Evolutionary History of Chromosome 20

Doriana Misceo, Maria Francesca Cardone,* Lucia Carbone,* Pietro D'Addabbo,*† Pieter J. de Jong,‡ Mariano Rocchi,* and Nicoletta Archidiacono**

*Sezione di Genetica, DAPEG, Italy; †Center for Research into Molecular Genetics- Fondazione CARISBO, Institute of Histology and General Embryology, University of Bologna, Bologna, Italy; and ‡Children's Hospital Oakland Research Institute, Oakland, California

The evolutionary history of human chromosome 20 in primates was investigated using a panel of human BAC/PAC probes spaced along the chromosome. Oligonucleotide primers derived from the sequence of each human clone were used to screen horse, cat, pig, and black lemur BAC libraries to assemble, for each species, a panel of probes mapping to chromosomal loci orthologous to the loci encompassed by the human BACs. This approach facilitated marker-order comparison aimed at defining marker arrangement in primate ancestor. To this goal, we also took advantage of the mouse and rat draft sequences. The almost perfect colinearity of chromosome 20 sequence in humans and mouse could be interpreted as evidence that their form was ancestral to primates. Contrary to this view, we found that horse, macaque, and two New World monkeys share the same marker-order arrangement from which the human and mouse forms can be derived, assuming similar but distinct inversions that fully account for the small difference in marker arrangement between humans and mouse. The evolutionary history of this chromosome unveiled also two centromere repositioning events in New World monkey species.

Introduction

Karyotype comparative studies in primates were first attempted in the 1970s using banding techniques. Banding pattern comparison, however, can be misleading when the species under study are not closely related. In addition, differences other than rearrangements can strongly affect the look of chromosomes. Lemur-human sequence comparison has shown, for instance, that repeat-expansion can account for up to 20% difference in DNA content between the two species (Liu et al. 2003). The fluorescence in situ hybridization (FISH) techniques, in the 1990s, introduced a more powerful tool in cytogenetic investigations and, being based on sequence homology, solved many of the problems posed by banding pattern comparison based on visual inspection. Most of comparative FISH studies made use of painting libraries (whole-chromosome or partial-chromosome paints of various origin) as a basic tool for karyotype comparison (Wienberg et al. 1990; Jauch et al. 1992). This approach has the advantage of producing rapid results but lacks resolution, and marker order remains frequently undetermined. The introduction of locus-specific clones, BAC/PAC probes in particular, has strongly contributed to a detailed analysis of chromosomal evolution. In the present paper, we report the evolutionary history of human chromosome 20 in primates studied using a panel of BAC/PAC probes spaced along the chromosome. The study was further extended to six mammalian species to better define the marker organization of primate ancestor.

Chromosome 20 in human, mouse, and rat are almost perfectly colinear (Waterston et al. 2002; Bourque, Pevzner, and Tesler 2004; Gibbs et al. 2004). This observation could be interpreted as strong evidence that this form was ancestral to mammals (Zhao et al. 2004). Contrary to this view, our data suggest that human and mouse/rat forms are derivative. Special attention was paid to the evolutionary history of 20p12 region, where the emergence of a neocentromere in a clinical case was reported (Voullaire et al. 1999).

Key words: chromosome 20, neocentromeres, chromosome evolution.

E-mail: archidiacono@biologia.uniba.it.

Mol. Biol. Evol. 22(2):360–366. 2005
doi:10.1093/molbev/msi021
Advance Access publication October 20, 2004

Methods

Metaphase preparations were obtained from lymphoblastoid or fibroblast cell lines of the following species. Great apes: common chimpanzee (*Pan troglodytes* [PTR]), gorilla (*Gorilla gorilla* [GGO]), and Borneo orangutan (*Pongo pygmaeus pygmaeus* [PPY]); Old World monkeys: rhesus monkey (*Macaca mulatta* [MMU, Cercopithecinae]), African green monkey (*Cercopithecus aethiops* [CAE, Cercopithecinae]), and silvered-leaf monkey (*Presbytis cristata* [PCR, Colobinae]); New World monkey: wooly monkey (*Lagothrix lagothricha* [LLA, Atelinae]), common marmoset (*Callithrix jacchus* [CJA, Callitricinae]), dusky titi (*Callicebus molloch* [CMO, Callicebinae]), and squirrel monkey (*Saimiri boliviensis boliviensis* [SBO, Callicebinae]); prosimians: black lemur (*Eulemur macaco* [EMA, Lemurinae]), ring-tailed lemur (*Lemur catta* [LCA, Lemurinae]), and brown lemur (*Eulemur fulvus* [EFU, Lemurinae]). Marker-order organization in these species was established by FISH experiments using 10 human BAC/PAC clones (table 1, clones in regular style) distributed along chromosome 20 and belonging to RPCI-4, RPCI-5, and RPCI-11 libraries (BACPAC Resources, http://bacpac.chori.org/). FISH experiments were also performed on metaphases obtained from fibroblasts of the following nonprimate mammals: cat (*Catus domesticus* [FCA]), horse (*Equus caballus* [ECA]), pig (*Sus scrofa* [SSC]), and cattle (*Bos taurus* [BTA]), using appropriate probes, obtained as described in *Results*.

DNA extraction from BACs has already been reported (Ventura, Archidiacono, and Rocchi 2001). FISH experiments were performed essentially as described by Lichter et al. (1990). Digital images were obtained using a Leica DMRXA epifluorescence microscope, equipped with a cooled CCD camera (Princeton Instruments, Princeton, NJ).

Molecular Biology and Evolution vol. 22 no. 2 © Society for Molecular Biology and Evolution 2005; all rights reserved.

Table 1
Human Probes Used in the Study

Code	Clone	Accession Number	Mapping	UCSC May 2004	Comment
A1	***RP11-371L19***	*AL118502*	*20p13*	*659,205–785,463*	
A2	***RP11-314N13***	*AL136531*	*20p13*	*1,186,672–1,305,743*	
A3	***RP11-586B11***	*BES*	*20p13*	*1,424,378–1,644,815*	Absent in MUS
A4	***RP11-366B7***	*BES*	*20p13*	*1,574,824–1,775,646*	Absent in MUS
A5	***RP11-636L22***	*BES*	*20p13*	*1,770,721–1,942,465*	Absent in MUS
A	**RP5-1187M17**	AL121891	20p13	3,013,541–3,139,396	
A6	*RP11-684P7*	*BES*	*20p12.3*	*5,343,692–5,528,006*	Splits in PPY
A7	***RP11-474H3***	*BES*	*20p12.3*	*8,766,805–8,940,976*	Defines break in PCR
B1	***RP11-357M12***	*BES*	*20p12.3*	*8,824,725–8,998,446*	Defines break in PCR
B	**RP5-1068F16**	AL023913	20p12.2	10,155,017–10,295,322	
neocen	*RP1-278O22*	*AL135937*	*20p12.2*	*10,662,942–10,722,172*	Neocentromeric region
neocen	***RP4-697P8***	*AL050403*	*20p12.2*	*10,722,073–10,866,996*	Neocentromeric region
neocen	*RP11-103J8*	*BES*	*20p12.2*	*10,793,139–10,947,545*	Neocentromeric region
neocen	*RP5-727I10*	*AL050322*	*20p12.2*	*10,950,574–11,064,725*	Neocentromeric region
neocen	*RP5-784K2*	*AL078588*	*20p12.2*	*11,064,626–11,127,046*	Neocentromeric region
C	RP4-813H11	AL079337	20p12.2	11,379,732–11,417,054	
C1	*RP11-643H5*	*BES*	*20p12.2*	*11,379,493–11,544,313*	Splits in CAE
C2	*RP11-432M9*	*BES*	*20p12.1*	*12,620,845–12,811,812*	Defines break in LLA
D1	*RP11-777O4*	*BES*	*20p12.1*	*12,831,165–13,061,282*	Defines break in LLA
D	RP5-1069O1	AL049633	20p12.1	15,126,843–15,219,199	
D2	*RP11-9L7*	*BES*	*20p12.1*	*16,053,273–16,203,763*	Defines break in PCR
E1	RP11-622P4	*BES*	*20p12.1*	*16,133,945–16,335,170*	Defines break in PCR
E2	*RP11-632C21*	*BES*	*20p11.21*	*24,111,438–24,297,462*	Duplicated in CAE
E3	**RP11-668m24**	*BES*	*20p11.21*	*24,350,347–24,541,210*	Duplicated in CAE
E4	*RP11-984E2*	*BES*	*20p11.21*	*24,417,209–24,575,047*	Duplicated in CAE
E5	*RP11-475N20*	*BES*	*20p11.21*	*24,583,369–24,756,231*	Duplicated in CAE
E	RP5-966J20	AL121925	20p11.21	24,698,120–24,737,379	Duplicated in CAE
E6	***RP4-568C11***	*AL035661*	*20p11.21*	*24,852,731–24,956,411*	Duplicated in CAE
E7	*RP4-738P15*	*AL035252*	*20p11.21*	*25,050,872–25,159,358*	Duplicated in CAE
E8	*RP4-691N24*	*AL031672*	*20p11.21*	*25,364,352–25,465,309*	Duplicated in CAE
E9	*RP4-694b14*	*AL031673*	*20p11.21*	*25,522,225–25,650,091*	Duplicated in CAE
cen					
F	**RP5-836N17**	AL049539	20q11.21	30,126,905–30,238,598	
F1	RP11-663D7	*BES*	*20q11.21*	*33,562,439–33,755,765*	Defines break in CAE
F2	RP11-353C18	AL357374	20q11.22	33,744,285–33,824,849	Defines break in CAE
G	RP5-954P9	AL359828	20q11.23	34,046,335–34,084,879	
G1	*RP11-888D20*	*BES*	*20q11.23*	*34,932,840–35,111,176*	Splits in CAE
G2	*RP11-826B14*	*BES*	*20q11.23*	*35,332,463–35,548,961*	Defines break in CJA
H1	*RP11-138A15*	*BES*	*20q11.23*	*35,387,228–35,535,624*	Defines break in CJA
H	RP5-906C1	AL133342	20q13.13	46,828,731–46,939,544	
I	**RP5-1059L7**	AL121913	20q13.32	55,665,561–55,815,784	
J	RP11-476I15	BES	20q13.33	62,378,079–62,432,679	
tel				62,435,964	

NOTE.—Probes in bold were used to characterize all species. Probes in italics were used to define specific rearrangements. BES stands for BAC ends.

Cy3-dCTP, FluorX-dCTP, DEAC, Cy5-dCTP, and DAPI fluorescence signals, detected with specific filters, were recorded separately as gray-scale images. Pseudocoloring and merging of images were performed using Adobe Photoshop software.

Evolutionary marker-order reconstruction was accomplished by use of the GRIMM software package (Bourque and Pevzner 2002) (http://www.cs.ucsd.edu/groups/bioinformatics/GRIMM/).

Overgo probes of 36 bp each were selected from the searchable database of Universal Probes (http://uprobe.genetics.emory.edu/). They are based on conserved sequences identified by the alignment of human, mouse, and rat genomes. The probes were hybridized to high-density filters following the procedures already described (McPherson et al. 2001), and the images were analyzed with ArrayVision version 6.0 (Imaging Research Inc.). The sequence and location of overgo probes are reported in table 2.

Results

The human BAC/PAC probes reported in table 1 (probes in regular style) were cohybridized on chimpanzee (PTR), gorilla (GGO), Borneo orangutan (PPY), rhesus monkey (MMU), African green monkey (CAE), silvered-leaf monkey (PCR), wooly monkey (LLA), common marmoset (CJA), dusky titi (CMO), and squirrel monkey (SBO) primate species. To better define the chromosome 20 arrangement in primate ancesto, we characterized the organization of this chromosome in the black lemur, ring-tailed lemur, brown lemur, and in six nonprimate mammals for which sequence or BAC libraries were available. To this goal, we screened the BAC libraries of the ring-tailed lemur (library LBNL-2), domestic cat (library RPCI-86), horse (library CHORI 241), and pig (library CHORI 242) for appropriate BAC probes. The sequence encompassed by each human BAC reported in

Table 2
Cat, Horse, Pig, Lemur, and Cattle BAC Probes Used in the Study

Probe	Cat	Horse	Horse	Pig	Lemur	Overgo Sequence	UCSC May 2004	Cattle	Cattle BAC Ends
A	216N10	324H10	324H10	26J24	27A11	TCCAGGAGAAGTGTGCCCAG-TACTGGCCATCTGATG	2,964,226–2,964,261	69M1	376,436–376,721
B	248J2	501B17	501B17		6E12	AAGTAGGTACTGGGTAC-CAGCTCTAATCTGTGGCGT	10,221,646–10,221,681	122I1	7,593,646–7,593,759
C	364P7	324O21	324O21	245M15	48L16	TGGCTAACCCCAAAGTAG-GATGTGGGTGCCAGCCTT	11,175,220–11,175,255	9H9	11,387,226–11,387,504
D	69F6	311K16	311K16	30H3*	—	GGGCCCAGAATACCTGAGT-CAATTTCAGAATCCAAT	15,225,942–15,225,977	30H3	13,446,448–13,446,728
E	196O1	380G19	380G19	88A1*	—	CGATGGCTGCATTTAAG-CAGTTGATAGCCTCTGAAG	23,323,793–23,323,828	88A1	23,925,771–23,925,832
F	148A12	241N14	241N14	115B1	22A19	GGGAATGGATTCACGGTGTG-TAAACAGTGTTCACTG	30,247,236–30,247,271	126P17	29,747,719–29,747,889
G	63A1	50C10*	50C10*	12N13	126N9	AAGGTGCAGCA-GATCGCCCTCTGCTGCTCAA-CATGA	33,998,928–33,998,963	50C10	35,069,845–35,070,199
H	537F4	345B18	345B18	11I23	42K22	ATGCAGGAGAGCCAGACCAA-GAGCATGTTCGTGTCC	46,971,834–46,971,869	49C15	47,067,159–47,067,340
I	—	—	—	—	—	GTTTCCATTTCCTGAGGATTAA-TAAGTGTGGAGGGG	55,591,123–55,591,158	66M22	53,087,093–53,087,128
J	232D3	499I10	499I10	121M4	53O10	AAGAGGACAGACAA-TAGCTGTGGGGAAAGG-TAAGTC	62,266,508–62,266,543	11O3	61,010,604–61,010,639

NOTE.—Species-specific BAC clones were identified by appropriate library screening using overgo sequences reported in column 7. Their positions in the human genome assembling are reported in the following column. Some oligos failed to identify a specific BAC. In pig and horse, the missing probes were occasionally replaced by the corresponding cattle probes, which gave satisfactory FISH signals (BACs with asterisk). Cattle probes were chosen among BACs whose position in the human sequence (last column) was reported by Larkin et al. (2003) (updated to the May 2004 UCSC release). For library names see text.

table 1 (BACs in regular style) was searched for conservation against mouse and rat genomes using the Universal Probes database (see *Methods*). "Overgo" probes (see *Methods* and table 2) were designed on the most conserved region of each BAC (or very close to it) and were then used to screen each specific library. This approach was aimed at assembling, for each species, a panel of BAC probes mapping to chromosomal loci orthologous to the loci encompassed by the human BACs, thus, facilitating mapping comparison. Human chromosome 20 sequence has been recently accomplished (Deloukas et al. 2001). The marker-order organization of this chromosome in mouse (*Mus musculus* [MUS]) and in rat (*Rattus norvegicus* [RNO]) (Waterston et al. 2002; Gibbs et al. 2004) was derived from the draft sequence assembly available at the UCSC (http://genome.ucsc.edu; mouse: May 2004 assembly; rat: June 2003 assembly). Cattle was also included in this study, taking advantage of the large number of BAC probes positioned on the human sequence by BAC end sequencing (Larkin et al. 2003; library CHORI 240).

Additional human BAC probes were utilized in reiterative FISH experiments to exactly define inversion breakpoints in primates. The most informative probes are reported in table 1 (italics) and in figure 1 (red), which summarize the overall results. Splitting signals were interpreted as caused by the occurrence of a breakpoint inside the marker. To reject the possible interpretation that splitting signals were just caused by the presence of duplicons, additional BAC probes, partially overlapping the splitting clone on both sides, were also used.

A neocentromere in the short arm of chromosome 20, located between markers B and C, has been reported by Voullaire et al. (1999) (red dot in figure 1). A 460-kb sequence responsible for the neocentromere functioning was subsequently precisely identified using chromatin immunoprecipitation and array analysis (Lo et al. 2001). To track the evolutionary history of this region, we used supplementary BAC/PAC probes encompassing the neocentromere domain (table 1, probes annotated as "neocen").

The first goal of the study was the assessment of chromosome 20 marker organization in primate ancestor (PA in figure 1). Mouse, rat, and human sequence comparison has shown that human chromosome 20 is an uninterrupted part of mouse chromosome 2 and rat chromosome 3 (Waterston et al. 2002; Gibbs et al. 2004; UCSC, http://genome.ucsc.edu). Human/mouse dot plots analysis indicated that the two sequences are colinear, with the exception of 1.3 Mb stretch (red segment in figure 2), telomerically located in humans (20pter; 70,697 to 1,395,942 bp at UCSC) and located, in mouse and rat, close to sequences corresponding to the 20q11.21, in inverted orientation with respect to humans (mouse 151,557,204 to 152,554,062 bp at UCSC). In humans, a small region (green in figure 2) of about 300 kb (1,395,942 to 1,692,749 bp at UCSC) is lacking in mouse and rat and includes the SIRPB1 and SIRPB2 genes, which appear as a partially duplicated copies of a portion of PTPNS1 gene (exons 2, 3, and 4, with the corresponding introns). Probes from this region (clones A3 to A5) were tested in PTR, GGO, PPY, MMU, PCR, and CJA. All species were found positive for these clones. Also the telomeric 71-kb region is lacking in mouse and rat (blue in figure 2). In this respect, it has to be considered that telomeric regions are highly plastic.

Marker-order definition of the different examined species, as reported in figure 1, shows that horse, macaque, and two New World monkey species (SBO and CMO)

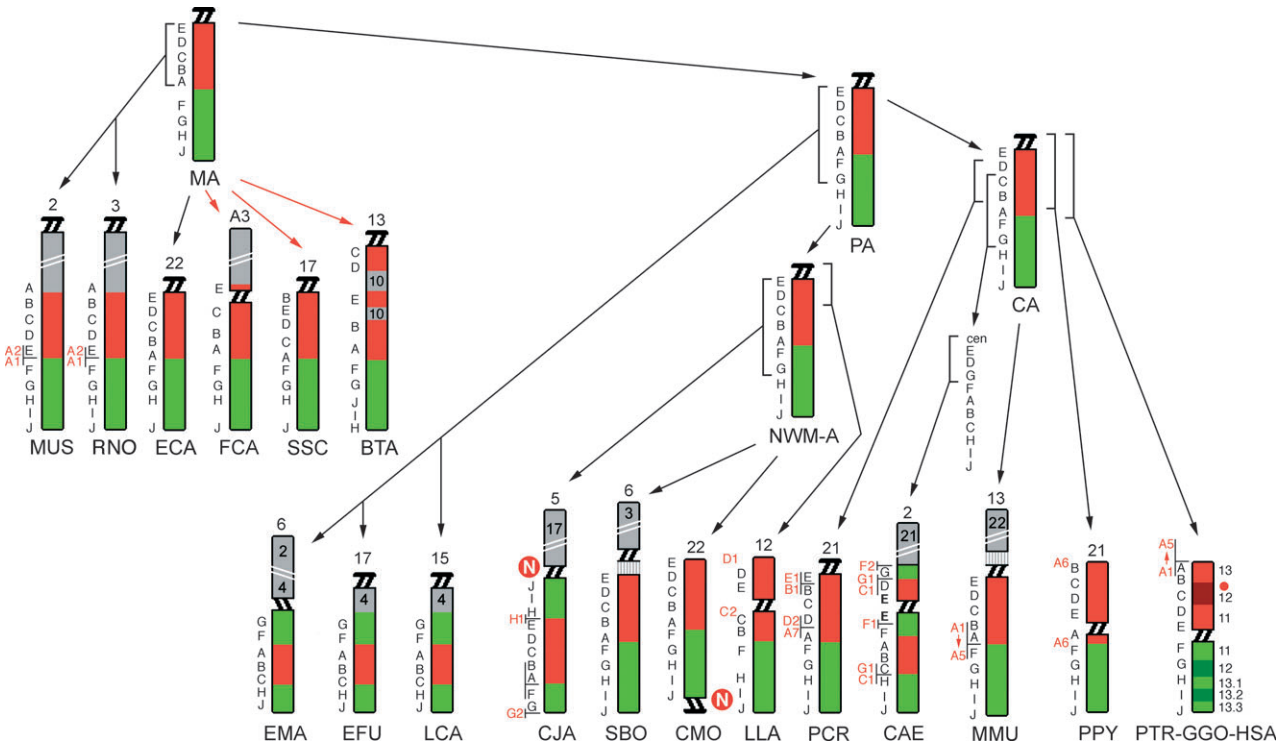

FIG. 1.—Reconstruction of the chromosome 20 evolutionary history in primates and in six mammalian species. Some chromosomes are upside down to facilitate comparison. PA = primate ancestor; CA = Catarrhini ancestor; NWM-A = New World monkey ancestor; MA = mammalian ancestor. The number that identifies the chromosome in each species is reported on top of the chromosome. The black letters on the left of each primate chromosome refer to the panel of BAC probes reported in normal style in table 1. Letters in red are the additional probes used to delimit breakpoints, centromeres, and telomeres, reported in italics in table 1. Cat, horse, pig, and lemur letters refer to probes that have been identified by library screening as reported in table 2. Cattle probes have been identified in the Larkin et al. (2003) BAC ends tables (see table 2). Letters in mouse and rat chromosomes are not based on FISH experiments but on in silico comparative sequence analysis. N in a red circle (new centromere) in CJA and CMO indicate the occurrence of a centromere repositioning. A red dot indicates the region, 20p12, where a neocentromere in a clinical case has been reported by Voullaire et al. (1999). Red arrows point to species for which an intermediate ancestor was not drawn. A single pericentric inversion and two consecutive paracentric inversions are required for cat and pig, respectively. For details see text.

substantially share the same marker arrangement. This arrangement was, therefore, hypothesized as ancestral (see *Discussion*). African great apes share the same chromosome 20 organization with humans. Their marker order can be easily derived from Catarrhini ancestor (CA) hypothesizing

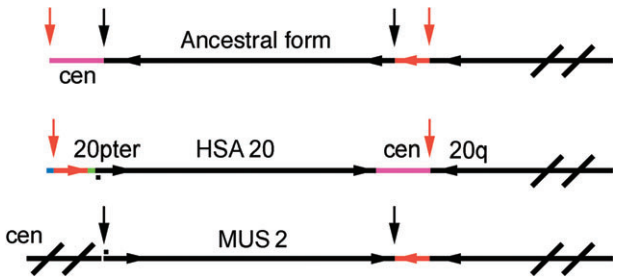

FIG. 2.—The figure summarizes the organization, in humans and mouse, of sequences corresponding to the short arm of chromosome 20, after the two distinct inversions that occurred in the ancestral form. Red arrows indicate inversion breakpoints leading to the human form, and black arrows indicate those leading to the mouse form. The red segment of the sequence was not involved in the mouse inversion. The green segment corresponds to the human duplicated region lacking in mouse. The small black squares indicate the region from which the duplications originated. The telomeric blue segment (71 kb) in human represents the 20p telomere, also lacking in mouse. For details see text.

a pericentric inversion involving the entire short arm. The inversion brought the centromere, telomerically located in CA, to the present-day location. In turn, euchromatic sequences were brought to the 20p telomere. In orangutan, a similar but shorter inversion generated the present-day orangutan chromosome 20. The breakpoint was identified inside marker A6 (BAC RP11-684P7) (fig. 3*a*).

Macaque chromosome 20 organization was unchanged with respect to the hypothesized CA. CAE and PCR, on the contrary, showed a rearranged marker-order organization. A single paracentric inversion generated the PCR arrangement (see figure 1 and table 1), with one breakpoint between A and B and the second between D and E markers. Both breaks were precisely characterized as occurring between overlapping markers D2/E1 and A7/B1 (see table 1). Two distinct inversions are necessary to reconcile the CAE chromosome 20 form with Catarrhini ancestor. Breakpoints of the first inversion (paracentric), delimited by markers C to D and G to H, were precisely located inside marker C1 and G1, respectively, both giving splitting signals (table 1 and figs. 1, 3*b*, and 3*c*). One breakpoint of the second inversion (pericentric) occurred inside the centromere; the second was delimited by markers G and F. The latter breakpoint was further restricted as occurring between the almost overlapping markers F1 and

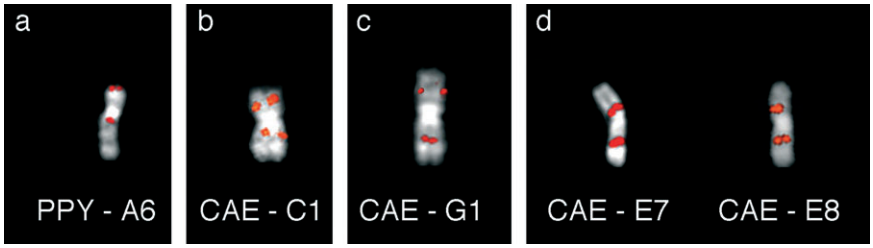

FIG. 3.—FISH examples showing the splitting signal, caused by a pericentric inversion, (*a*) of marker A6 in orangutan, (*b*) of marker C1 in African green monkey (CAE), and (*c*) of marker G1 in CAE. (*d*) Two examples (markers E7 and E8) of FISH signals revealing the large duplication around the centromere of CAE chromosome 2.

F2 (table 1 and fig. 1). Marker E detected signals on both sides of the CAE centromere (data not shown). Several additional BAC clones flanking marker E on both sides (markers E2 to E9 in table 1) were used to delimit the extension of the duplication. The duplicated region was found spanning a large region, approximately 1.5 Mb in size (chr20: 24,111,438 to 25,650,091; FISH examples in figure 3*d*). CAE-duplicated clones were tested in all primates used in this study, but no hint of duplication was found. This large pericentromeric duplication event, therefore, appears to be species specific.

SBO and CMO revealed a conserved marker-order arrangement as compared with the hypothesized PA ancestor. CMO centromere, however, was found located at the opposite telomeric chromosomal end with respect to the hypothesized ancestor, close to marker J, which is very telomeric (table 1). This centromere, therefore, was considered as a new evolutionary centromere (N in a red circle in figure 1). A similar repositioning occurred in CJA (N in a red circle in figure 1). A single inversion differentiates CJA arrangement from PA. One breakpoint was located inside the centromere; the second breakpoint was encompassed by the two overlapping probes G2/H1 (table 1). Also in CJA, the centromere is located close to marker J. A complete marker-order characterization of LLA was not achieved because probes A and G repeatedly failed to yield appreciable FISH signals. There is no doubt, however, on the occurrence of a pericentric inversion that brought the centromere inside the chromosome, with one breakpoint occurring between markers C and D. The breakpoint was restricted to a chromosomal segment defined by the two overlapping probes D1/C2 (table 1 and fig. 1).

EMA, EFU, and LCA prosimians share an identical marker arrangement, which differ from PA for a single paracentric inversion, with one breakpoint between markers F/G and the second break in the pericentromeric area. In all the three species, the segment corresponding to human chromosome 20 is part of a bigger chromosome, marker G flanking sequences belonging to HSA4p16 (Cardone et al. 2002). Horse, as hypothesized above, shares an identical marker order with PA and could also represent the ancestral mammalian form. The position of the centromere in cat slightly differs from the horse and could be the result of a small pericentric inversion.

Discussion

We have reconstructed the evolutionary history of chromosome 20 in primates using a panel of BAC/PAC probes spaced along this chromosome. Six nonprimate mammalian species were also included in the study. The reconstruction made use of the GRIMM software package, designed to outline the most parsimonious scenario of marker-order evolutionary changes (Bourque and Pevzner 2002). In all examined species, chromosome 20 was a unique chromosome or a contiguous part of a bigger chromosome, the only exception being cattle, in which sequences belonging to chromosome 20, located in BTA chromosome 13, are interrupted by sequences of chromosome 10. Differences in marker-order arrangements are only caused by pericentric or paracentric inversions. These observations further support the common opinion that inversions are by far the most common evolutionary rearrangements.

The availability of draft sequence of mouse and rat genomes (Waterston et al. 2002; Bourque, Pevzner, and Tesler 2004; Gibbs et al. 2004) allowed an unparalleled detailed comparison of marker-order arrangement of these species with humans. The nearly perfect conservation of chromosome 20 sequence order along almost the entire length could suggest that this arrangement is ancestral to primates (Zhao et al. 2004). This hypothesis does not fit with present data. An identical marker-order arrangement was found in horse, macaque (MMU and OWM), and two New World monkey species (SBO and CMO). These data strongly support the hypothesis that this form, and not the human form, is ancestral to primates. Chromosome 20 evolutionary relationships reported in figure 1 were drawn according to this assumption. This figure also shows the rearrangements necessary to reconcile this ancestral form with the form of the extant species we have investigated. According to this hypothesis, two similar but distinct inversions generated the human and mouse/rat order arrangements (fig. 2). In humans, the inversion involved the entire short arm, whereas in mouse and rat common ancestor, the inversion did not include the 1.3-Mb stretch whose orientation, therefore, appears to be inverted (red in figure 2).

Segmental duplications are biased against pericentromeric and telomeric regions (Bailey et al. 2002). A more detailed and accurate picture of pericentromeric and subtelomeric duplications has been recently achieved (Riethman et al. 2004; She et al. 2004). Both works have shown that the amount of segmental duplication varies considerably among pericentromeric or subtelomeric regions and that some of them appear almost devoid of duplicons. This fact could be the result of incomplete

sequencing or inadequacy in sequence assembling (Bailey et al. 2001). In specific cases, however, the evolutionary history of human chromosomes seems to provide a more adequate explanation. Evolution of human chromosome 3, for instance, has shown that its centromere is an evolutionary new centromere whose seeding occurred before the great apes divergence (Ventura et al. 2004). Duplicon scarcity of the pericentromeric regions of this chromosome appear, therefore, quite predictable. Chromosome 18 organization differs in humans and great apes for a pericentromeric inversion that brought the centromere, telomerically located in human ancestor, to its present-day location (Dennehey et al. 2004; Goidts et al. 2004). This inversion perfectly accounts for the duplicons absence in the long arm–side pericentromeric region of this chromosome (She et al. 2004). Pericentromeric duplicons on the 20q side are considerably less abundant than on the 20p side. The inversion that occurred after orangutan divergence could adequately account for the reduced size of centromeric transition region of 20q with respect to 20p.

Armengol et al. (2003) and Bailey et al. (2004) have reported a peculiar association between duplicons in humans and lineage-specific breakpoints in mouse. One of the two inversion breakpoints we have hypothesized in mouse map to the boundary of a region that includes duplicated sequences not represented in the mouse/rat genome, thus, providing an interesting example of this intriguing association. Markers A3 to A5 (table 1) span this region, with no counterpart in mouse and rat. They gave FISH signals in all tested primates (PTR, GGO, PPY, MMU, PCR, and CJA), thus, suggesting that the duplications occurred early in primate evolution. Duplications inside these BACs are of small size and absent in marker A4. Therefore, FISH signals, at least in case of A4, cannot be attributed to sequence cross-hybridization. Several distinct pieces of evidence support the conclusion that the duplicated 300 kb in humans (1,395,942 to 1,692,749 bp) are absent in mouse: (1) the mouse regions corresponding to the breakpoints in humans are continuous (no gap present); (2) SIRPB1 and SIRPB2 genes appear to be composed of duplicated exons of the PTPNS1 gene, which is single copy in mouse; (3) SIRPB1 and SIRPB2 do not appear in the very rich mouse EST collection; and (4) the sequence similarity among the different duplicated exons suggests that they emerged during distinct rounds of duplication and dates them after rodents/primates divergence.

Yunis and Prakash (1982) suggested that in orangutan, the terminal band 20p13 was inserted into 8q, close to the centromere, and that a paracentric inversion occurred in the long arm. We have documented a single rearrangement consisting in a species-specific pericentric inversion that brought the telomeric centromere to the present-day location, with the second inversion breakpoint inside marker A6 (fig. 1). With respect to the similar but distinct and larger inversion that occurred in PTR-GGO-HSA ancestor, a 5.5-Mb chromosomal segment, now telomeric in humans (20p13), was not involved in the PPY inversion. As a result of the inversion, this region now flanks the centromere on PPY 20q.

Evolutionary centromere repositioning is a biological phenomenon we have recently described (Montefalcone et al. 1999). It appears to be relatively common in primates (Ventura et al. 2001; Carbone et al. 2002; Eder et al. 2003; Ventura et al. 2004), and examples have been reported in non-primate mammals and in birds (Band et al. 2000; Kasai et al. 2003; Yang et al. 2004). Data from the present paper suggest the occurrence of centromere repositioning in CJA and CMO (see figure 1). Chromosome 3 evolutionary studies in primates (Ventura et al. 2004) provided examples of centromeres alternatively positioned at opposite telomeric location of orthologous chromosomes in three different New World monkey species (CJA, CMO, and LLA). The two centromere repositionings we have documented in the present paper, in CJA and CMO, add further examples to the latter observation and seem to indicate that New World monkeys appear to be particularly prone to centromere repositioning events.

Voullaire et al. (1999) have reported a neocentromere at 20p12 and, as stated, the 460 kb responsible of the neocentromeric function were later on precisely defined (Lo et al. 2001). This 460-kb region is located at 10,709 to 11,174 kb, close to marker C. We have recently reported that neocentromeres at 15q24-26 map to duplicons that flanked an ancestral inactivated centromere (Ventura et al. 2003). We extensively searched to determine whether the 20p12 neocentromeric region was the site of an ancestral, inactivated centromere in other primates or in nonprimate mammals. Interestingly, marker C is close to the centromere in cattle. The centromere location, however, appears to be derivative with respect to the horse ancestral form. It could be hypothesized that an evolutionary centromere repositioning event and a neocentromere seeding in a clinical case occurred in the same region. Any relationship, however, could not be established with certainty. Pericentromeric duplicons, after centromere inactivation, can disperse in a very large area (Eder et al. 2003; Ventura et al. 2003), making their evolutionary tracking a very difficult task.

Acknowledgments

CEGBA (Centro di Eccellenza Geni in campo Biosanitario e Agroalimentare), MIUR (Ministero Italiano della Universita' e della Ricerca; Cluster C03, Prog. L.488/92), and European Commission (INPRIMAT, QLRI-CT-2002–01325) are gratefully acknowledged for financial support. Mammal cell lines were kindly provided by the Cambridge Resource Centre (http://www.vet.cam.ac.uk/genomics).

Literature Cited

Armengol, L., M. A. Pujana, J. Cheung, S. W. Scherer, and X. Estivill. 2003. Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements. Hum. Mol. Genet. **12**:2201–2208.

Bailey, J. A., R. Baertsch, W. J. Kent, D. Haussler, and E. E. Eichler. 2004. Hotspots of mammalian chromosomal evolution. Genome Biol. **5**:R23.

Bailey, J. A., Z. Gu, R. A. Clark, K. Reinert, R. V. Samonte, S. Schwartz, M. D. Adams, E. W. Myers, P. W. Li, and E. E. Eichler. 2002. Recent segmental duplications in the human genome. Science **297**:1003–1007.

Bailey, J. A., A. M. Yavor, H. F. Massa, B. J. Trask, and E. E. Eichler. 2001. Segmental duplications: organization and impact within the current human genome project assembly. Genome Res. **11**:1005–1017.

Band, M. R., J. H. Larson, M. Rebeiz et al. (11 co-authors). 2000. An ordered comparative map of the cattle and human genomes. Genome Res. **10**:1359–1368.

Bourque, G., and P. A. Pevzner. 2002. Genome-scale evolution: reconstructing gene orders in the ancestral species. Genome Res. **12**:26–36.

Bourque, G., P. A. Pevzner, and G. Tesler. 2004. Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. Genome Res. **14**:507–516.

Carbone, L., M. Ventura, S. Tempesta, M. Rocchi, and N. Archidiacono. 2002. Evolutionary history of chromosome 10 in primates. Chromosoma **111**:267–272.

Cardone, M. F., M. Ventura, S. Tempesta, M. Rocchi, and N. Archidiacono. 2002. Analysis of chromosome conservation in *Lemur catta* studied by chromosome paints and BAC/PAC probes. Chromosoma **111**:348–356.

Deloukas, P. L., H. Matthews, J. Ashurst et al. (127 co-authors). 2001. The DNA sequence and comparative analysis of human chromosome 20. Nature **414**:865–871.

Dennehey, B. K., D. G. Gutches, E. H. McConkey, and K. S. Krauter. 2004. Inversion, duplication, and changes in gene context are associated with human chromosome 18 evolution. Genomics **83**:493–501.

Eder, V., M. Ventura, M. Ianigro, M. Teti, M. Rocchi, and N. Archidiacono. 2003. Chromosome 6 phylogeny in primates and centromere repositioning. Mol. Biol. Evol. **20**:1506–1512.

Gibbs, R. A., G. M. Weinstock, M. L. Metzker et al. (229 co-authors). 2004. Genome sequence of the brown Norway rat yields insights into mammalian evolution. Nature **428**: 493–521.

Goidts, V., J. M. Szamalek, H. Hameister, and H. Kehrer-Sawatzki. 2004. Segmental duplication associated with the human-specific inversion of chromosome 18: a further example of the impact of segmental duplications on karyotype and genome evolution in primates. Hum. Genet. **115**: 116–122.

Jauch, A., J. Wienberg, R. Stanyon, N. Arnold, S. Tofanelli, T. Ishida, and T. Cremer. 1992. Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting. Proc. Natl. Acad. Sci. USA **89**:8611–8615.

Kasai, F., C. Garcia, M. V. Arruga, and M. A. Ferguson-Smith. 2003. Chromosome homology between chicken (*Gallus gallus* domesticus) and the red-legged partridge (*Alectoris rufa*); evidence of the occurrence of a neocentromere during evolution. Cytogenet. Genome Res. **102**:326–330.

Larkin, D. M., A. Everts-van der Wind, M. Rebeiz et al. (15 co-authors). 2003. A cattle–human comparative map built with cattle BAC-ends and human genome sequence. Genome Res. **13**:1996–1972.

Lichter, P., C.-J. Tang Chang, K. Call, G. Hermanson, G. A. Evans, D. Housman, and D. C. Ward. 1990. High resolution mapping of human chromosomes 11 by in situ hybridization with cosmid clones. Science **247**:64–69.

Liu, G., N. C. Program, S. Zhao, J. A. Bailey, S. C. Sahinalp, C. Alkan, E. Tuzun, E. D. Green, and E. E. Eichler. 2003. Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. Genome Res. **13**:358–368.

Lo, A. W., D. J. Magliano, M. C. Sibson, P. Kalitsis, J. M. Craig, and K. H. Choo. 2001. A novel chromatin immunoprecipitation and array (CIA) analysis identifies a 460-kb cenp -a-binding neocentromere DNA. Genome Res. **11**:448–457.

McPherson, J. D., M. Marra, L. Hillier et al. (113 co-authors). 2001. A physical map of the human genome. Nature **409**:934–941.

Montefalcone, G., S. Tempesta, M. Rocchi, and N. Archidiacono. 1999. Centromere repositioning. Genome Res. **9**:1184–1188.

Riethman, H., A. Ambrosini, C. Castaneda, J. Finklestein, X.-L. Hu, U. Mudunuri, S. Paul, and J. Wei. 2004. Mapping and initial analysis of human subtelomeric sequence assemblies. Genome Res. **14**:18–28.

She, X., J. E. Horvath, Z. Jiang et al. (18 co-authors). 2004. The structure and evolution of centromeric transition regions within the human genome. Nature **430**:857–864.

Ventura, M., S. Weigl, L. Carbone et al. (14 co-authors). 2004. Recurrent sites for new centromere seeding. Genome Res. **14**:1696–1703.

Ventura, M., N. Archidiacono, and M. Rocchi. 2001. Centromere emergence in evolution. Genome Res. **11**:595–599.

Ventura, M., M. Boniotto, M. F. Cardone, L. Fulizio, N. Archidiacono, M. Rocchi, and S. Crovella. 2001. Characterization of a highly repeated DNA sequence family in five species of the genus Eulemur. Gene **275**:305–310.

Ventura, M., J. M. Mudge, V. Palumbo, S. Burn, E. Blennow, M. Pierluigi, R. Giorda, O. Zuffardi, N. Archidiacono, M. S. Jackson, and M. Rocchi. 2003. Neocentromeres in 15q24-26 map to duplicons which flanked an ancestral centromere in 15q25. Genome Res. **13**:2059–2068.

Voullaire, L., R. Saffery, J. Davies, E. Earle, P. Kalitsis, H. Slater, D. V. Irvine, and K. H. Choo. 1999. Trisomy 20p resulting from inverted duplication and neocentromere formation. Am. J. Med. Genet. **85**:403–408.

Waterston, R. H., K. Lindblad-Toh, E. Birney et al. (222 co-authors). 2002. Initial sequencing and comparative analysis of the mouse genome. Nature **420**:520–562.

Wienberg, J., A. Jauch, R. Stanyon, and T. Cremer. 1990. Molecular cytotaxonomy of primates by chromosomal in situ suppression hybridization. Genomics **8**:347–350.

Yang, F., B. Fu, P. C. M. O'Brien, W. Nie, O. A. Ryder, and M. A. Ferguson-Smith. 2004. Refined genome-wide comparative map of the domestic horse, donkey and human based on cross-species chromosome painting: insight into the occasional fertility of mules. Chromosome Res. **12**:65–76.

Yunis, J. J., and O. Prakash. 1982. The origin of man: a chromosomal pictorial legacy. Science **215**:1525–1530.

Zhao, S., J. Shetty, L. Hou, A. Delcher, B. Zhu, K. Osoegawa, P. De Jong, W. C. Nierman, R. L. Strausberg, and C. M. Fraser. 2004. Human, mouse, and rat genome large-scale rearrangements: stability versus speciation. Genome Res. **14**:1851–1860.

Mark Ragan, Associate Editor

Accepted October 5, 2004