The following resources related to this article are available online at
www.sciencemag.org (this information is current as of April 13, 2007):

Updated information and services, including high-resolution figures, can be found in the online
version of this article at:
http://www.sciencemag.org/cgi/content/full/316/5822/238

Supporting Online Material can be found at:
http://www.sciencemag.org/cgi/content/full/316/5822/238/DC1

A list of selected additional articles on the Science Web sites **related to this article** can be
found at:
http://www.sciencemag.org/cgi/content/full/316/5822/238#related-content

This article **cites 27 articles**, 9 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/316/5822/238#otherarticles

This article has been **cited by** 1 articles hosted by HighWire Press; see:
http://www.sciencemag.org/cgi/content/full/316/5822/238#otherarticles

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce**
this article in whole or in part can be found at:
http://www.sciencemag.org/about/permissions.dtl

REPORT

Mobile DNA in Old World Monkeys: A Glimpse Through the Rhesus Macaque Genome

Kyudong Han,[1]* Miriam K. Konkel,[1]* Jinchuan Xing,[1]*† Hui Wang,[1]* Jungnam Lee,[1] Thomas J. Meyer,[1] Charles T. Huang,[1] Erin Sandifer,[1] Kristi Hebert,[1] Erin W. Barnes,[1] Robert Hubley,[2] Webb Miller,[3] Arian F. A. Smit,[2] Brygg Ullmer,[4] Mark A. Batzer[1]‡

The completion of the draft sequence of the rhesus macaque genome allowed us to study the genomic composition and evolution of transposable elements in this representative of the Old World monkey lineage, a group of diverse primates closely related to humans. The L1 family of long interspersed elements appears to have evolved as a single lineage, and *Alu* elements have evolved into four currently active lineages. We also found evidence of elevated horizontal transmissions of retroviruses and the absence of DNA transposon activity in the Old World monkey lineage. In addition, ~100 precursors of composite SVA (short interspersed element, variable number of tandem repeat, and *Alu*) elements were identified, with the majority being shared by the common ancestor of humans and rhesus macaques. Mobile elements compose roughly 50% of primate genomes, and our findings illustrate their diversity and strong influence on genome evolution between closely related species.

Old World monkeys (OWMs) represent one of the most closely related primate groups to humans. The rhesus macaques (*Macaca mulatta*), along with other OWMs, have been extensively used in biomedical studies (*1*). An improved understanding of their genomic architecture could hold important implications for medicine, evolutionary understanding, and beyond. Similar to the human and chimpanzee genomes, roughly 50% of the rhesus macaque genome consists of various repetitive sequences (*2–4*). The majority of these repeats are mobile elements, which can be divided into class I DNA transposons (*5*) and class II retrotransposons (*6*). Related transposable elements are further categorized into families, with each family further classified into subfamilies on the basis of their sequence relationships. The insertion of mobile elements can alter gene expression (*7*), generate genomic deletions (*8*), and even create new genes and gene families (*9*). Existing repetitive elements can also mediate recombinations between similar elements at different genomic locations (ectopic recombination) (*10*). In addition, the GC-rich nature of certain mobile elements {e.g., *Alu* and SVA [short interspersed element (SINE), variable number of tandem repeat (VNTR), and *Alu*] elements} can introduce new GC islands through their insertion (*3*). Despite the overall similarity in retrotransposon mobilization activity in the OWM and hominoid (human and ape) lineages, mobile elements have continued to evolve independently in both lineages. Close examination of the overall mobile-element composition in OWMs, with the rhesus macaque genome used as a reference, allows an understanding of their lineage-specific expansion and illustrates their overall contribution to genome evolution.

Without any detected lineage-specific copies, DNA transposons, which mobilize through a cut-and-paste mechanism, appear to have been inactive in the rhesus macaque lineage since their speciation from humans. The paucity of DNA transposon mobilization in mammals, and in amniotes in general, is noteworthy by comparison with other organisms (e.g., plants) and may result from the relative difficulty in horizontal transfer into animals' germ lines (*11*).

Similar to the human genome, the rhesus macaque genome contains over half a million recognizable copies of endogenous retroviruses (ERVs) and their nonautonomous derivatives, with the great majority being present or fixed before the hominoid-OWM split (*12*). We found evidence for at least eight instances of horizontal transmission of ERVs in the OWM lineage resulting in 2750 extant copies (table S1 and SOM Text). This is much higher than in the human lineage, where there is evidence for only one or two invading elements leaving fewer than 10 extant copies (*13*). Five of the eight horizontally transmitted ERVs belong to class I retroviruses, and the remaining three belong to class II retroviruses (shown in red letters in Fig. 1). Apart from these new invasions, at least seven ERV families already entered the genome before the hominoid-OWM split and remained active afterward. There are over 3500 copies of these ERV subfamilies in the OWM lineage, similar to the number of lineage-specific ERV copies in humans.

The L1PA (primate A) family of long interspersed elements (LINES) represents the dominant active L1 lineage throughout primate evolution. In our analysis, L1PA5 was the most commonly recovered L1 subfamily, and ~19,000 L1PA5 elements specific to the OWM lineage were identified in the rhesus macaque genome. Most of these elements represent insertion events that occurred along the OWM lineage leading to rhesus macaques and are therefore present in multiple OWM species (fig. S2). A total of 32 OWM-specific L1 subfamilies were identified with the use of diagnostic substitutions present in these elements (table S2). To investigate the relationship of L1s, we constructed a median-joining network with their consensus sequences (Fig. 2 and SOM Text) and estimated the age of each subfamily (table S2). The network results indicated that the OWM-specific L1 lineage rooted with the L1PA6 consensus sequence, and several lineages roughly followed a sequential order, with little overlap in their amplification period. The sequential evolution of L1 elements appears to follow a general trend seen in mammalian L1s (*14*) and may result from amplification competition between two distinct L1 lineages (*15*). Altogether, we identified nine putative retrotransposition-competent L1s in the rhesus macaque genome, and they belonged to the L1CER-3 or L1CER-4 subfamilies; each L1 subfamily name is identified by "CER" (which stands for Cercopithecidae, indicating the origin of the consensus sequence) and an Arabic numeral indicating its lineage (*12*). Nine was a considerably lower number of potentially active L1 elements than that in the human genome, which has 80 to 100 active copies (*16*). Nevertheless, it is likely that additional retrotransposition-competent L1 elements will be recovered in more refined drafts of the rhesus macaque genome.
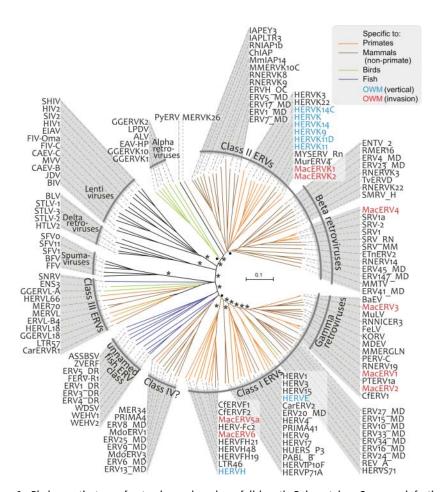
Retrotransposon-mediated DNA sequence transduction is a process whereby a retrotransposon carries a flanking genomic sequence during its mobilization that can result in exon or gene duplication (*17*). Three L1 elements with 5′ transduced exon–derived sequences were identified in the rhesus macaque genome. Moreover, detailed analysis indicated that one of the three insertions occurred in an exon of another gene (table S3 and SOM Text). These three events empirically demonstrate that exon-derived sequences can be transferred via 5′ L1–mediated transduction within primate genomes and that 5′ transduction constitutes a second mechanism of retrotransposon-mediated "exon shuffling."

[1]Department of Biological Sciences, Biological Computation and Visualization Center, Center for Bio-Modular Multi-Scale Systems, Louisiana State University, Baton Rouge, LA 70803, USA. [2]Institute for Systems Biology, Seattle, WA 98103, USA. [3]Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, University Park, PA 16802, USA. [4]Department of Computer Science, Center for Computation and Technology (CCT), Louisiana State University, Baton Rouge, LA 70803, USA.

*These authors contributed equally to this work.
†Present address: Department of Human Genetics, University of Utah Health Sciences Center, Salt Lake City, UT 84112, USA
‡To whom correspondence should be addressed. E-mail: mbatzer@lsu.edu

Fig. 1. Phylogenetic tree of retroviruses based on full-length Pol proteins. Common infectious retroviruses and endogenous retroviruses, present in fish, birds, mammals (nonprimate), and primates, were included in the analysis. Color identifications for each group are shown in the upper right corner. Asterisks and circles show deep-rooted branches with >95 and >75% bootstrap values, respectively. The ERVs identified in this study that invaded the OWM genome horizontally (i.e., through external germline infection) are indicated with red letters. For all ERVs shown in blue letters, the original insertion occurred in the common ancestor of humans and rhesus macaques (i.e., vertically) and is present in both genomes. All ERVs indicated with blue letters also generated new insertions in the OWM lineage. The scale bar indicates 10% divergence in the amino acid sequence.
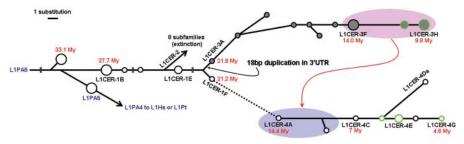

Fig. 2. Median-joining network of OWM-specific L1 subfamilies. Subfamilies are represented by circles, with the circle size symbolizing the relative size of each subfamily. The length of the lines corresponds to the number of substitutions. The scale of a single substitution is shown in the upper left corner. Broken lines indicate segments not drawn to scale. Gray circles represent the subfamilies belonging to the L1CER-3 lineage, which include an 18–base pair (bp) duplication in their 3′ untranslated region (3′UTR), and green-edged circles contain intact full-length L1 elements. The dashed line and red arrow represent two alternative pathways for the origin of the L1CER-4 subfamily. The subfamilies in the blue and pink ovals share the same diagnostic mutations but do not share the 18-bp duplication. My, million years.
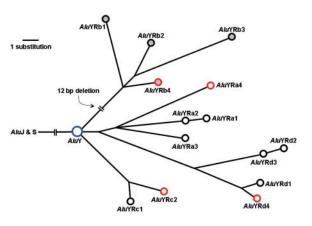
Alu elements are the most successful SINEs in primate genomes (*18*), and ~110,000 *Alu* insertions are specific to OWMs. Fourteen different OWM lineage–specific *Alu*Y subfamilies fell into four lineages, shown in a median-joining network analysis (Fig. 3), and were identified with estimated copy numbers (table S4). All subfamilies were estimated to have originated after the hominoid-OWM divergence and were congruent with our phylogenetic analyses showing that all of these *Alu* subfamilies were restricted to OWMs (SOM Text). The simultaneous retrotransposition activity of multiple *Alu* subfamilies is similar to that in the human genome, and the activity of multiple "source genes" may have contributed to the amplification success of *Alu* elements despite their reliance on L1 enzymatic machinery for mobilization (*19*).

About 100 precursors of SVA were identified in the rhesus macaque genome. The variable number of tandem repeat (VNTR) regions of these elements share >90% identity with the VNTR unit in hominoid SVA elements (*20*), although they have no sequence homology with other components of SVA elements. Thus, these elements appear to have contributed a portion of the genetic material required to form the SVA composite retrotransposon family in hominoids. The majority of these elements are shared between human and rhesus macaque, indicating that these elements were active before the divergence of hominoids and OWMs. The low number of lineage-specific elements (~20 in the OWM lineage) suggests a very low retrotransposition rate of SVA precursor elements over the past 25 million years.

Composing nearly half of all sequenced primate genomes, mobile elements—especially retrotransposons—are major components of genomic variation and a driving force of primate evolution. Although the overall number of mobile elements is similar in the human, chimpanzee, and rhesus macaque genomes (*2–4*), a large fraction of the elements inserted independently into different locations within each genome and thus shaped the genomes differently (*21*). Whereas most retrotransposon insertions remain neutral in the genome, many insertions can have deleterious effects of varying severity. Mobile elements can cause genetic diseases not only by direct gene disruption or by the deletion of exonic sequence upon insertion but also by mediating subsequent recombination between existing retrotransposons. Indeed, more than 118 human genetic disorders are caused by retrotransposons, including hemophilia B, breast cancers, and congenital muscular dystrophy [see (*22*) and (*23*) for reviews]; they are likely to have a similar impact on the rhesus genome. Yet, retrotransposons are also responsible for creating a variety of genomic novelties. They are involved in mediating gene duplication, exon shuffling, and RNA-editing–mediated exonization (*9, 17, 24*). All these mechanisms can contribute to new gene formation, as well as potentially altering DNA methylation patterns and contributing to X chro-

The Rhesus Macaque Genome

Fig. 3. Median-joining network of OWM-specific *Alu* subfamilies. Subfamilies are represented by circles. The length of the lines corresponds to the number of substitutions, and the scale of a single substitution is shown in the upper left corner. Broken lines indicate segments not drawn to scale. Gray circles represent all subfamilies belonging to the *Alu*YRb lineage containing a 12-bp deletion. Red-edged circles denote the youngest *Alu* subfamily within each lineage, and the blue-edged circle indicates the *Alu*Y subfamily consensus sequence.

mosome inactivation in females (*25, 26*). In addition, retrotransposons provide highly valuable genetic systems for primate population and phylogenetic studies, because they have a known ancestral (i.e., insertion-absent) state, and the chance that the same type of element would integrate at precisely the same location in multiple individuals is essentially zero (i.e., the insertions are identical by descent) (*27, 28*). Altogether, understanding the mobile-element landscape in primates is not only important for biologists but also crucial for biomedical researchers using primate animal models.

References and Notes

1. H. E. Carlsson, S. J. Schapiro, I. Farah, J. Hau, *Am. J. Primatol.* **63**, 225 (2004).
2. Chimpanzee Sequencing and Analysis Consortium, *Nature* **437**, 69 (2005).
3. E. S. Lander *et al.*, *Nature* **409**, 860 (2001).
4. Rhesus Macaque Genome Sequencing and Analysis Consortium, *Science* **316**, 222 (2007).
5. A. F. Smit, *Curr. Opin. Genet. Dev.* **6**, 743 (1996).
6. P. L. Deininger, M. A. Batzer, *Genome Res.* **12**, 1455 (2002).
7. P. L. Deininger, J. V. Moran, M. A. Batzer, H. H. Kazazian Jr., *Curr. Opin. Genet. Dev.* **13**, 651 (2003).
8. K. Han *et al.*, *Nucleic Acids Res.* **33**, 4040 (2005).
9. J. Xing *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17608 (2006).
10. S. K. Sen *et al.*, *Am. J. Hum. Genet.* **79**, 41 (2006).
11. A. F. Smit, *Curr. Opin. Genet. Dev.* **9**, 657 (1999).
12. Materials and methods are available as supporting material on *Science* Online.
13. L. Benit, A. Calteau, T. Heidmann, *Virology* **312**, 159 (2003).
14. A. V. Furano, D. D. Duvernell, S. Boissinot, *Trends Genet.* **20**, 9 (2004).
15. H. Khan, A. Smit, S. Boissinot, *Genome Res.* **16**, 78 (2006).
16. B. Brouha *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5280 (2003).
17. J. V. Moran, R. J. DeBerardinis, H. H. Kazazian Jr., *Science* **283**, 1530 (1999).
18. M. A. Batzer, P. L. Deininger, *Nat. Rev. Genet.* **3**, 370 (2002).
19. R. Cordaux, D. J. Hedges, M. A. Batzer, *Trends Genet.* **20**, 464 (2004).
20. H. Wang *et al.*, *J. Mol. Biol.* **354**, 994 (2005).
21. R. E. Mills *et al.*, *Am. J. Hum. Genet.* **78**, 671 (2006).
22. P. A. Callinan, M. A. Batzer, in *Genome Dynamics*, J. N. Volff, Ed. (Karger, Basel, Switzerland, 2006), vol. 1, pp. 104–115.
23. J. M. Chen, C. Ferec, D. N. Cooper, *J. Biomed. Biotechnol.* **2006**, 56182 (2006).
24. G. Lev-Maor *et al.*, *Genome Biol.* **8**, R29 (2007).
25. J. A. Bailey, L. Carrel, A. Chakravarti, E. E. Eichler, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6634 (2000).
26. R. S. Hansen, *Hum. Mol. Genet.* **12**, 2559 (2003).
27. A. M. Shedlock, K. Takahashi, N. Okada, *Trends Ecol. Evol.* **19**, 545 (2004).
28. D. A. Ray, J. Xing, A. H. Salem, M. A. Batzer, *Syst. Biol.* **55**, 928 (2006).
29. Thanks to the RMGSAC for the rhesus macaque genome sequence and to S. Brandt, W. Scullin, and S. White for computational support. This project was facilitated in part by high-performance computing allocations from Louisiana State University CCT and supported by the NSF grants BCS-0218338 (M.A.B.) and EPS-0346411 (M.A.B.), NIH GM59290 (M.A.B.), and the State of Louisiana Board of Regents Support Fund (M.A.B.).

REPORT

Demographic Histories and Patterns of Linkage Disequilibrium in Chinese and Indian Rhesus Macaques

Ryan D. Hernandez,[1] Melissa J. Hubisz,[2] David A. Wheeler,[3] David G. Smith,[4,5] Betsy Ferguson,[6,7] Jeffrey Rogers,[8] Lynne Nazareth,[3] Amit Indap,[1] Traci Bourquin,[3] John McPherson,[3] Donna Muzny,[3] Richard Gibbs,[3] Rasmus Nielsen,[9] Carlos D. Bustamante[1]*

To understand the demographic history of rhesus macaques (*Macaca mulatta*) and document the extent of linkage disequilibrium (LD) in the genome, we partially resequenced five Encyclopedia of DNA Elements regions in 9 Chinese and 38 captive-born Indian rhesus macaques. Population genetic analyses of the 1467 single-nucleotide polymorphisms discovered suggest that the two populations separated about 162,000 years ago, with the Chinese population tripling in size since then and the Indian population eventually shrinking by a factor of four. Using coalescent simulations, we confirmed that these inferred demographic events explain a much faster decay of LD in Chinese ($r^2 \approx 0.15$ at 10 kilobases) versus Indian ($r^2 \approx 0.52$ at 10 kilobases) macaque populations.

Rhesus macaques (*Macaca mulatta*) and humans shared a most recent common ancestor (MRCA) ~25 million years ago (Ma), and our genomes differ at <7% of nucleotide bases (*1*). Rhesus and humans, therefore, share a large number of fundamental biological characteristics, including many underlying genetic and physiological processes that lead to disease. For this reason, rhesus macaques have become a model organism for vaccine research (*2, 3*), as well as

studies of normal human physiology and disease. Although previous studies of genetic variation in rhesus have described >300 microsatellite polymorphisms (*4, 5*), identifying specific genetic risk factors for disease requires a much greater resolution of genetic variation across the genome.

The current geographic range of rhesus macaques is larger than any other nonhuman primate, stretching from western India and Pakistan to the eastern shores of China (Fig. 1). Fossil records suggest that the genus *Macaca* originated

[1]Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14850, USA. [2]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA. [3]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA. [4]Department of Anthropology, Davis, CA, USA. [5]California National Primate Research Center, Davis, CA, USA. [6]Genetics Research and Informatics Program, Oregon National Primate Research Center, Oregon Health and Sciences University, Beaverton, OR 97006, USA. [7]Washington National Primate Research Center, University of Washington, Seattle, WA 98195, USA. [8]Department of Genetics, Southwest Foundation for Biomedical Research, and Southwest National Primate Research Center, San Antonio, TX 78227, USA. [9]Center for Comparative Genomics, Department of Biology, University of Copenhagen, Universitetsparken 15, 2100 Kbh Ø, Denmark.

*To whom correspondence should be addressed. E-mail: cdb28@cornell.edu