and ScLCD1). Indeed, 20.7% (35/169) of the genes that did not confer any DDR phenotypes gave Emb or Ste phenotypes (data not shown).

20. S. Xanthoudakis, T. Curran, EMBO J. 11, 653 (1992).
21. L. Michaux et al., Genes Chromosomes Cancer 15, 38 (1996).
22. R. L. Cai, Y. Yan-Neale, M. A. Cueto, H. Xu, D. Cohen, J. Biol. Chem. 275, 27909 (2000).
23. F. al-Khodairy, T. Enoch, I. M. Hagan, A. M. Carr, J. Cell Sci. 108, 475 (1995).
24. S. Muller, C. Hoege, G. Pyrowolakis, S. Jentsch, Nature Rev. Mol. Cell Biol. 2, 202 (2001).
25. J. C. Venter et al., Science 291, 1304 (2001).
26. E. S. Lander et al., Nature 409, 860 (2001).
27. S. M. Elbashir et al., Nature 411, 494 (2001).
28. M. Vidal, P. Legrain, Nucleic Acids Res. 27, 919 (1999).
29. A. Davy et al., EMBO Rep. 2, 821 (2001).
30. A. J. M. Walhout, M. Vidal, Nature Rev. Mol. Cell. Biol. 2, 55 (2001).
31. A. F. Dernburg, J. Zalevsky, M. P. Colaiacovo, A. M. Villeneuve, Genes Dev. 14, 1578 (2000).
32. We thank J. Reid, A. J. M. Walhout, J. Lamb, M. Tewari, R. Abraham, and S. van den Heuvel for comments on the

Construction and Analysis of a Human-Chimpanzee Comparative Clone Map

Asao Fujiyama,[1,2]*† Hidemi Watanabe,[1]*† Atsushi Toyoda,[1]*
Todd D. Taylor,[1]* Takehiko Itoh,[3]* Shih-Feng Tsai,[4,5]*
Hong-Seog Park,[6]* Marie-Laure Yaspo,[7]* Hans Lehrach,[7]
Zhu Chen,[8]* Gang Fu,[8]* Naruya Saitou,[2]* Kazutoyo Osoegawa,[9]
Pieter J. de Jong,[9] Yumiko Suto,[10] Masahira Hattori,[1]*†
Yoshiyuki Sakaki[1,11]*†

The recently released human genome sequences provide us with reference data to conduct comparative genomic research on primates, which will be important to understand what genetic information makes us human. Here we present a first-generation human-chimpanzee comparative genome map and its initial analysis. The map was constructed through paired alignment of 77,461 chimpanzee bacterial artificial chromosome end sequences with publicly available human genome sequences. We detected candidate positions, including two clusters on human chromosome 21 that suggest large, nonrandom regions of difference between the two genomes.

With the draft human genome sequences (1, 2), comparative genomics has become a powerful approach to extract genetic information from large stretches of nucleotide sequences through identification of conserved regions that are most likely functionally important. Genomic information is also the most valuable resource for understanding the genetic differences between species, a basis for deciphering how genome information is processed into phenotypes. Because chimpanzees are our closest relatives, the differences between us are less than with any other species, yet these differences are more likely to

be important. It has been estimated that the sequence identity between human and chimpanzee is within the range of 98 to 99% (3–12). Thus, comparisons between humans and chimpanzees are the most efficient and effective approach to understand what makes us human.

In this report we present the construction and analysis of a first-generation human-chimpanzee comparative genomic map based on the alignments of 77,461 chimpanzee bacterial artificial chromosome (BAC) end sequences (BESs) to human genomic sequences obtained from the public databases. To prepare the BESs, we used two independently prepared BAC libraries, PTB1 and RPCI-43 (Table 1) (13). Briefly, we sequenced 64,116 BAC clones (roughly 3.3 times coverage of the currently available human contiguous genomic sequence) that produced 114,421 valid BESs (13). The BESs were then aligned with the RefSeq human genome contigs [National Center for Biotechnology Information (NCBI)] through NCBI-BLAST. The number of BESs having an alignment longer than 50 base pairs (bp) with $\geq$90% identity was 77,461 (13). Out of this number, 49,160 BESs from 24,580 clones formed paired ends where each pair was derived from the same clone. Only one end could be successfully aligned from the remaining 28,301 clones. The remaining 36,960 BESs that were not mapped to the human genome were categorized into three different classes: (i) those corresponding to repeat sequences (1168

[1]RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. [2]National Institute of Genetics, Yata 1111, Mishima, Shizuoka 411-8450, Japan. [3]Mitsubishi Research Institute, 3-6, Otemachi 2-chome Chiyoda-ku, Tokyo 100-8141, Japan. [4]National Yang-Ming University, 128, Yenchiu, Yuan Road, Sec. 2, Taipei 115, Taiwan. [5]National Health Research Institutes, 128, Yen-chiu, Yuan Road, Sec. 2, Taipei 115, Taiwan. [6]Genome Research Center, KRIBB, 52 Oun-dong, Yusong-gu, Daejeon 305-333, Korea. [7]Max Planck Institute for Molecular Genetics, Ihnestrasse 73, D-14195 Berlin-Dahlem, Germany. [8]Chinese National Human Genome Center at Shanghai, 351, Guo Shou Jing Road, Shanghai 201203, China. [9]Children's Hospital Oakland Research Institute, BACPAC Resources, 747 52nd Street, Oakland, CA 94609–1809, USA. [10]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Hongou, Bunkyou-ku, Tokyo 113-0033, Japan. [11]Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.

*These authors form the International Consortium for the Sequencing of Chimpanzee Chromosome 22.
†To whom correspondence should be addressed. E-mail: afujiyam@gsc.riken.go.jp, watanabe@gsc.riken.go.jp, hattori@gsc.riken.go.jp, sakaki@gsc.riken.go.jp

Table 1. Summary of BES readings and mapping.

		Number of BAC clones*	Number of BESs
Sequenced		64,116	114,421†
Mapped		52,881	77,461‡
	Paired ends§	24,580	49,160
	Singletons‖	28,301	28,301
Unmapped			36,960
	Repeats		1,168
	Human¶		20,376
	Nonhuman		515
	No hit		14,901

*These clones were selected from the PTB1 or RPCI-43 chimpanzee male BAC libraries, and both ends of the inserts were sequenced (53,240 BESs from PTB1 and 24,221 BESs from RPCI-43 are mapped). †Number of total BESs. ‡Number of mapped BESs. §Number of mapped BESs derived from the same BAC clone. Long BES pairs (>300 kb) were not used for the mapping. ‖Number of the singleton (lone) BESs mapped on the human genome sequence. ¶Contains hits to sequences not included in the NCBI contigs or hits to mRNAs only.

BESs) or showing hits to human sequences not included in the NT contigs (20,376 BESs), (ii) those matched only with sequences from several species other than human (515 BESs), and (iii) the 14,901 BESs that did not match with human sequences, which either correspond to unsequenced human regions or are from chimpanzee regions that have diverged substantially from humans or did not match for other unknown reasons.

The BESs mapped with high confidence (13) were used to calculate the difference between the chimpanzee and human genomes at the nucleotide level. The number of sites in valid alignments (nucleotide sites that have PHRED quality values $q \geq 30$) was 19,813,086. Out of this number, 19,568,394 sites were identical to their human counterparts for a mean percent identity of 98.77. This value is consistent with previous observations (3, 9, 10); however, our calculation comes from a much larger random comparison of slightly less than 1% of the total genome. The distribution of the percent identity of BESs ($q \geq 30$) that have ≥ 100 sites per bin is depicted in Fig. 1. The distribution was skewed to 100% identity instead of the normal distribution, and hence the mode of the difference is at around 0.8%. Although most of the BESs have higher identity, we also found the existence of many low-identity BESs in the genome.

To construct a human-chimpanzee comparative map, we aligned 49,160 paired BESs (practically equivalent to BAC clones) from 24,580 chimpanzee BACs to the RefSeq human genome contigs (build 24, July 2001) constructed by NCBI (13). In addition, 28,301 singleton BAC ends were mapped onto the human genome, even though sequences of the opposite ends were either repeats, low-similarity hits, or not found in the current database (Table 1). The exact positions of these clones will become clearer through the progress of the human genome sequencing and/or the sequencing of the corresponding chimpanzee clones in the near future. The entire map showing the relative positions of the chimpanzee BAC clones (with links to the corresponding human BAC se-

quences) along the human genome contigs (NT contigs, NCBI) and LocusLink (NCBI) information and a comprehensive cross reference table that includes corresponding positions in the human genome and links to the human BAC sequences can be freely accessed through the Web (14).

We found that 48.6% of the whole human genome was covered by the chimpanzee BACs (Table 2). One of the reasons for this apparently low coverage is that we used rather stringent conditions for the calculation; that is, BAC clones were incorporated into the calculation only when they had two sequenced ends in the same NT contig with the correct orientation. Probably because the orientation of draft sequences within the NT contig is sometimes incorrect, 70% of the total paired ends fit the condition. The cov-

erage for chromosomes 14, 20, 21, and 22 was substantially higher. This difference correlates closely with the quality of the human genome sequences used as reference where finished chromosomes and those with longer contigs display higher BAC coverage. We also tested 1 Mb of the human draft sequences corresponding to positions 8012178 to 8426236 and 18502342 to 19012063 of chromosome 21, which we were able to retrieve from the public portion of the Celera database; however, we observed no substantial differences between the mapping results obtained through the public database and that of Celera. Theoretically speaking, the probability of coverage assuming totally random nucleotide distribution was calculated at around 0.7 (15); thus, we concluded that the actual coverage of about 70% for these essentially

Table 2. Estimated coverage of chromosomes by the chimpanzee BACs. The RefSeq human genome contigs (NT sequences) of NCBI, build 24 (July 2001, ftp://ncbi.nlm.nih.gov/genomes/H_sapiens) were used for the calculation.

Chromosome	Sizes (kb)	Covered (kb)	Number of clones*	Coverage (%)	Total BES†
1	222,590	114,161	1,503	51.3	6,561
2	219,687	107,196	1,375	48.8	6,848
3	180,073	70,238	845	39.0	5,071
4	173,018	55,707	670	32.2	4,693
5	163,831	71,411	947	43.6	4,679
6	170,571	104,412	1,378	61.2	5,072
7	149,184	84,480	1,095	56.6	4,408
8	127,942	56,156	692	43.9	3,640
9	111,625	64,427	833	57.7	3,372
10	130,908	72,039	909	55.0	3,771
11	130,943	66,419	830	50.7	4,185
12	126,843	58,650	709	46.2	3,411
13	97,659	56,934	729	58.3	2,750
14	88,311	59,389	789	67.3	2,668
15	75,981	38,992	515	51.3	2,309
16	73,282	34,094	467	46.5	2,390
17	74,182	34,143	487	46.0	2,444
18	72,437	36,303	440	50.1	2,084
19	57,681	19,494	241	33.8	1,388
20	59,154	41,738	570	70.6	1,720
21	33,543	22,614	309	67.4	931
22	33,705	22,351	317	66.3	1,059
X	137,753	34,991	329	25.4	1,642
Y	23,055	1,097	15	4.8	174
Total	2,733,958	1,327,436	16,994	48.6	77,270

*Here, BAC clones are counted only when they have two sequenced ends in the same NT contig with the correct orientation. †Number of BESs (paired and single) allocated to the particular chromosome. There are 191 BESs remaining that are not assigned at this time point.
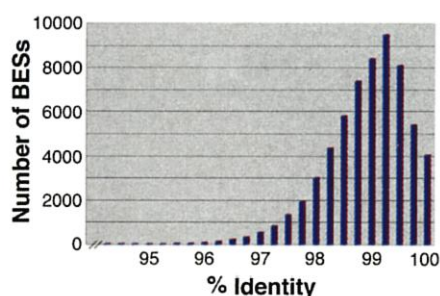

Fig. 1. Distribution of identity scores of the chimpanzee-BES alignments to human genome sequences. Frequencies of BESs with percent identity scores from 94 to 100% are shown.
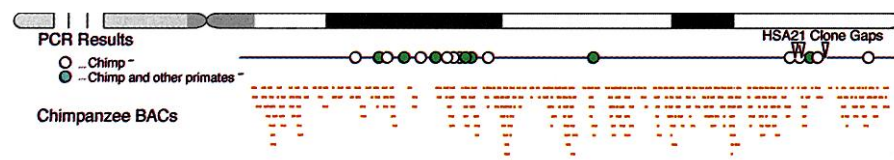

Fig. 2. Distribution of chimpanzee BESs on human chromosome 21 and detection of human chromosome 21–specific STSs in the chimpanzee genome. BAC clones based on the BES alignment are positioned on the human chromosome 21 sequence and shown as red bars (the shortest bar denotes 40 kb). Positions of human STSs that were not detected in the chimpanzee genome are shown in the middle bar with open circles. Filled circles designate STSs that were detected only in human DNA (see text and notes for details). Positions of clone gaps in the human chromosome 21 sequence are also shown as vertical arrowheads.

finished chromosomes is reasonable. Relatively lower coverage of chromosome X (about half of the other unfinished chromosomes) can be explained by the haploid status of the chromosome in the chimpanzee BAC libraries. In contrast, we can only speculate as to why the Y chromosome coverage is so much lower (4.8%) as compared with the other chromosomes (Table 2). One possibility is that the pseudoautosomal region of the Y chromosome confounded our matching al-

gorithm. Another hypothesis may be that the human Y sequences in the DNA databank are so different that the BESs in this study could not find many valid matches. We will have to wait until sequencing of the human Y chromosome is finished to answer some of these questions. Alternatively, we will also need an independent approach, such as the construction and analysis of chimpanzee Y chromosome–specific libraries prepared from sorted chromosomes, generating a complementary

tool to help fully resolve the comparative mapping for the Y chromosome.

Because the genomic sequence of human chromosome 21 is finished to a high degree (16), we analyzed the chimpanzee-human relationship of this chromosome by combining the BES mapping information with a sequence-tagged site (STS)–based approach (17). We identified 18 STSs that amplified products from human DNA but not from that of chimpanzee (shown as circles in Fig. 2). Because we used genomic DNA isolated from three chimpanzee individuals, two males and one female, the effects of relatively larger polymorphisms among the chimpanzee genomes should be minimized. These 18 primer sets, together with the flanking STSs, were further tested with other primates including gorilla. Out of these, amplification products appeared exclusively in humans from seven primer pairs (filled green circles in Fig. 2 whose positions in human chr21 are about 7.2, 8.5, 10.0, 11.6, 11.8, 18.1, and 29.3 Mb from the centromeric end, respectively) (16, 17), suggesting that these loci might correspond to insertions that are specific to the human lineage. Nonhuman primate specific deletions cannot be ruled out but seem less likely because this deletion would have had to occur in all primates but humans. The remaining 11 primer pairs fail to amplify any products from chimpanzee DNAs but showed positive signals in some of the other primates, suggesting the existence of deletions or mutation sites at those positions in the chimpanzee genome. Although one cannot exclude the possibility that these findings are mere reflection of the statistical variation, the region around these sites remains to be the primary target of further investigations in the future. Simple extrapolation of these results suggests that there might be more than several hundred such sites in the entire chimpanzee or human genomes, again including statistical variations.

To identify the boundaries of possible genomic rearrangements, we then searched for candidate clones containing chromosomal breakpoints by inspecting the mapping results. If a clone contains one of the breakpoints of a large genomic rearrangement, the paired BESs of the clone should map to sites separated by an interval longer than expected or the experimentally determined insert size (13). We identified one such example, clone PTB-053J22. One of the BESs and 62% of the finished sequence of this clone matched (>99%) to the 12q15 human BAC clone (AC005294; clone GSHB-410F4 contained in NT_029419.1), and the other BES and 37% of the sequence matched (>98%) to the 12p12 BAC (AC011604; clone RP11-80N2 contained in NT_009700.5). Sequences of these three clones are in finished status and the location of the human clones has been
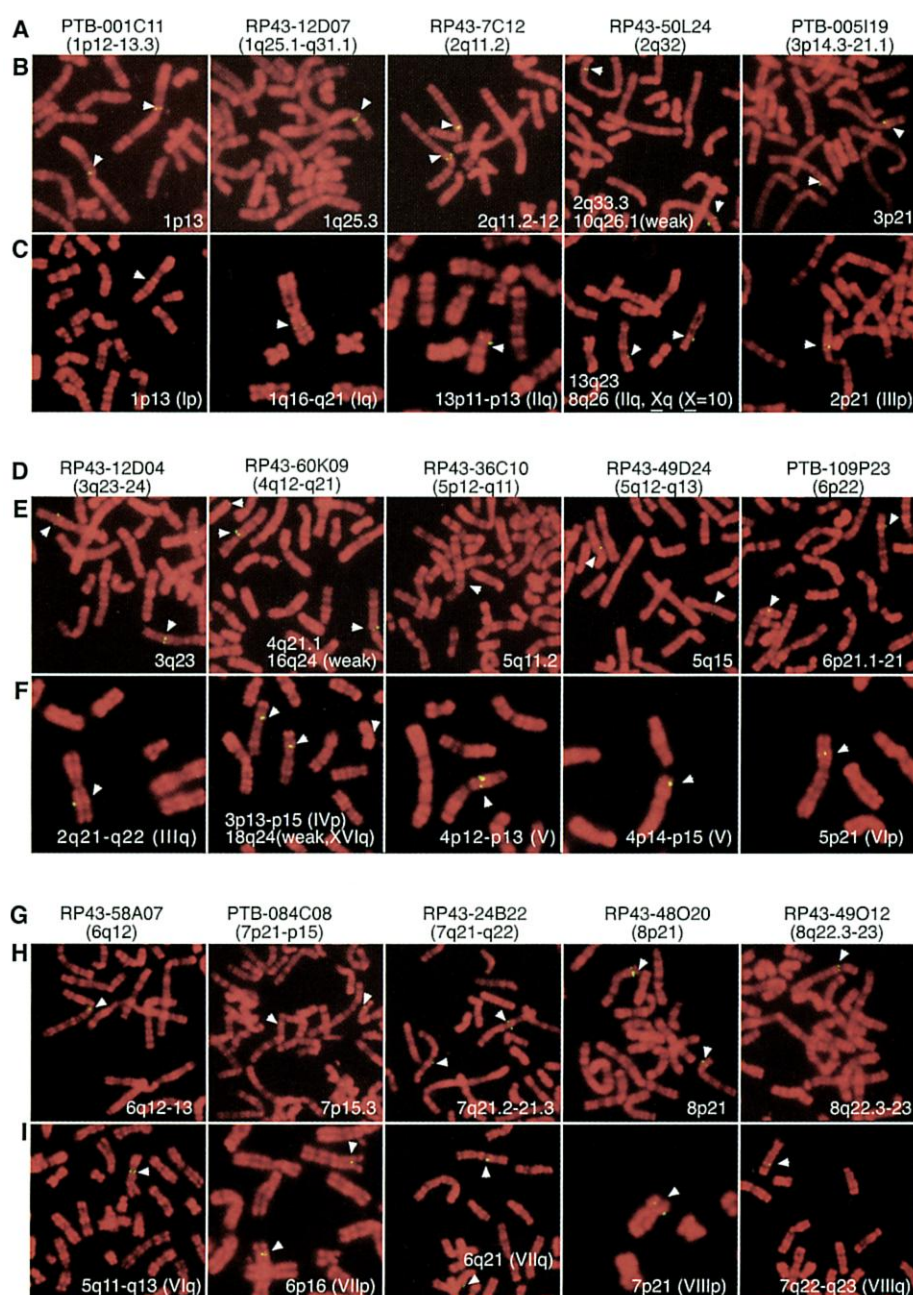

Fig. 3. FISH analysis of selected chimpanzee BAC clones against human and chimpanzee chromosomes. (A, D, and G) Clone names and positions (in parentheses) on human chromosomes determined by the BES analysis (13) [FISH data found in the clone server (www.ncbi.nlm.nih.gov/genome/clone/) were also used]. (B, E, and H) FISH mapping onto human chromosomes. (C, F, and I) FISH mapping onto chimpanzee chromosomes. Chromosome numbers in the parentheses (expressed in Roman letters) are expressed as phylogenetic chromosome number (X = 10) (18). Arrowheads designate the positions of hybridization signals. Magnification of each panel is variable.

confirmed by both electron–polymerase chain reaction (e-PCR) and fluorescent in situ hybridization (FISH) analyses (http://ncbi.nlm.nih.gov/genome/clone). In addition, PTB-053J22 is fully included in another independently sequenced chimpanzee BAC clone (AC007214, RP43-135M19). Thus, it is highly likely that PTB-053J22, which we detected through this study, contains one of the breakpoints corresponding to the human (or vice versa in chimpanzee) chromosomal inversion between 12p12 and 12q15. In addition, this region in human chromosome 12 or in chimpanzee chromosome 10 is known to be inverted in gorilla and chimpanzee as opposed to human and orangutan (8, 18). We found several genes around the PTB-053J22 BESs in the following order, SCL21A14 (solute carrier, organic anion transporter, family 21, member 14), <36 kb>, PTB-053J22-F, <5 kb>, SLC21A8 (solute carrier, organic anion transporter, family 21, member 8, <cen>, DYRK2 (dual-specificity tyrosine-Y-phosphorylation regulated kinase 2), <330 kb>, PTB-053J22-R, <168 kb>, and IFNG (interferon-γ), based on the annotations on the corresponding NT contigs. The effect of the inversion on these genes should be the target of future studies.

To independently test our mapping procedure, we selected 15 chimpanzee BAC clones mapped to human chromosomes 1 to 8 by the BES alignment procedure (13) and subjected them to FISH analysis with both human and chimpanzee M-phase cell spreads (Fig. 3). As shown, 13 clones showed single locus signals at the corresponding positions on both human and chimpanzee chromosomes, and two clones, RP43-50L24 and RP43-60K09, showed similar signals at two loci on the human and chimpanzee chromosomes, suggesting that the mapping procedure we used in this study is reliable. We believe that the whole genome chimpanzee/human comparative map built here by the BES alignment procedure is reasonably accurate and useful for future studies. Recent development of the human-mouse comparative map (19, 20) also supports our approach.

Users of this map should still be careful in applying the information because the possibility remains that assignment of particular clones in the NT contig is incorrect or that inter- or intrachromosomally duplicated regions may be included within an insert. However, the quality of the map, and thus its usefulness, should increasingly improve as the finishing of the human genome sequence proceeds.

References and Notes

1. International Human Genome Sequencing consortium, *Nature* **409**, 860 (2001).
2. J. C. Venter *et. al.*, *Science* **291**, 1304 (2001).
3. M.-C. King, A. C. Wilson, *Science* **188**, 107 (1975).
4. A. Jauch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8611 (1992).
5. B. Dutrillaux, *Hum. Genet.* **48**, 251 (1979).
6. W. Burrows, O. A. Ryder, *Nature* **385**, 125 (1997).
7. M. Goodman *et al.*, *Mol. Phylogenet. Evol.* **9**, 585 (1998).
8. E. Nickerson, D. L. Nelson, *Genomics* **50**, 368 (1998).
9. J. G. Hacia *et al.*, *Nature Genet.* **18**, 155 (1998).
10. H. Kaessmann, V. Wiebe, S. Paabo, *Science* **286**, 1159 (1999).
11. P. Gagneux, A. Varki, *Mol. Phylogenet. Genet.* **18**, 2 (2001).
12. S. Paabo, *Science* **291**, 1219 (2001).
13. Methods to construct the BAC end sequences (BESs) library and the comparative map can be seen at http://hgp.gsc.riken.go.jp or *Science* Online (14). All sequence reads were submitted to the DNA Data-bank of Japan (DDBJ) under accession numbers AG029037 to AG186569 and AG186783 to AG187837.
14. The supplemental data are available at hgp.gsc.riken.go.jp or www.sciencemag.org/cgi/content/full/295/5552/131/DC1.
15. The probability equation, $N = \ln(1 - P)/\ln(1 - 1/n)$, where N is the number of clones, in this case 24,580, and n is the total size (2.7 Gb in this case) divided by the size of fractions. The estimated insert size of the BAC libraries is about 130 to 150 kb. The calculated P value is 0.694 (130-kb insert) or 0.745 (150-kb insert).
16. M. Hattori *et al.*, *Nature* **405**, 311 (2000).
17. In total, 752 primer pairs were designed from the human chromosome 21 sequence at distances of about 50 kb (sequences of the primer pairs can be seen at *Science* Online) (14). For the reactions successful only for human DNA, we expanded the test to include two additional genomic DNA preparations from different individuals of chimpanzee (male and female) and five other primates, including gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*), Old World monkey (*Macaca fascicularis*), New World monkey (*Ateles geoffroyi*), and prosimian (*Lemur catta*).
18. International System for Human Cytogenetic Nomenclature (1985) (ISCN1985) (Karger, Basel, Switzerland, 1985), pp. 95–109.
19. D. Butler, *Nature* **413**, 444 (2001).
20. See http://mouse.ensembl.org.
21. We thank all the members and associates (individual names can be seen at our Web site) of the Human Genome Research Group, RIKEN Genomic Sciences Center, for their enthusiastic support and stimulating discussions. Special thanks to RIKEN GSC and the National Institute of Genetics for supporting the international workshop on chimpanzee genomics, held in Tokyo, March 2001. This work was supported in part by special grants from the Ministry of Education, Culture, Sports, Science and Technology for Genome Research.

Nucleotide Variation Along the *Drosophila melanogaster* Fourth Chromosome

Wen Wang,[1] Kevin Thornton,[2] Andrew Berry,[3] Manyuan Long[1,2]*

The *Drosophila melanogaster* fourth chromosome, believed to be nonrecombining and invariable, is a classic example of the effect of natural selection in eliminating genetic variation in linked loci. However, in a chromosome-wide assay of nucleotide variation in natural populations, we have observed a high level of polymorphism in a ~200-kilobase region and marked levels of polymorphism in several other fragments interspersed with regions of little variation, suggesting different evolutionary histories in different chromosomal domains. Statistical tests of neutral evolution showed that a few haplotypes predominate in the 200-kilobase polymorphic region. Finally, contrary to the expectation of no recombination, we identified six recombination events within the chromosome. Thus, positive Darwinian selection and recombination have affected the evolution of this chromosome.

Detecting evolutionary forces that shape the structure of genetic variation at the genomic level often relies on understanding the effects of natural selection on nearby linked loci (1–4), for which the fourth chromosome of *Drosophila melanogaster* is a classical model system (5–7). It has been thought to undergo no meiotic recombination except under certain experimental conditions (e.g., the interchromosomal effect introduced for the purpose of mapping)

[1]Department of Ecology and Evolution, [2]Committee on Genetics, University of Chicago, 1101 East 57 Street, Chicago, IL 60637, USA. [3]Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed. E-mail: mlong@midway.uchicago.edu

(8, 9). Two genetic models—"selective sweep" by the hitchhiking effect (1, 10), in which an advantageous allele is selected for and fixed in species rapidly, and the background selection model, in which deleterious mutations are selected against (2, 10)—predicted a lack of variation throughout the chromosome, which has been supported by limited data (6, 11). We have reexamined the level of variation and recombination in the fourth chromosome using a chromosome-wide assay of nucleotide variation.

We first investigated within-species nucleotide variation in two adjacent regions from the 102F cytological position of the fourth chromosome of *D. melanogaster*: (i) 4257 base pairs (bp) of the *CG11091* locus, and (ii) 847 bp of an intron of the *toy* gene. *CG11091* and *toy* are separated by ~10 kb. We directly