

The following resources related to this article are available online at www.sciencemag.org (this information is current as of December 2, 2009):

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/316/5822/240>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/316/5822/240/DC1>

This article **cites 17 articles**, 4 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/316/5822/240#otherarticles>

This article has been **cited by** 15 article(s) on the ISI Web of Science.

This article has been **cited by** 7 articles hosted by HighWire Press; see:

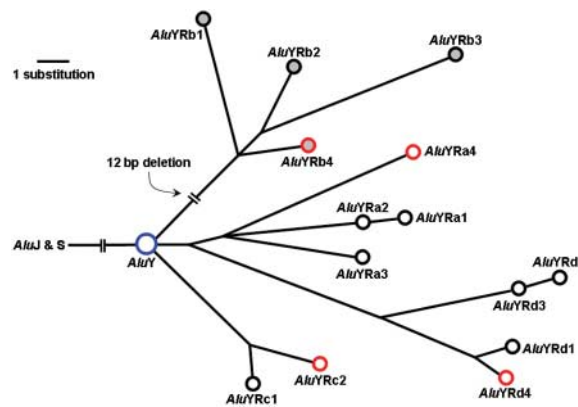
<http://www.sciencemag.org/cgi/content/full/316/5822/240#otherarticles>

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>

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 *AluYRb* lineage containing a 12-bp deletion. Red-edged circles denote the youngest *Alu* subfamily within each lineage, and the blue-edged circle indicates the *AluY* 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

- H. E. Carlsson, S. J. Schapiro, I. Farah, J. Hau, *Am. J. Primatol.* **63**, 225 (2004).
- Chimpanzee Sequencing and Analysis Consortium, *Nature* **437**, 69 (2005).

- E. S. Lander *et al.*, *Nature* **409**, 860 (2001).
- Rhesus Macaque Genome Sequencing and Analysis Consortium, *Science* **316**, 222 (2007).
- A. F. Smit, *Curr. Opin. Genet. Dev.* **6**, 743 (1996).
- P. L. Deininger, M. A. Batzer, *Genome Res.* **12**, 1455 (2002).
- P. L. Deininger, J. V. Moran, M. A. Batzer, H. H. Kazazian Jr., *Curr. Opin. Genet. Dev.* **13**, 651 (2003).
- K. Han *et al.*, *Nucleic Acids Res.* **33**, 4040 (2005).
- J. Xing *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17608 (2006).
- S. K. Sen *et al.*, *Am. J. Hum. Genet.* **79**, 41 (2006).
- A. F. Smit, *Curr. Opin. Genet. Dev.* **9**, 657 (1999).
- Materials and methods are available as supporting material on Science Online.
- L. Benit, A. Calteau, T. Heidmann, *Virology* **312**, 159 (2003).
- A. V. Furano, D. D. Duvernell, S. Boissinot, *Trends Genet.* **20**, 9 (2004).
- H. Khan, A. Smit, S. Boissinot, *Genome Res.* **16**, 78 (2006).
- B. Brouha *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5280 (2003).

- J. V. Moran, R. J. DeBerardinis, H. H. Kazazian Jr., *Science* **283**, 1530 (1999).
- M. A. Batzer, P. L. Deininger, *Nat. Rev. Genet.* **3**, 370 (2002).
- R. Cordaux, D. J. Hedges, M. A. Batzer, *Trends Genet.* **20**, 464 (2004).
- H. Wang *et al.*, *J. Mol. Biol.* **354**, 994 (2005).
- R. E. Mills *et al.*, *Am. J. Hum. Genet.* **78**, 671 (2006).
- P. A. Callinan, M. A. Batzer, in *Genome Dynamics*, J. N. Volf, Ed. (Karger, Basel, Switzerland, 2006), vol. 1, pp. 104–115.
- J. M. Chen, C. Ferec, D. N. Cooper, *J. Biomed. Biotechnol.* **2006**, 56182 (2006).
- G. Lev-Maor *et al.*, *Genome Biol.* **8**, R29 (2007).
- J. A. Bailey, L. Carrel, A. Chakravarti, E. E. Eichler, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6634 (2000).
- R. S. Hansen, *Hum. Mol. Genet.* **12**, 2559 (2003).
- A. M. Shedlock, K. Takahashi, N. Okada, *Trends Ecol. Evol.* **19**, 545 (2004).
- D. A. Ray, J. Xing, A. H. Salem, M. A. Batzer, *Syst. Biol.* **55**, 928 (2006).
- 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.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/316/5822/238/DC1
Materials and Methods
SOM Text
Figs. S1 and S2
Tables S1 to S7
References
Source Code

3 January 2007; accepted 16 March 2007
10.1126/science.1139462

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

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

in northern Africa approximately 5.5 Ma, followed by migration through the Middle East and into northern India by ~3 Ma (6). By ~2 Ma, macaques had traversed most of China and reached the Indonesian archipelago, where the putative ancestral species of rhesus macaque, *M. fascicularis*, is thought to have originated (6, 7).

Previous studies of mitochondrial DNA (8), major histocompatibility complex (MHC) alleles (9), and single-nucleotide polymorphisms (SNPs)

in gene-linked regions (10) suggest moderate levels of genetic differentiation between captive-born Indian and Chinese rhesus populations. Developing a more thorough understanding of genetic variation within and between these two populations has important implications for biomedical research. For example, when infected with the simian immunodeficiency virus, animals from Chinese populations develop AIDS-like symptoms more slowly than animals from Indian populations (3).

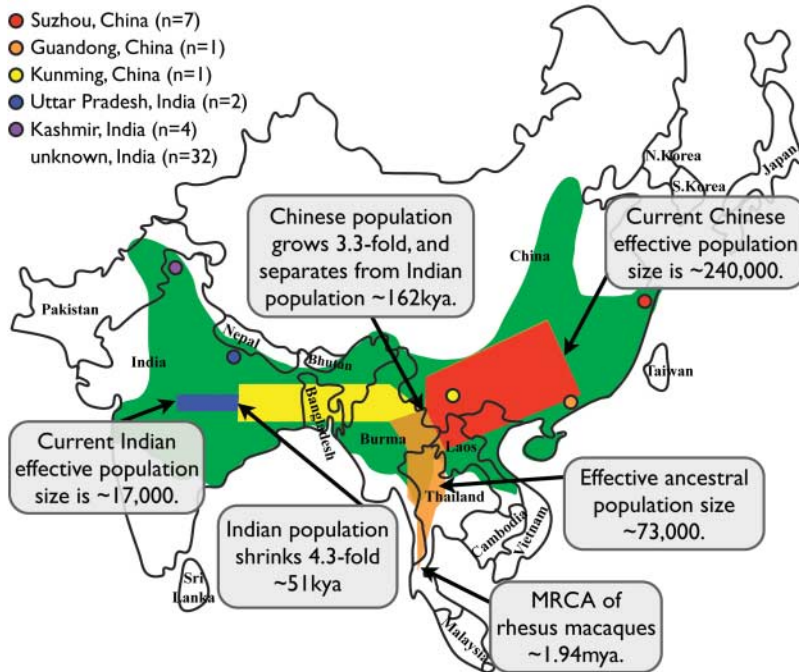


Fig. 1. The current geographic range of rhesus macaques [green, redrawn from (20)] with the inferred demographic history and the sample locations superimposed. The geographic location of the MRCA is based on (4).

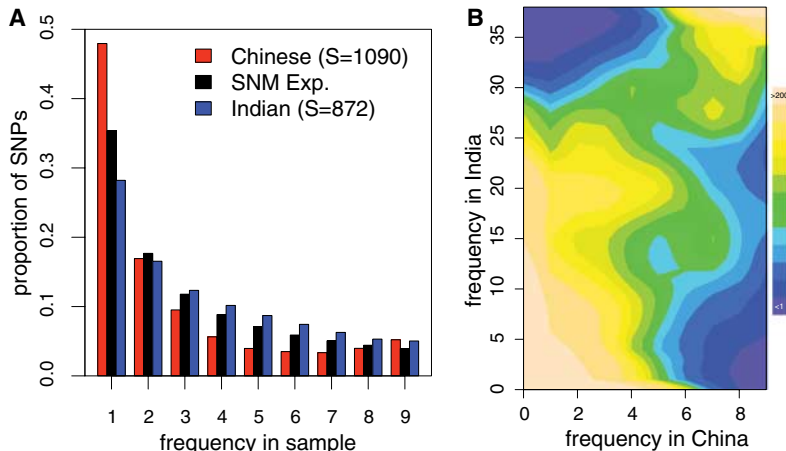


Fig. 2. (A) The marginal frequency spectrum of derived mutations for each population (shown as expected proportions in a subsample of 10 chromosomes by integrating over possible configurations of observed and missing data, with the total number of SNPs in parentheses) and the expected distribution under the standard neutral model (SNM) of constant size. (B) A "topographical map" of the joint site-frequency spectrum for the two populations, with darker tones representing frequency pairs with few SNPs, and lighter tones representing frequency pairs with many SNPs.

We have identified 1476 SNPs by sequencing >150 kb of DNA across five Encyclopedia of DNA Elements (ENCODE) (11–13) regions located on separate autosomal chromosomes in nine captive-born from wild-caught Chinese and 38 captive-born Indian rhesus macaques. The Chinese animals derive from three distinct geographical sites, whereas the Indian animals came from three different colonies in the United States (Fig. 1). Individuals were chosen to represent rhesus macaque populations that are currently being studied by the international community and to minimize relatedness in the sample [with most individuals in the study being unrelated back to the founding of the colony into which they were born, and none having a shared grandparent (13)]. In our sample of 1476 SNPs discovered, only 486 (33%) were shared across both populations, whereas 604 were found only in the Chinese population (61% of 1090 SNPs observed) and 386 were found only in the Indian population (39% of 872 SNPs observed). The frequency distribution of derived mutations across SNPs [using DNA sequence from the ENCODE project for baboon, *Papio cynocephalus anubis*, to infer the putative ancestral state (13)] shows that the Chinese population harbors an excess of rare SNPs relative to a population of constant size, whereas the Indian population has too few rare and too many intermediate- and high-frequency-derived SNPs (Fig. 2A). The observed disparity in SNP density (7.25 SNPs per kb for Chinese versus 5.8 SNPs per kb for Indian) in the two populations suggests that the effective size of the Chinese population is much larger than the Indian population, given that the Indian sample size is four times as large as that of the Chinese.

We observed a moderate level of population structure between the Indian and Chinese samples, as measured by Wright's F_{ST} statistic (average $F_{ST} = 0.14$; SD = 0.11; range = -0.024 to 0.645) (Fig. 3A). Furthermore, the Bayesian clustering program STRUCTURE (14) clearly separates Chinese and Indian individuals when assuming two clusters (Fig. 3B), and considering more clusters does not significantly improve the fit of the model. We found only one Chinese individual with a marginal amount of Indian ancestry (8.5%, sampled from Suzhou) and eight Indian individuals with more than 5% Chinese ancestry [max 16.8%, including animals from all three primate centers (13)]. These low levels of admixture suggest that recurrent migration between the populations has been minimal. Moreover, the two populations were clearly distinguished by principal components analysis (15) along the first two axes of variation (Fig. 3C). Interestingly, the second component also separates one Chinese individual (sampled from Suzhou) from the others, which suggests that further population substructure may exist. Although this individual is not differentiated from other Chinese-origin animals in the STRUCTURE analysis, it may, nonetheless, harbor alleles from an unsampled

The Rhesus Macaque Genome

Fig. 3. (A) The distribution of F_{ST} between Indian and Chinese rhesus, calculated with the average pairwise-difference across each nonoverlapping window (13). (B) STRUCTURE results. Individuals are represented by vertical lines, and sorted by their amount of Chinese ancestry (black vertical line separates animals with Indian and Chinese origins). Colors correspond to the proportion of an individual's ancestry attributable to a given population (blue, Indian; red, Chinese). (C) Principal component 1 (PC1) and PC2 separate Indian from Chinese individuals. PC2 also isolates a single Chinese individual [corresponding to an individual sampled from Suzhou and shown as the fourth individual from the right in (B)].

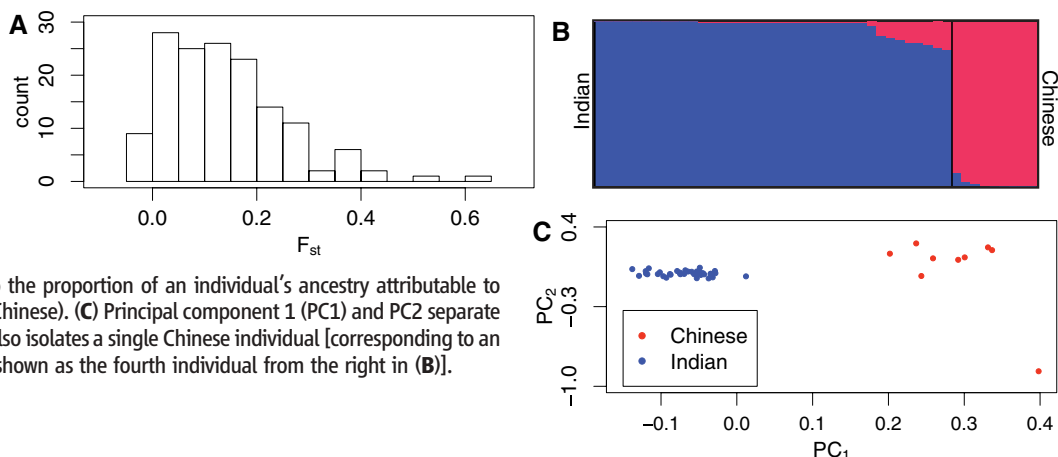
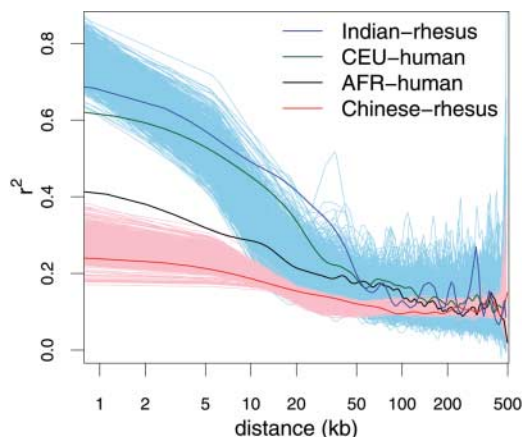


Fig. 4. The decay of LD for Indian and Chinese rhesus macaques versus European and African humans ($n = 9$ for all samples), along with the decay of LD for 1000 neutral simulations of our inferred demographic history for rhesus macaque. Human data are from three ENCODE regions orthologous to the rhesus data (13, 21).



Chinese subpopulation (i.e., the two wild-caught parents may be from different subpopulations).

Using maximum likelihood under the assumption that the animals in this study form a random sample from their respective population (13), we fit a two-population demographic model to the joint distribution of SNP frequencies, or site-frequency spectrum, shown in Fig. 2B. Our model suggests that the Chinese population expanded by a factor of 3.3 and separated from the Indian population ~162 thousand years ago (ka) (95% confidence interval, CI = 183 to 132 ka). After separating, the Indian population maintained its ancestral population size until ~51 ka CI = 72 to 21 ka], when it was reduced by a factor of 4.3. The population genetic model, although a very simplistic approximation to the rich and complex history of the species, fits the data well, as indicated by a goodness-of-fit test ($P = 0.133$). Coalescent simulations (13) on the basis of the inferred demographic history for Indian and Chinese rhesus macaques suggest that the MRCA of the two populations lived ~1.94 Ma (SE 14 Ky). This estimate places the MRCA of rhesus near the divergence time from *M. fascicularis*, inferred from mitochondrial DNA to be 1.83 to 5 Ma (16, 17). Moreover, our simulations suggest that the effective size of the ancestral population of rhesus

macaques was ~73,070 (SE 231) individuals, implying that the current effective size of the Chinese population is ~239,704, whereas the Indian population is estimated to be ~17,014.

The recent demographic events that caused these differences in effective population sizes of Indian and Chinese rhesus macaques have also had a large impact on linkage disequilibrium (LD). To quantify the extent of LD in Indian and Chinese rhesus macaques, we measured the correlation coefficient (r^2) of alleles from frequency-matched SNPs (13, 18). Figure 4 shows substantial differences between the Indian and Chinese rhesus macaque populations, which are more extreme than the patterns observed among humans. For example, within the Indian rhesus population, LD extends much further than LD observed for European humans, whereas the Chinese rhesus population shows little LD, even for SNPs that are physically very close. Coalescent simulations (13) show that the observed patterns of LD are consistent with our inferred demographic history of this species (shown in Fig. 4 as light blue and pink curves for Indian and Chinese rhesus, respectively). However, LD in the Indian population extends slightly further than expected. This observation may be consistent with recent admixture with a Burmese rhesus population not sam-

pled in this study (8), because admixture between populations with allele frequency differences is known to generate long-range LD.

In this study, we analyzed noncoding data in rhesus macaques to characterize their underlying demographic history and to quantify the extent of LD relative to humans. The genetic differences that we have observed between Indian and Chinese rhesus macaques are consistent with a recent report on the distribution of SNPs in these populations (10), as well as previous studies of protein coding, microsatellite STR (short tandem repeat), MHC loci, and mitochondrial and Y-chromosome DNA haplotypes (8). Without samples from wild-caught Indian rhesus monkeys, however, these data must be regarded as estimates, because they may reflect a sampling bias toward those macaques that are available for study in the United States as a result of international restrictions on exportation of primates.

Extending these studies to whole-genome association mapping in captive-bom animals could be fruitful for identifying genes involved in human diseases. On the basis of the patterns of LD that we have observed, such an association study would likely require many fewer markers to identify common disease-causing variants in rhesus macaques than in humans. Because LD in captive Indian rhesus macaque populations extends much further than in humans, a SNP map with roughly 1 SNP every 35 kb (82,000 SNPs total) would suffice to achieve the same threshold ($r^2 = 0.4$) as a marker every 6 kb in humans (13, 19). Furthermore, because LD decays much faster in Chinese rhesus monkeys than in humans, Chinese macaques provide an ideal platform for localizing mutations that are difficult to map in either Indian macaques or humans as a result of extensive LD among candidate mutations in a particular region.

References and Notes

1. Rhesus Macaque Genome Sequencing and Analysis Consortium, *Science* **316**, 222 (2007).
2. R. A. Weiss, *Nature* **410**, 1035 (2001).
3. B. Ling et al., *AIDS* **16**, 1489 (2002).
4. J. Rogers et al., *Genomics* **87**, 30 (2006).

5. M. Raveendran *et al.*, *Genomics* **88**, 706 (2006).
6. E. Delson, in *The Macaques: Studies in Ecology, Behavior, and Evolution*, D. D. Lindburg, Ed. (van Nostrand Reinhold, New York, 1980), pp. 10–30.
7. C. Abegg, B. Thierry, *Biol. J. Linn. Soc.* **75**, 555 (2002).
8. D. G. Smith, J. McDonough, *Am. J. Primatol.* **65**, 1 (2005).
9. J. Viray, B. Rolfs, D. G. Smith, *Comp. Med.* **51**, 555 (2001).
10. B. Ferguson *et al.*, *BMC Genom.* **8**, 43 (2007).
11. ENCODE Project Consortium, *Science* **306**, 636 (2004).
12. ENCODE regions were chosen because they have been widely studied across several mammals, including rhesus and baboon.
13. Materials and methods are available as supporting material on Science Online.
14. D. Falush, M. Stephens, J. K. Pritchard, *Genetics* **164**, 1567 (2003).
15. A. L. Price *et al.*, *Nat. Genet.* **38**, 904 (2006).
16. K. Hayasaka, K. Fujii, S. Horai, *Mol. Biol. Evol.* **13**, 1044 (1996).
17. J. C. Morales, D. J. Melnick, *J. Hum. Evol.* **34**, 1 (1998).
18. M. A. Eberle, M. J. Rieder, L. Kruglyak, D. A. Nickerson, *PLoS Genet.* **2**, 1319 (2006).
19. L. Kruglyak, *Nat. Genet.* **22**, 139 (1999).
20. J. Fooden, in *The Macaques: Studies in Ecology, Behavior, and Evolution*, D. D. Lindburg, Ed. (van Nostrand Reinhold, New York, 1980), pp. 1–9.
21. HapMap, *Nature* **437**, 1299 (2005).
22. We thank the Yerkes, Oregon, and California National Primate Research Centers for contributing samples, and D. G. Torgerson for comments. Funded by NIH grant RR05090 to D.G.S., NIH RR00163 to B.F., NIH RR015383 to J.R., NSF0516310 to C.D.B., and 1R01HG003229 to C.D.B., R.N., A. G. Clark, and T. Mattise. Trace Index numbers are consecutively numbered from 1664051535 to 1664070335 and can be retrieved using the following query: PROJECT_NAME='ENCODE' STRATEGY='Re-sequencing' TRACE_TYPE_CODE='PCR' SPECIES_CODE='MACACA MULATTA'.

26 January 2007; accepted 16 March 2007
10.1126/science.1140462

REPORT

Evolutionary Formation of New Centromeres in Macaque

Mario Ventura,^{1*} Francesca Antonacci,^{1*} Maria Francesca Cardone,¹ Roscoe Stanyon,² Pietro D'Addabbo,¹ Angelo Cellamare,¹ L. James Sprague,³ Evan E. Eichler,³ Nicoletta Archidiacono,¹ Mariano Rocchi^{1†}

A systematic fluorescence in situ hybridization comparison of macaque and human synteny organization disclosed five additional macaque evolutionary new centromeres (ENCs) for a total of nine ENCs. To understand the dynamics of ENC formation and progression, we compared the ENC of macaque chromosome 4 with the human orthologous region, at 6q24.3, that conserves the ancestral genomic organization. A 250-kilobase segment was extensively duplicated around the macaque centromere. These duplications were strictly intrachromosomal. Our results suggest that novel centromeres may trigger only local duplication activity and that the absence of genes in the seeding region may have been important in ENC maintenance and progression.

Evolutionary new centromeres (ENCs) can appear during evolution in a novel chromosomal region with concomitant inactivation of the old centromere. The new centromere then becomes fixed in the species while inevitably progressing toward the complexity typical of a mature mammalian centromere, with intra- and interchromosomal pericentromeric segmental duplications and a large core of satellite DNA (*1*). Unambiguous examples of ENCs were initially reported in primates (*2*) and then described in various other mammalian orders (*3*). A similar phenomenon, well known from clinical cases, is the mitotic rescue of an acentric chromosomal fragment by the opportunistic de novo emergence of a neocentromere (*4*). Recently, two cases of neocentromeres in normal individuals with otherwise normal karyotypes were fortuitously discovered (*5, 6*). These two “in progress” centromeres can be regarded as ENCs at the initial stage, thus reinforcing the opinion that ENCs and clinical neocentromeres are two faces of the same coin. The goal of the research presented here

was to gain insight into the processes and mechanisms of ENC evolution. First, we systematically compared macaque and human synteny organization in search of ENCs. Then, we characterized in detail a macaque ENC and compared it to the orthologous domain in humans, which represents the ancestral genomic structure before ENC seeding.

Multicolor hybridization on rhesus macaque chromosomes [*Macaca mulatta* (MMU) $2n = 42$, where n is the haploid number of chromosomes] of about 500 evenly spaced human bacterial artificial chromosome (BAC) clones

revealed that seven macaque/human homologs (chromosomes 6/5, 8/8, 11/12, 17/13, 19/19, 20/16, and X/X, respectively) were colinear when the position of the centromere was excluded. However, human chromosomes 7/21, 14/15, and 20/22 form syntenic associations as part of three compound macaque chromosomes (3, 7, and 10, respectively). Differences in marker order between macaque and humans were accounted for by 20 chromosome rearrangements. Reiterative fluorescence in situ hybridization (FISH) experiments with additional BAC clones more precisely defined rearrangement breakpoints (table S1). A summary of all results is graphically displayed at www.biologia.uniba.it/macaque. Tables S2 and S3 provide a comprehensive list of the ~600 clones that were used in FISH experiments, from the perspective of both macaque and human chromosomes, respectively.

This comprehensive marker-order comparison revealed that the centromeres of many orthologous chromosomes were embedded in different genomic contexts. To distinguish whether ENC events had occurred in the human or macaque ortholog, or in both, we took into account previous reports that attempted to establish the ancestral form for each chromosome (*2, 3, 6–15*). (Most of these papers use a different macaque chromosome nomenclature; here we follow the nomenclature used by the macaque genome sequencing consortium. For a comparison, see www.biologia.uniba.it/macaque.) The

Table 1. Macaque chromosomes with neocentromeres. The two noncontiguous positions defining, in human, the ENC of chromosome 1 are due to the colocalization of the ENC with a macaque-specific inversion breakpoint.

MMU (HSA)	Clones	Position of the neocentromere on the human sequence	Reference
1 (1)	RP4-621015 RP11-572K18	chr1:226,810,735–226,866,653 chr1:160,918,751–161,035,790	Present study
2 (3)	RP11-355I21/RP11-418B12	chr3:163,822,353–164,707,155	(6)
4 (6)	RP11-474A9	chr6:145,651,644–145,845,896	(9)
12 (2q)	RP11-343I5/RP11-846E22	chr2:138,659,884–138,908,673	Present study
13 (2p)	RP11-722G17	chr2:86,622,638–86,827,260	Present study
14 (11)	RP11-625D10/RP11-661M13	chr11:5,667,339–6,043,020	(10)
15 (9)	RP11-542K23/RP11-64P14	chr9:124,189,785–124,493,134	Present study
17 (13)	RP11-543A19/RP11-527N12	chr13:61,111,769–62,699,203	(3)
18 (18)	RP11-61D1/RP11-289E15	chr18:50,155,761–50,526,34	Present study

¹Department of Genetics and Microbiology, University of Bari, 70126 Bari, Italy. ²Department of Animal Biology and Genetics, University of Florence, Florence 50125, Italy. ³Howard Hughes Medical Institute, Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: rocchi@biologia.uniba.it