* exon of *SETMAR* starts with codon ACT, located immediately before and in frame with the putative start codon (S) of the ancestral *Hsmar1* transposase.

conserved in the *Hsmar1* consensus sequence (Fig. 2c) and (ii) we identified three unrelated chimeric RNA transcripts encoded elsewhere in the human genome in which a 3' acceptor splice site located at the same position is used to fuse *Hsmar1* transposase sequences to different upstream exons (Fig. 6, which is published as supporting information on the PNAS web site). However, none of these three chimeric transcripts have significant coding capacity in the transposase region because of premature stop codons. By contrast, the positions of the cryptic splice sites in *SETMAR*, together with the size of the deletion preceding the donor splice site, coincidentally allowed for the translational fusion of the *SET* and *MAR* ORFs. Equally remarkable is the fact that all of the mutational events and mechanisms leading to the assembly of the novel *SETMAR* gene took place within a very narrow evolutionary time window of <18 million years, after the emergence of the tarsier lineage and before the diversification of anthropoid primates (ref. 18; Fig. 1).

Functional Contribution of the *MAR* Transposase. To gain insight into what property of the transposase may have provided a selective advantage to be recruited as part of *SETMAR*, we compared the rate of nonsynonymous (K_A) and synonymous (K_S) nucleotide

substitutions per site of the *MAR* region of each anthropoid lineage by using a likelihood ratio test (21). This analysis revealed that the K_A/K_S ratios among the different primate lineages are not significantly different (Table 2, which is published as supporting information on the PNAS web site). The best fit to the data was obtained for $K_A/K_S = 0.3$, which is significantly <1 (Table 2). This result suggests that the *MAR* region has evolved consistently under purifying selection in all anthropoid lineages. However, when the analysis was performed separately for the 5' and 3' halves of the *MAR* region, the 5' half displayed a very strong signal of purifying selection ($K_A/K_S = 0.1$, significantly <1; Table 2), whereas the 3' half displayed a signal of neutral evolution ($K_A/K_S = 0.7$, not significantly different from 1; Table 2). Thus, the *MAR* region may have been recruited for a function located in the N-terminal region of the transposase.

All eukaryotic transposases studied thus far contain two major functional domains. The N-terminal region is responsible for DNA binding to the TIRs of their cognate transposons, whereas the C-terminal region contains the catalytic domain responsible for cleavage and joining reaction of the so-called cut-and-paste transposition reaction (10). Therefore, the *MAR* region might have been recruited for its DNA-binding capabilities rather than

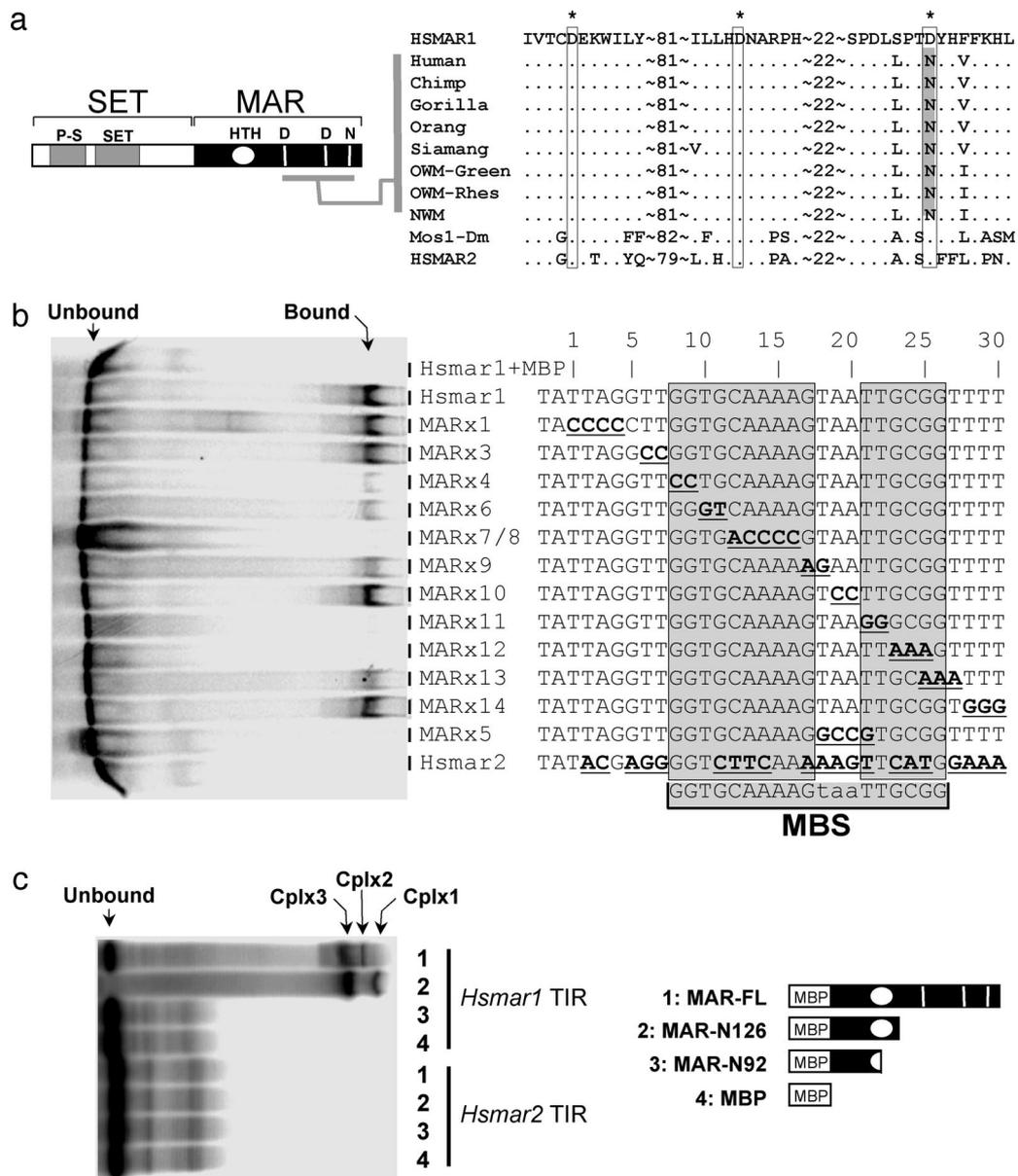


Fig. 3. *In vitro* DNA-binding activity and specificity of the MAR domain of SETMAR. (a) Schematic representation of the SETMAR protein and its predicted features: pre-SET (*p-S*), helix–turn–helix motif (HTH), and DDN triad (positions of the original catalytic amino acid triad of the MAR region). The protein multiple alignment on the right shows that the triad is DD34N (*) in all of the SETMAR protein sequences examined (naming convention as in Fig. 2) instead of the typical DD34D motif of the *Hsmar1* and *Hsmar2* consensus transposase sequences (14, 22) and all known active *mariner* transposases, such as *mos1* from *Drosophila melanogaster* (*Mos1-Dm*). Dots indicate identity with top sequence, and numbers indicate the number of amino acids between the sequence portions shown. (b) *In vitro* DNA-binding activity and specificity of purified MAR protein domain. EMSA of various TIR double-stranded oligonucleotides mixed with a purified recombinant peptide corresponding to MBP domain alone (top lane) or to the entire MAR region fused to a N-terminal MBP domain (all other lanes). The TIR oligonucleotides were designed by using the consensus *Hsmar1* or *Hsmar2* sequences (14, 22) and their characteristic flanking TA target site duplication. Base substitutions relative to the *Hsmar1* TIR are in bold and underlined. The EMSA autoradiography shows shifted DNA (bound) on the right side of the gel, whereas input DNA (unbound) is on the left side. MARx7/8 corresponds to a mixture of two oligonucleotides, none of which are bound by the purified protein. (c) Mapping of the MAR region involved in DNA binding. EMSA of either the *Hsmar1* or *Hsmar2* TIR oligonucleotides with four recombinant purified peptides corresponding to the entire MAR peptide (lane 1), the first 126 (lane 2) or 92 (lane 3) aa of the MAR peptide fused to a N-terminal MBP tag, or the MBP alone (lane 4). Two shifted bands can be seen when the *Hsmar1* TIR oligonucleotide is mixed with either peptide 1 or peptide 2. Based on previous *in vitro* studies of *mariner* DNA-binding activities (23–26), we interpret complex (Cplx 3) as a single oligonucleotide with a protein dimer, whereas the upper bands may correspond to tetramers of protein bound to single (Cplx 2) or paired (Cplx 1) oligonucleotides.

for its catalytic activities. In support of this hypothesis, the third D amino acid constituting the DD34D catalytic triad of *mariner* transposases (10) is mutated to N in all MAR primate sequences examined, whereas the triad is conserved in the consensus *Hsmar1* sequence (Fig. 3a; see also Fig. 7, which is published as supporting information on the PNAS web site). Alteration of this amino acid in the *Drosophila mos1 mariner* transposase abolishes

its catalytic activity (27). Furthermore, we were unable to detect transposition of an artificial *Hsmar1* transposon upon forced expression of the MAR transposase by using an *in vivo* assay in *Escherichia coli* (28), whereas the same assay showed high frequency of transposition of the active *Himar1* element upon expression of its cognate transposase (data not shown).

By contrast with the C-terminal region, the N-terminal half of

MAR is highly conserved in all anthropoid lineages (Fig. 7), suggesting that the MAR protein may have retained its ancestral DNA-binding activity. Because all *mariner*-like transposases studied so far specifically bind to the TIRs of their cognate transposons (24, 25, 27, 30), we tested the ability of the MAR domain to recognize DNA sequence motifs identical or very similar to the TIR of *Hsmar1* elements. The entire human MAR peptide (343 aa) was expressed in *E. coli*, purified as a fusion protein with a N-terminal maltose-binding protein (MBP)-MAR domain and incubated with radiolabeled double-stranded DNA oligonucleotides corresponding to the consensus *Hsmar1* TIR (14). EMSA resulted in a strong DNA shift when MBP-MAR was mixed to the *Hsmar1* TIR (Fig. 3*b*, second lane), but no shift when only MBP was incubated (Fig. 3*b*, first lane). These results demonstrate that the MAR peptide interacts *in vitro* with the consensus TIR sequence of *Hsmar1* transposons. Additional EMSA with a series of mutant *Hsmar1* TIR oligonucleotides and the *Hsmar2* TIR (22) refined the location of the MAR-binding site (MBS) to a 19-bp motif within the consensus *Hsmar1* TIR sequence (Fig. 3*b*). The binding appears highly specific because replacing virtually any consecutive dinucleotides within this motif (except for TIR variant MARx10) drastically reduces the amount of shifted DNA (Fig. 3*b* and data not shown). The location of the MAR-binding site within the *Hsmar1* TIR is consistent with those of other *mariner* transposase-binding sites (26, 31).

To map the MAR region involved in DNA-binding activity, two deleted MAR recombinant peptides were purified and tested for DNA binding by using EMSA. The peptide MAR-N126 encompassing the first 126 aa of the MAR peptide and including a predicted helix–turn–helix (HTH) motif at amino acid positions 86–107 of MAR (Fig. 7), interacted specifically with the *Hsmar1* TIR (Fig. 3*c*). By contrast, the peptide MAR-N92 encompassing the first 92 aa of the MAR region and lacking the recognition helix of the predicted HTH did not yield any detectable protein–DNA interaction (Fig. 3*c*). Together these results demonstrate that the N-terminal domain of the MAR region has retained the ability to bind specifically *in vitro* to its ancestral binding site and that the region encompassing amino acid positions 93–126, which contains a predicted HTH motif, is critical for this interaction.

Conclusion

In sum, our results show that the transposase of a mobile element has become part of a functional primate gene through a stepwise evolutionary process involving transposition and subsequent transcriptional and translational fusion. Comparative sequence analysis and functional assays strongly suggest that selection has acted to preserve the specific DNA-binding activity of the ancestral transposase, whereas its catalytic activity has likely been lost. Interestingly, BLAST searches of the human genome sequence revealed the presence of 752 and 760 sequences identical and with a single mismatch, respectively, to the 19-bp MAR-binding site. These data suggest that the human genome contains an enormous reservoir of potential SETMAR binding sites, $\approx 97\%$ of which lie within the TIRs of recognizable *Hsmar1* transposons and their derivatives (data not shown). This observation raises the possibility that the recruitment of the MAR DNA-binding domain may have provided an opportunity for the corecruitment of a network of DNA binding sites to which the fusion SETMAR protein now could be tethered. The SET domain of SETMAR methylates histone H3 predominantly at lysine 36 (15), an epigenetic mark that in yeast has a repressive impact on transcription elongation by RNA polymerase II and prevents spurious intragenic transcription from cryptic promoters (32, 33). Thus, the transposase-derived DNA-binding domain of SETMAR may have provided a means to target methylation

of histone H3 at lysine 36 to particular sites in the genome where it could affect gene expression or other biological processes.

Materials and Methods

Evolutionary Experiments and Analyses. The 3' end of the *SET* exon 2 and the orthologous *MAR* region were amplified by PCR in eight primate species by using different primer sets (Tables 3 and 4, which are published as supporting information on the PNAS web site). PCR conditions and cycling were as described in ref. 34, except for the amount of DNA (50–100 ng) and number of cycles (Table 3). PCR products were cloned into TOPO-TA cloning vectors (Invitrogen), according to the manufacturer's recommendations. At least three clones of each PCR product were randomly selected and amplified by PCR with primers M13F and M13R (Invitrogen). DNA sequences were determined by using primers M13F and M13R (along with MAR-Fint and MAR-Rint for the MAR region) and resolved on an ABI3100 automatic DNA sequencer.

The sequences generated in this study have been deposited in GenBank under accession numbers DQ341316–DQ341331. All other sequences used were downloaded from the University of California, Santa Cruz, genome browser by using the latest available freezes of the human, rhesus macaque, dog, cow, mouse, rat, and opossum genome sequences. Sequences were aligned by using CLUSTALW, as implemented in BIOEDIT 7.0 (35), followed by manual adjustments. The alignment of the sequence flanking the *SET* exon 2 and the *MAR* region, along with BLAST searches against the human genome revealing unique significant matches corresponding to the expected SETMAR locus, confirmed the orthology between the different sequences obtained for each of the two loci.

The K_A/K_S tests of selection of the *MAR* region were performed by using PAML 3.14 (21). A model of a single rate for all sites was specified, and the tree was provided as follows: (((((human, chimpanzee), gorilla), orangutan), siamang), (rhesus macaque, African green monkey)), owl monkey, X), where X = consensus sequences for *Hsmar1* or *Mmmar1* (a mouse *mariner*-like transposon family) or no sequence. Different K_A/K_S ratio models were tested by using maximum-likelihood ratio tests. Log likelihoods of the models were compared with a χ^2 distribution with as many degrees of freedom as the difference in number of parameters of the compared models (21). Log likelihood of a model with different K_A/K_S ratio in every branch was compared with the model with a single K_A/K_S ratio to be estimated from the data. Next, log likelihood of a model with equal K_A/K_S ratio in all branches was compared with the model with that ratio fixed to one (i.e., neutrality). In addition, we compared log likelihoods of models allowing positive selection on individual sites (M2a) or not (M1a) (see Table 2).

Functional Experiments and Analyses. The plasmid pBM3.1, a gift from D. Lampe (Duquesne University, Pittsburgh), contains the entire *MAR* region of SETMAR (positions 985–2,013 of the SETMAR coding sequence in the human cDNA AF054989; 343 aa) cloned in frame with the MBP domain, between the XbaI and HindIII sites of the expression vector pMAL-c2x (New England Biolabs). The MAR deletion constructs used for mapping the DNA-binding domain were generated by PCR with pBM3.1 as a template with the forward primer XbaI-MAR (5'-ACTCTA-GAATGAAAATGATGTTAGACAAAAAGC-3') corresponding to an XbaI site fused to the first 25 bp of the MAR region with the reverse primer MARdelR1-HindIII (5'-ACAAGCTTTCAATTTTCAGTCAGTCATGAGGC-3') or MARdelR2-HindIII (5'-ACAAGCTTTCAAGCAACT-TCTCGTGTAGTTGTAAGG-3'). The corresponding PCR products were gel-purified and cloned in between the XbaI and HindIII sites of pMAL-c2x, resulting in pMAR-N92 and pMAR-N126, respectively. The integrity of all coding regions and

in-frame fusion with the MBP sequence was verified by DNA sequencing.

The plasmids pMAL-c2x, pBM3.1, pMAR-N126, and pMAR-N92 were transformed into *E. coli* strain BL21. For each recombinant clone, 100 ml of log-phase cultures were induced with 0.3 mM isopropyl β -D-thiogalactoside and grown for 2 h at 37°C. Cells were pelleted, frozen overnight, resuspended the next day in 5 ml of B-PER extraction reagent (Pierce), pelleted again, and the supernatant was mixed with 500 μ l of amylose resin (New England Biolabs) for 1 h at 4°C. After four washes with TWB buffer (20 mM Tris-HCl, pH 7.4/200 mM NaCl/1 mM EDTA/2 mM DTT), the bound protein was eluted with 500 μ l of buffer TWB containing 10% glycerol and 10 mM maltose. Purified proteins were analyzed by Coomassie staining and Western blotting with an MBP monoclonal antibody (R29.6 from Santa Cruz Biotechnology). This analysis shows that each purification procedure yielded a major peptide species corresponding in molecular mass to full-length MBP (42.5 kDa), MBP-MAR (83 kDa), MBP-MAR-N92 (53 kDa), and MAR-N126 (57 kDa) fusion proteins (data not shown).

EMSAs were performed by using synthetic double-stranded oligonucleotides (Integrated DNA Technologies, Coralville, IA), whose sequences are shown in Fig. 3. Complementary oligonucleotides were mixed at a concentration of 1.8 mM and end-labeled with gamma-P33 or gamma-P32 by using T4 kinase

(Invitrogen). The labeling reaction was stopped by heating to 95°C for 7 min, followed by slow cooling at room temperature to allow the annealing of the oligonucleotides. EMSA reactions were carried out in a total volume of 15 μ l containing 1 μ l of labeled DNA and 0.5 μ l of purified protein (\approx 0.5 ng) in a buffer containing 15 mM Tris (pH 7.5), 0.1 mM EDTA, 1 mM DTT, 0.3 mg/ml BSA, 0.1% Nonidet P-40, 10% glycerol, and 33 mg/ml single-stranded DNA. Reactions were incubated for 1 h at 25°C, and samples were separated by electrophoresis on 6% native polyacrylamide gels. Gels were visualized by autoradiography or by using a phosphorimager (FLA-3000G; Fujifilm). The two EMSA experiments shown in Fig. 3 were each repeated three times independently to verify the consistency of the results.

We thank D. Lampe for providing the SETMAR and MAR expression plasmids; T. Disotell (New York University) for the tarsier DNA sample; Z. Abdallah for technical support; E. Betrán, E. Pritham, and two anonymous reviewers for comments on the manuscript; and S. W. Herke and members of the Genome Biology Group at the University of Texas at Arlington for discussions. This research was supported by Louisiana Board of Regents Millennium Trust Health Excellence Fund Grants (2000-05)-05, (2000-05)-01, and (2001-06)-02; National Institutes of Health Grant GM59290; National Science Foundation Grants BCS-0218338 and EPS-0346411; the State of Louisiana Board of Regents Support Fund (M.A.B.); and by funds from the University of Texas at Arlington (C.F.).

1. Long, M., Betran, E., Thornton, K. & Wang, W. (2003) *Nat. Rev. Genet.* **4**, 865–875.
2. Johnson, M. E., Viggiano, L., Bailey, J. A., Abdul-Rauf, M., Goodwin, G., Rocchi, M. & Eichler, E. E. (2001) *Nature* **413**, 514–519.
3. Marques, A. C., Dupanloup, I., Vinckenbosch, N., Reymond, A. & Kaessmann, H. (2005) *PLoS Biol.* **3**, e357.
4. Miller, W. J., Hagemann, S., Reiter, E. & Pinsky, W. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4018–4022.
5. Kidwell, M. G. & Lisch, D. R. (2001) *Evolution Int. J. Org. Evolution* **55**, 1–24.
6. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001) *Nature* **409**, 860–921.
7. Britten, R. J. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 16825–16830.
8. Zdobnov, E. M., Campillos, M., Harrington, E. D., Torrents, D. & Bork, P. (2005) *Nucleic Acids Res.* **33**, 946–954.
9. Bundock, P. & Hooykaas, P. (2005) *Nature* **436**, 282–284.
10. Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M. (2002) *Mobile DNA II* (Am. Soc. Microbiol., Washington, DC).
11. Doolittle, W. F. & Sapienza, C. (1980) *Nature* **284**, 601–603.
12. Deininger, P. L., Moran, J. V., Batzer, M. A. & Kazazian, H. H., Jr. (2003) *Curr. Opin. Genet. Dev.* **13**, 651–658.
13. Cordaux, R. & Batzer, M. A. (2006) *Proc. Natl. Acad. Sci. USA* **103**, 1157–1158.
14. Robertson, H. M. & Zuppano, K. L. (1997) *Gene* **205**, 203–217.
15. Lee, S.-H., Oshige, M., Durant, S. T., Rasila, K. K., Williamson, E. A., Ramsey, H., Kwan, L., Nickoloff, J. A. & Hromas, R. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 18075–18080.
16. Kouzarides, T. (2002) *Curr. Opin. Genet. Dev.* **12**, 198–209.
17. Cheng, X., Collins, R. E. & Zhang, X. (2005) *Annu. Rev. Biophys. Biomol. Struct.* **34**, 267–294.
18. Goodman, M., Porter, C. A., Czelusniak, J., Page, S. L., Schneider, H., Shoshani, J., Gunnell, G. & Groves, C. P. (1998) *Mol. Phylogenet. Evol.* **9**, 585–598.
19. Lim, L. P. & Burge, C. B. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 11193–11198.
20. Brunak, S., Engelbrecht, J. & Knudsen, S. (1991) *J. Mol. Biol.* **220**, 49–65.
21. Yang, Z. (1997) *Comput. Appl. Biosci.* **13**, 555–556.
22. Robertson, H. M. & Martos, R. (1997) *Gene* **205**, 219–228.
23. Lipkow, K., Buisine, N., Lampe, D. J. & Chalmers, R. (2004) *Mol. Cell. Biol.* **24**, 8301–8311.
24. Zhang, L., Dawson, A. & Finnegan, D. J. (2001) *Nucleic Acids Res.* **29**, 3566–3575.
25. Auge-Gouillou, C., Hamelin, M. H., Demattei, M. V., Periquet, M. & Bigot, Y. (2001) *Mol. Genet. Genomics* **265**, 51–57.
26. Auge-Gouillou, C., Brillet, B., Germon, S., Hamelin, M. H. & Bigot, Y. (2005) *J. Mol. Biol.* **351**, 117–130.
27. Lohe, A. R., De Aguiar, D. & Hartl, D. L. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 1293–1297.
28. Bender, J. & Kleckner, N. (1992) *EMBO J.* **11**, 741–750.
29. Lampe, D. J., Churchill, M. E. & Robertson, H. M. (1996) *EMBO J.* **15**, 5470–5479.
30. Feschotte, C., Osterlund, M. T., Peeler, R. & Wessler, S. R. (2005) *Nucleic Acids Res.* **33**, 2153–2165.
31. Bigot, Y., Brillet, B. & Auge-Gouillou, C. (2005) *J. Mol. Biol.* **351**, 108–116.
32. Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W. J., Anderson, S., Yates, J., Washburn, M. P., et al. (2005) *Cell* **123**, 581–592.
33. Keogh, M. C., Kurdستاني, S. K., Morris, S. A., Ahn, S. H., Podolny, V., Collins, S. R., Schuldiner, M., Chin, K., Punna, T., Thompson, N. J., et al. (2005) *Cell* **123**, 593–605.
34. Cordaux, R., Lee, J., Dinoso, L. & Batzer, M. A. (March 6, 2006) *Gene*, 10.1016/j.gene.2006.01.020.
35. Hall, T. A. (1999) *Nucleic Acids Symp. Ser.* **41**, 95–98.