

Serial segmental duplications during primate evolution result in complex human genome architecture

Paweł Stankiewicz,¹ Christine J. Shaw,¹ Marjorie Withers,¹ Ken Inoue,^{1,4}
James R. Lupski^{1,2,3,5}

Departments of ¹Molecular & Human Genetics, and ²Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA;

³Texas Children's Hospital, Houston, Texas 77030, USA

The human genome is particularly rich in low-copy repeats (LCRs) or segmental duplications (5%–10%), and this characteristic likely distinguishes us from lower mammals such as rodents. How and why the complex human genome architecture consisting of multiple LCRs has evolved remains an open question. Using molecular and computational analyses of human and primate genomic regions, we analyzed the structure and evolution of LCRs that resulted in complex architectural features of the human genome in proximal 17p. We found that multiple LCRs of different origins are situated adjacent to one another, whereas each LCR changed at different time points between >25 to 3–7 million years ago (Mya) during primate evolution. Evolutionary studies in primates suggested communication between the LCRs by gene conversion. The DNA transposable element MER1-Charlie3 and retroviral ERVL elements were identified at the breakpoint of the t(4;19) chromosome translocation in *Gorilla gorilla*, suggesting a potential role for transpositions in evolution of the primate genome. Thus, a series of consecutive segmental duplication events during primate evolution resulted in complex genome architecture in proximal 17p. Some of the more recent events led to the formation of novel genes that in human are expressed primarily in the brain. Our observations support the contention that serial segmental duplication events might have orchestrated primate evolution by the generation of novel fusion/fission genes as well as potentially by genomic inversions associated with decreased recombination rates facilitating gene divergence.

Recent computational analyses of the available DNA sequencing data and fluorescence in situ hybridization (FISH) studies have estimated that low-copy repeats (LCRs) (paralogous segmental duplications, or duplicons), which are ≥95%–97% identical, occupy 5%–10% of the human genome (Bailey et al. 2001; Cheung et al. 2001; Eichler 2001; International Human Genome Sequencing Consortium 2001; Venter et al. 2001). In some regions, they constitute up to 23%–45% of the examined sequence (Hurles and Jobling 2003; Stankiewicz et al. 2003). Because of their repeat nature, LCRs have been a significant challenge to the Human Genome Project and many of the remaining gaps in euchromatic regions are associated with LCRs (Stankiewicz and Lupski 2002; Eichler et al. 2004).

The genome architecture resulting from the size, orientation, and arrangement of LCRs was shown to be responsible for genomic instability (Lupski 1998; Emanuel and Shaikh 2001; Stankiewicz and Lupski 2002a). Serving as substrates for non-allelic (or “ectopic”) homologous recombination (NAHR), LCRs facilitate meiotic and potentially mitotic DNA rearrangements associated with several diseases that are referred to as genomic disorders. Genomic rearrangements can be responsible for Mendelian traits, contiguous gene syndromes, and whole-arm chromosome aberrations (Lupski 1998, 2003; Stankiewicz and Lupski

2002a). During the last decade, substantial molecular data have emerged documenting that LCR-mediated NAHR is a major mechanism for human disease (Emanuel and Shaikh 2001; Stankiewicz and Lupski 2002a; Shaw and Lupski 2004).

Several studies have shown that LCRs arose recently, apparently during primate evolution (Stankiewicz and Lupski 2002b). Recent data suggest that LCR-associated genome architecture does not represent simple segmental duplications but rather reflects complex rearrangements that are potentially the end product of multiple events (van Geel et al. 2002). These serial segmental duplications can result in a complex shuffling of genomic sequences (Babcock et al. 2003; Lupski 2003).

The gene- and LCR-rich human genomic region 17p11.2-p12 is rearranged in a variety of different constitutional, evolutionary, and cancer-associated structural chromosome aberrations, and thus appears to be an excellent model to investigate the role of genome architecture in DNA rearrangements (Pentao et al. 1992; Chance et al. 1994; Chen et al. 1997; Reiter et al. 1998; Potocki et al. 2000; Stankiewicz et al. 2001; Park et al. 2002; Bi et al. 2003; Barbouti et al. 2004). Recently, we documented that LCRs in proximal 17p constitute >23% of ~7.5 Mb of genome sequence (Fig. 1). We also obtained evidence that genome architecture is responsible for rearrangements with nonrecurrent breakpoints, thus supporting the notion that chromosomal rearrangements are not random events but rather reflect structural features of the human genome (Stankiewicz et al. 2003). How such complex architecture arose, often with two, three, or more copies and also as complex mosaic repeats, remains unknown.

Using parallel approaches consisting of FISH studies in several primate species and computational analyses of human ge-

⁴Present address: Department of Mental Retardation and Birth Defect Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo 187-8502, Japan.

⁵Corresponding author.

E-mail: jlupski@bcm.tmc.edu; fax: (713) 798-5073.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.2746604>.

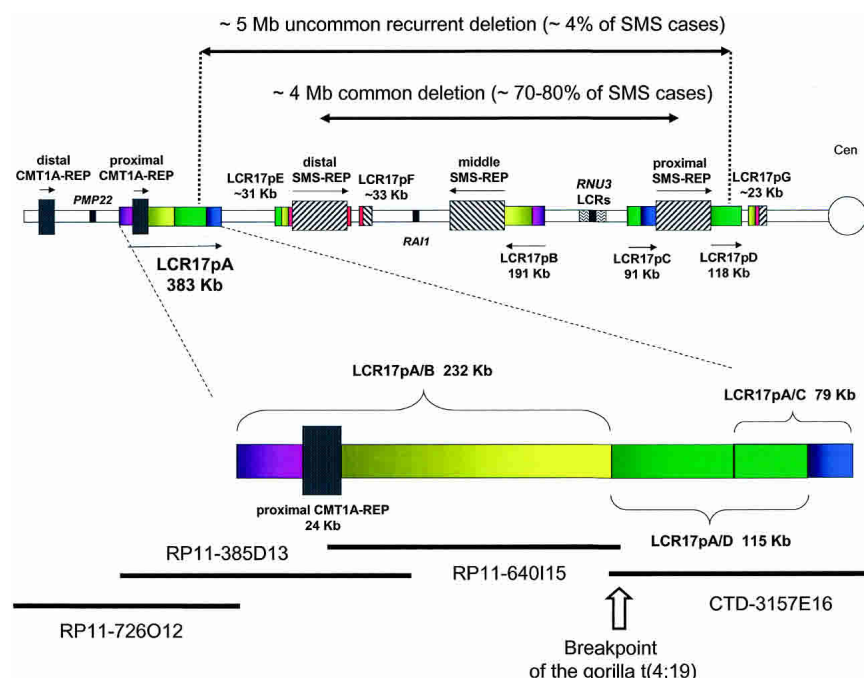


Figure 1. Schematic diagram of the LCR-rich human proximal chromosome 17p. (Top) Depiction of proximal chromosome 17p showing the position and orientation of LCRs. The LCR17p structures are depicted as rectangles with colors signifying shared homology and horizontal arrows showing relative orientation. (Bottom) Magnification of the LCR17pA structure with four BAC clones covering the region. The gorilla translocation breakpoint is indicated by an open vertical arrow between BAC clones RP11-640I15 and CTD-3157E15 and LCR17pA/B and LCR17pA/D subunits of the LCR17pA copy. The LCR17pE, LCR17pF, and LCR17pG have been described elsewhere (Stankiewicz et al. 2003). The LCRs that are >20 Kb are depicted.

nome sequence (molecular clock analysis), we investigated the complex structure and evolution of the genome architecture in an ~7.5 Mb region of proximal 17p. Our analyses support a model of serial segmental duplication events of LCRs during primate evolution.

Results

Structure and orientation of LCR17p repeats

The proximal short arm of human chromosome 17 contains two sets of well-characterized LCRs that play a major role in recurrent disease-causing genomic rearrangements, namely proximal and distal CMT1A-REPs (associated with Charcot-Marie-Tooth type 1A disease, [CMT1A], and hereditary neuropathy with liability to pressure palsies, [HNPP]) and proximal, middle, and distal SMS-REPs (associated with Smith-Magenis syndrome [SMS] and dup(17)(p11.2p11.2) syndrome). Recently, we provided evidence for the existence of yet another set of LCRs in 17p11.2-p12 termed LCR17pA, B, C, D, E, F, and G (Stankiewicz et al. 2003). We have analyzed genomic sequence spanning LCR17pA, LCR17pB, LCR17pC, and LCR17pD in 17p11.2-p12 (Fig. 1). Based on the genomic sequence information spanning these LCRs, we determined the size, orientation, and extent of homology for and between each copy. BLAST comparisons of LCR17p repeats revealed that LCR17pA is composed of

three subunits, which we termed LCR17pA/B (~232 Kb, homologous and in an inverted orientation with respect to LCR17pB) and adjacent to it on the centromeric side two overlapping sequences LCR17pA/C (~79 Kb) and LCR17pA/D (~115 Kb) (both homologous and directly oriented with respect to LCR17pC and LCR17pD, respectively). Interestingly, the LCR17pC and LCR17pD copies flank the proximal SMS LCR, termed proximal SMS-REP (Fig. 1). Also of interest, the directly oriented LCR17pA/D and LCR17pD are 98.7% identical (Table 1) and can act as alternative homologous recombination substrates resulting in an ~5 Mb deletion associated with SMS in about 4% of patients (Shaw et al. 2004b), whereas the directly oriented distal and proximal SMS-REP copies are the substrates for the common ~4 Mb deletion observed in 70%–80% of SMS patients (Chen et al. 1997; Park et al. 2002).

Evolution of LCR17pB, LCR17pC, and LCR17pD copies from LCR17pA

DNA sequence comparison excluding insertion/deletion (indels), or molecular clock analysis, among LCR17p copies showed 98.70% nucleotide identity between LCR17pA/B and LCR17pB and 98.57% between LCR17pA/D and LCR17pD, similar to the 98.15% (98.7% including indels; [Reiter et al. 1997; Inoue et al. 2001]) homology between CMT1A-REPs (Table 1). LCR17pC shares less sequence identity with LCR17pA/C (88.45%) and LCR17pD (88.03%), suggesting that LCR17pC may be older than LCR17pB, LCR17pD, and proximal CMT1A-REP (Table 1). Such molecular clock estimates, however, are based on the neutral theory for mutation rate in evolution, with a presumption that paralogous sequences evolve independently. Molecular clock analysis may be confounded by LCRs that are potentially interacting with each other genetically (i.e., by gene conversion) during evolution. Depending on the size, homology, distance, orientation, and frequency of continuous non-allelic misalignments between different LCR copies for millions of years, homologous recombination may have resulted in gene conversion and/or DNA sequence homogenization.

As an independent assessment of the timing of segmental duplication events during primate evolution, we performed hybridization-based evolutionary studies using FISH on various primate cell lines. Hybridization of human BAC clone CTD-

Table 1. DNA sequence homology comparison among different LCR17p copies

	LCR17pB	LCR17pC	LCR17pD	Proximal CMT1A-REP	Middle SMS-REP	Distal SMS-REP
LCR17pA	98.70%	88.45%	98.57%	—	—	—
LCR17pC	—	—	88.03%	—	—	—
distal CMT1A-REP	—	—	—	98.15%	—	—
proximal SMS-REP	—	—	—	—	98.20%	98.13%
middle SMA-REP	—	—	—	—	—	98.29%

3157E16, derived from the LCR17pC and LCR17pD subunits of the LCR17pA copy, revealed the presence of at least two signals in cell lines analyzed from different human (Fig. 2A) and primate species (chimpanzee, gorilla, orangutan [Fig. 2B], Old World monkey baboon [Fig. 2C], and New World monkey squirrel). In addition, we identified a 3-kb fragment overlapping proximal SMS-REP and LCR17pD. A copy of this fragment was also found in distal SMS-REP (which arose from proximal SMS-REP, Park et al. 2002), indicating that proximal SMS-REP and LCR17pD existed at the time of SMS-REP duplication. The overlap within LCR17pA of the regions similar to LCR17pC and LCR17pD indicates that LCR17pA is a progenitor to both LCR17pC and LCR17pD.

FISH with LCR17pA/B-derived human genomic clone RP11-640I15 (Fig. 1) in cell lines from human, chimpanzee, gorilla, orangutan, and Old World monkey baboon showed the presence of two signals on each chromosome homologue in human (Fig. 2D), chimpanzee (Fig. 2E), and gorilla (Fig. 2F) and only one signal in more distant primate species (Fig. 2G,H,I), indicating that the segmental duplication event resulting in the LCR17pB copy occurred after the divergence of gorilla and orangutan, ~7–12 Mya (Kumar and Hedges 1998). To determine which copy was the progenitor, using dual-color FISH on orangutan cell lines, we cohybridized human BAC clones RP11-640I15 and LCR-free RP11-726O12; the latter is ~50–100 kb telomeric to the former in the human genome (Fig. 1). The presence of distinct signals would indicate that the LCR17pB copy was a progenitor copy,

whereas overlapping signals would be most consistent with LCR17pA/B representing the original copy. We observed overlapping fluorescence signals in the multiple interphase nuclei examined (Fig. 2G). Thus, after the divergence of orangutan and gorilla, the ~191-Kb LCR17pB copy resulted from an insertional duplication associated with two smaller deletions (41 kb). Interestingly, the LCR17pB segmental duplication was positioned directly adjacent to the middle SMS-REP (Fig. 3).

Complex DNA exchange occurred at the gorilla t(4;19) translocation

Using the year 2000 human genome sequence from NCBI, we mapped the gorilla t(4;19) breakpoints by FISH to human BAC clone CITC-428H11 (HSA 5q13.3) and overlapping clones RP11-640I15 and CTD-3157E16 (HSA 17p12) (Stankiewicz et al. 2001). The current build 34 (July 2003) alignment of BAC clones in the human chromosome region 5q13-q14 revealed several changes when compared with the 2000 assembly; for example, BAC clone CITC-428H11 currently maps to chromosome 5q14.3. Interestingly, this chromosome region has been shown recently to be LCR-rich and evolutionarily unstable (Courseaux et al. 2003). Repeated FISH experiments showed that the BAC clone CITC-428H11 mapped only to the GGO 19 and did not span the breakpoint, suggesting mis-mapping or contamination in our previous breakpoint mapping experiment. Instead, two overlapping human BAC clones CTD-2327L5 and RP11-90A9 (5q14.1) were found to span the GGO 19 breakpoint (Fig. 4). A Genescan-predicted gene NT_006713.229 associated with the EST BI601141 expressed in hypothalamus was found to be disrupted by the translocation breakpoint.

To identify a gorilla BAC genomic clone that spans the evolutionary translocation breakpoint, we designed oligonucleotide probes from the human BAC clone CTD-2327L5 (HSA 5q14.1). These probes were used to screen the *Gorilla gorilla gorilla* BAC library CHORI-255 and identified six clones: 333K3, 363F1, 413A16, 419A7, 481B24, and 484N3. These gorilla genomic clones were subsequently used as FISH probes on human and gorilla lymphoblasts to confirm their chromosomal location. FISH and BAC-end sequencing of the clones showed that 413A16, 419A7, and 484N3 span the gorilla chromosome 19 breakpoint, with end sequences assigned to HSAs 5 (CTD-2327L5 and RP11-2015H6) and 17 (SMS-REP sequence). Our previous work has shown that the GGO 4 breakpoint localized within LCR17pA, near the junction of LCR17pD subunit (Fig. 1) (Stankiewicz et al. 2001). However, end sequence data from the three gorilla translocation breakpoint-spanning BAC clones placed their distal ends in DNA sequence homologous to the SMS-REP. This suggests that a complex chromosomal rearrangement event was associated with the origin of the evolutionary gorilla translocation t(4;19).

Supporting this notion, the analysis of the DNA sequences of the gorilla BAC clone 413A16 revealed the presence of two >20 Kb DNA segments homologous to chromosome 17p13.2 (RP11-46I8) and 17q23.2 (RP11-3K24; RP11-178C3), and a 229-bp fragment of the 2710-bp DNA transposon MER1-Charlie3 with an adjacent 9 bp insertion GTGGCTTGG, precisely at the translocation breakpoint (Fig. 4). These data indicate that the DNA transposon MER1-Charlie might have mediated the origin of the evolutionary gorilla translocation t(4;19). BLAST searches against human genomic sequences with the 229 bp fragment identified its paralogous sequences within all three SMS-REPs (two copies),

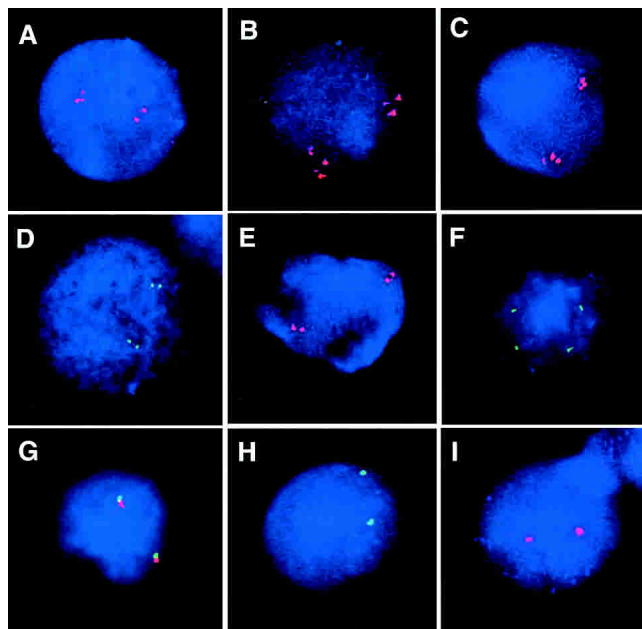


Figure 2. Evolutionary FISH analyses of LCR17ps. Interphase FISH with the LCR17pA/C- and LCR17pA/D-derived BAC clone CTD-3157E16 showed the presence of at least 2 signals on each chromosome for human (A), orangutan (B), and baboon (C). LCR17pA/B-derived clone RP11-640I15 showed four signals (green or red), two from each chromosome, in interphase nuclei of cell lines from human (D), chimpanzee (E), gorilla (F) [the separation of the signals is due to the presence of t(4;19) with LCR17pA/B and LCR17pB being present on separate derivative chromosomes]. Only one signal per chromosome homolog was observed in orangutan (G), baboon (H), and squirrel monkey (I). (G) The cohybridization of BAC clones RP11-640I15 (red) and, adjacent to it on the telomeric side, LCR-free RP11-726O12 (green) demonstrated that the LCR17pA/B was the progenitor copy (see Results).

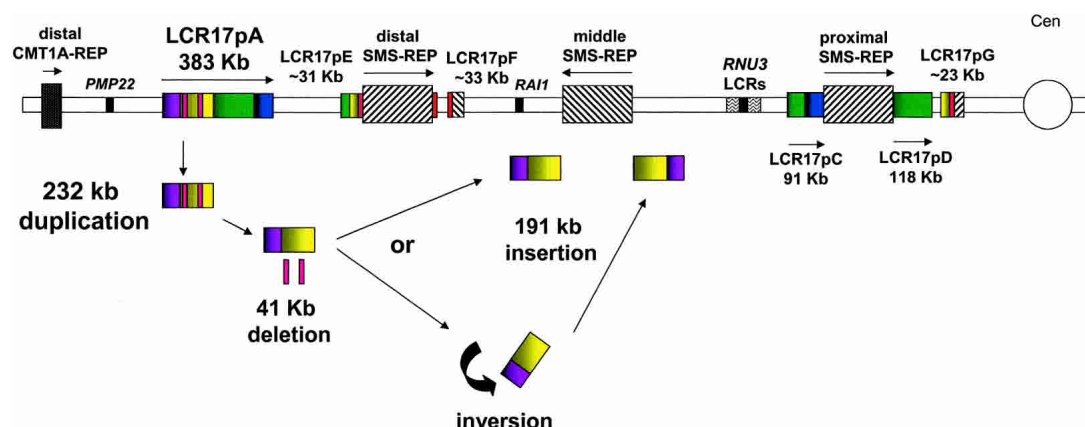


Figure 3. Insertional duplication of LCR17pB. After the divergence of gorilla and orangutan, the LCR17pA/B portion of LCR17pA was duplicated and inserted near middle SMS-REP to give LCR17pB (pink and yellow). The comparison between LCR17pA and LCR17pB copies reveals that two DNA segments (shown as red rectangles) were deleted and are not present in the LCR17pB copy. They were lost during the duplication event or after the LCR17pB insertion. The LCR17pB segment was either inserted directly adjacent to the middle SMS-REP on the telomeric side (and subsequently inverted together with the middle SMS-REP; Fig. 6) or inverted and inserted directly adjacent on its centromeric edge (see Discussion).

LCR17pE (Fig. 1), as well as in chromosome 17 BAC clones RP11-4618/AC012146 (17p13.2), RP11-3K24/AC104763 (17q23.2), chromosome 5 BAC clones RP11-810C16/AC093307 (5p15.31), RP11-167G20/AC020980 (5p15.1), CTD-210809/AC091839 (5p13.2), CTD-2022B6/AC010351 (5p13.2), CTD2317K6/AC010633 (5p13.1), CTD-2140K22/AC008829 (5q11.2), RP11-219L6/AC099519 (5q12.1-q12.2), CTD-2248H3/AC026436 (5q14.1), CTC-204P14/AC091818 (5q31.1), RP11-170L13/AC114959 (5q33.2), CTC-231O11/AC008388 (5q33.3-q34), and a few other shorter and less homologous loci on other chromosomes. Remarkably, the identified HSA 5 BAC clone CTD-2248H3 maps only ~700 Kb distal from the gorilla translocation breakpoint. Thus, in addition to RNA-mediated retrotransposon *Alu* elements (Bailey et al. 2003), these DNA transposons (considered as *sleeping beauty* elements) also may be responsible for the evolution of the primate genome. Interestingly, BLAST analysis of the gorilla BAC clone 413A16 revealed the presence of a concentration of retroviral ERVL-MER74A elements within ~4 Kb of the HSA5 syntenic segment directly adjacent to the evolutionary breakpoint.

To confirm the predicted gorilla translocation junction fragment, human, chimpanzee, gorilla, and orangutan genomic DNAs were amplified using PCR with primers designed from the sequence of the gorilla genomic BAC clone CHORI-255-413A16 to flank the identified breakpoint. As predicted, the PCR product of expected size was obtained only in gorilla and not in the genomic DNA from other primate species (Fig. 4). The DNA sequence of this PCR product matched that of the breakpoint-spanning gorilla BAC clone 413A16. Computer analysis of the breakpoint of gorilla t(4;19) revealed a sequence IMAGE:5265056 that is expressed in the human brain (hippocampus) (GenBank accession no. BC037339). Interestingly, this sequence was also identified in several LCRs including: LCR17pE, distal SMS-REP, LCR17pF, proximal SMS-REP, and LCR17pD (Fig. 4).

Complex genomic events accompany the CMT1A-REP segmental duplication

CMT1A-REP is an LCR with one copy located within LCR17pA. DNA sequence analysis of the CMT1A-REPs in humans reveals

that the distal copy, which is part of the *COX10* gene, is the progenitor copy (Murakami et al. 1997; Reiter et al. 1997). The proximal CMT1A-REP represents a duplicated copy of exon 6 of *COX10* and surrounding intronic sequences and arose after the divergence of gorilla and chimpanzee, ~3–7 Mya (Kiyosawa and Chance 1996; Boerkoel et al. 1999). The CMT1A-REP insertional/duplicational event appears to have split an ancestral gene, *AGIP* (ancestral gene before the integration of proximal CMT1A-REP), into two genes: *HREP* (*FAM18B*) expressed in heart and skeletal muscle (Kennerson et al. 1998) and *CDRT1* expressed in the pancreas and heart (Inoue et al. 2001).

Our previous comparative analysis of human CMT1A-REP genome sequence with mouse ESTs suggested a potential genomic deletion event accompanied the insertion of proximal CMT1A-REP during divergence between gorilla and chimpanzee (Inoue et al. 2001). We obtained further evidence for this proposed deletion that accompanied the evolutionary genomic rearrangement. Because the LCR17pB segmental duplication is older than the proximal CMT1A-REP integration, we hypothesized that the LCR17pB might have retained a copy of the genomic fragment that was lost in the original LCR17pA/B. Using BLAST2 sequence comparison between human LCR17pA/B and LCR17pB paralogs, at the site of insertion of proximal CMT1A-REP in LCR17pA, we identified a 4960-bp DNA fragment, present in the LCR17pB copy but absent in the LCR17pA copy (Fig. 5). This enabled a determination of a complete gene structure for human *AGIP*, including the exon that was removed from LCR17pA during CMT1A-REP segmental duplication. These findings suggest that the two serial segmental duplication events generated a duplicated copy of a gene (*AGIP*), and by exon fission, generated a new gene (*HREP*). Interestingly, BLAST analysis of the DNA sequence of this 4960-bp fragment against the human ESTs database revealed a human cDNA clone IMAGE:5170212 (GenBank accession no. BI828401), from mRNA expressed in the brain. We confirmed the brain expression of this transcript by PCR from the brain tissue cDNA library; the expected ~550-bp product was amplified (data not shown). This suggests that the duplicated copy of the ancestral *AGIP* gene (Inoue et al. 2001) may remain active in the LCR17pB copy and further supports the notion that ge-

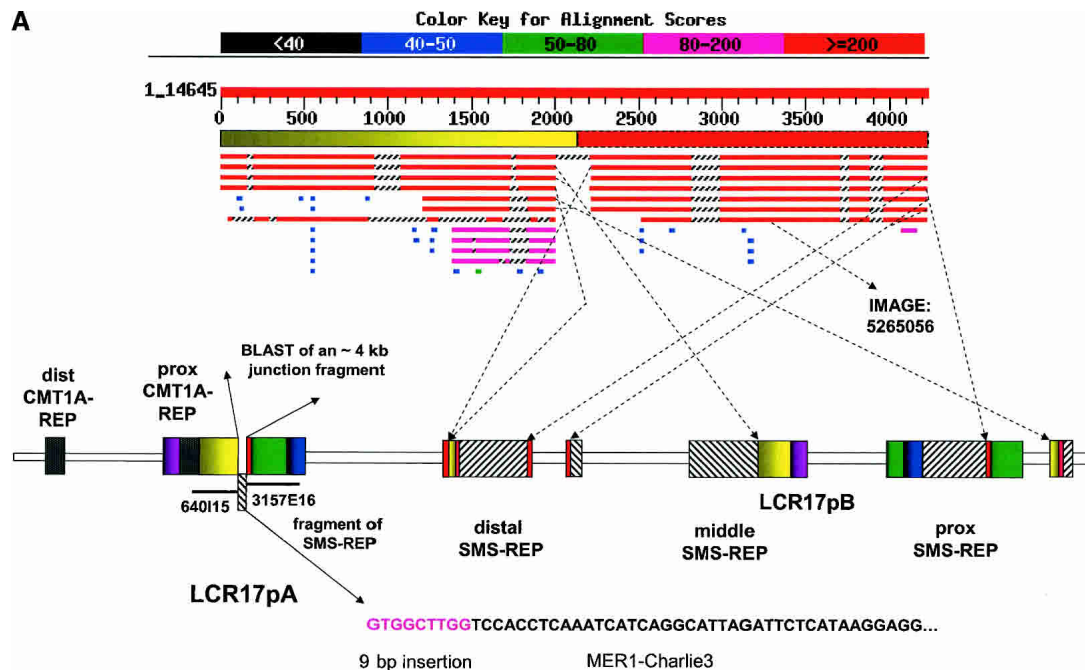


Figure 4. (Continued on next page)

nommic rearrangements may be partly responsible for primate evolution.

BLAST analysis identified a mouse BAC clone RP23-428I5 (AL645847) that is syntenic to the human LCR17pA copy. We narrowed the mouse chromosome 11 region syntenic to the insertional duplication site of the proximal CMT1A-REP to ~3.5 Kb (np. 41,500–45,034 in RP23-428I5).

To determine whether the deletion of this ~5-Kb fragment was concurrent with the proximal CMT1A-REP segmental duplication and insertional event, we designed PCR primers by comparisons between human LCR17pA and LCR17pB sequences for the analysis of the gorilla genomic substrate sequences present prior to the segmental duplication of CMT1A-REP at the proximal CMT1A-REP integration site. Using overlapping LCR17pA/B-specific oligonucleotide probes, we identified several BAC clones from the gorilla CHORI-255 genomic library. To distinguish LCR17pA/B- versus LCR17pB-derived clones, we tested them by PCR with primers specific for each copy and selected LCR17pA/B-derived gorilla BAC clones 412M18, 565H18, and 475B12. PCR amplification using the LCR17pA/B-derived gorilla BAC clones as templates resulted in a product spanning the predicted fission fragments of the 4960-bp DNA segment in LCR17pA from gorilla. DNA sequence analysis revealed the junctions at the ends of the 4960-bp DNA segment deleted in gorilla before its divergence from chimpanzee (Fig. 5). A 312-bp *AluSx* sequence was found at one end of 4960-bp DNA segment, suggesting its potential role in this indel event (Fig. 5).

A model of serial segmental duplications to evolve complex genome architecture

Cumulative data from studies of different segmental duplications in proximal 17p suggest a potential model that most parsimoniously explains how the complex genomic architecture evolved. The two copies of CMT1A-REP and three SMS-REPs are absent in rodent species (Fig. 6A) (Pentao et al. 1992; Chen et al.

1997; Probst et al. 1999; Bi et al. 2002). As an initial segmental duplication event, before the divergence of New World monkeys ~40–65 Mya, the ~151-Kb proximal portions of LCR17p, LCR17pA/C, and LCR17pA/D were duplicated and inserted, resulting in the adjacent LCR17pC (Fig. 6B) and LCR17pD (Fig. 6C) copies.

By an insertional event, a progenitor SMS-REP must have arisen in an ancient chromosome. The insertion of the progenitor proximal SMS-REP split the LCR17pC and LCR17pD copies (Fig. 6D) and the middle and distal SMS-REPs evolved by tandem duplication of the proximal SMS-REP copy (Fig. 6E,F) (Park et al. 2002). After the divergence between orangutan and gorilla ~7–12 Mya, the distal portion of LCR17pA, LCR17pA/B was tandemly duplicated and inserted directly adjacent to the middle SMS-REP on the centromeric side (Fig. 6G). Following that event, both middle SMS-REP and LCR17pB were likely inverted (Fig. 6H). Alternatively, it is possible that both middle SMS-REP and LCR17pB were generated by inverted duplications from proximal SMS-REP and LCR17pA, respectively. Because the evolutionary gorilla t(4;19) is not present in orangutan, chimpanzee, and human, we propose that it must have arisen independently, after the LCR17pB duplication (open vertical arrow on Fig. 6I). Finally, after the divergence of gorilla and chimpanzee, the distal CMT1A-REP was duplicated and resulted in the proximal copy (Fig. 6J).

Discussion

From the analysis of different genomes it has been proposed that whole-genome duplications occurred during early evolution, whereas more recent single-gene duplications have enabled the creation of novel protein functions (Li 1997). Much of what we know about duplication events has been inferred from studies of the end products viewed in a single or limited number of species. These studies have primarily consisted of computational se-

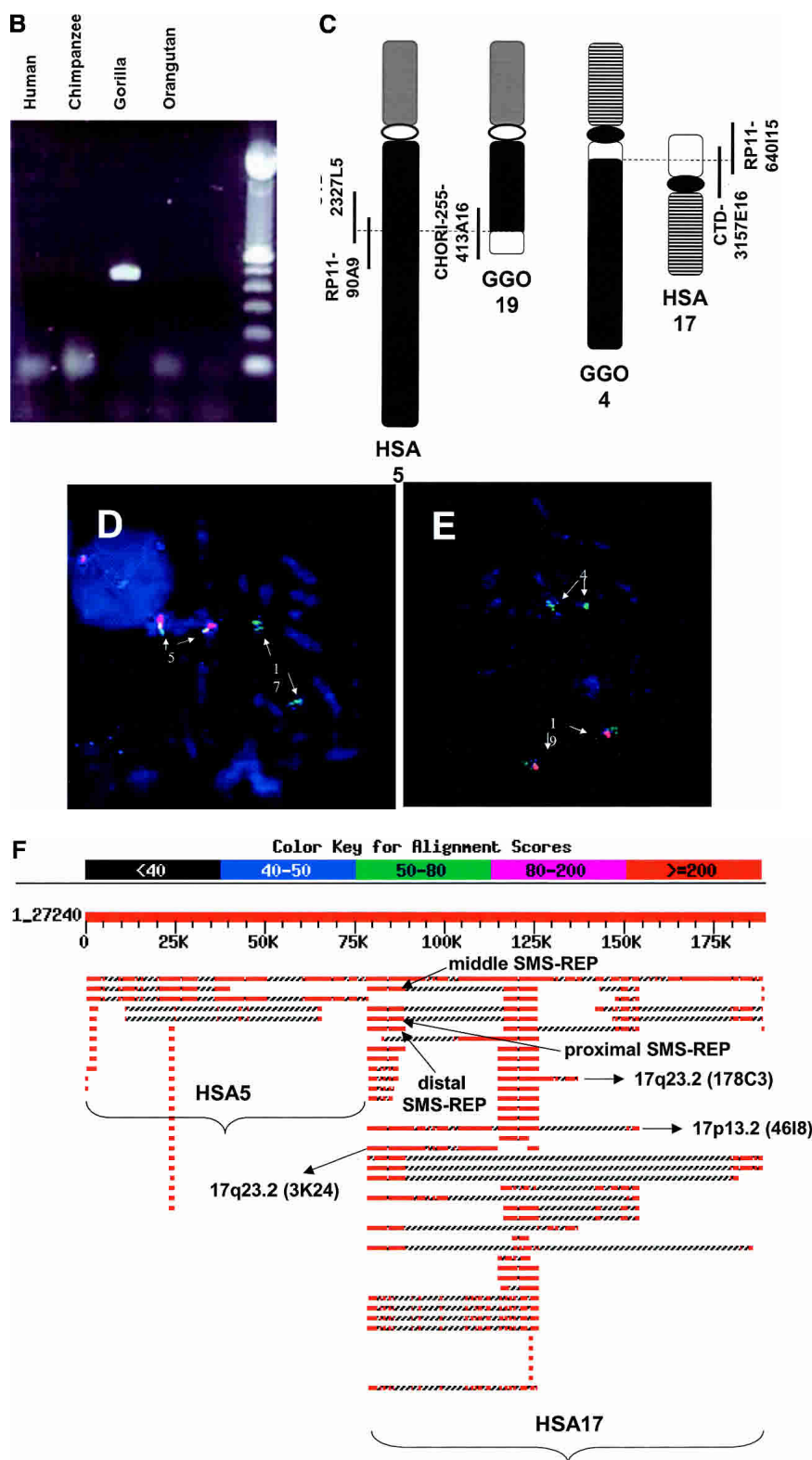


Figure 4. (Legend on next page)

quence analyses without other approaches to verify the inferences independently and experimentally. Molecular clock analysis and calculations based on sequence comparisons are con-

founded by gene conversion events occurring among repeated sequences. Therefore, a multifaceted approach was used to reconstruct the sequence of rearrangement events that result in complex genome architecture. We found that the complex genome architecture in proximal 17p changed through a series of consecutive segmental duplications during primate evolution. Both repetitive sequences and transposable elements were identified at the breakpoints of genome rearrangements, suggesting a potential role in their generation. Furthermore, the rearrangements resulted in the creation of novel genes at the junctions. DNA sequence analysis of LCRs created through segmental duplications suggests an ongoing gene conversion.

Genomic rearrangements and primate evolution

At the nucleotide sequence level, human and chimpanzee functionally important nonsynonymous sequence shows 99.4% identity (Wildman et al. 2003). It has been proposed that genomic rearrangements, rather than single nucleotide mutations/polymorphisms, through creation of novel fusion/fission genes may have been the major driving force for primate chromosome evolution (Inoue et al. 2001; Inoue and Lupski 2002; Samonte and Eichler 2002; Stankiewicz and Lupski 2002b). Early studies using comparative genomics among bacterial species revealed substantive evidence for genome rearrangements and indel events that accompany evolution (Versalovic et al. 1993; Weinstock 1994). Preliminary primate genome sequencing studies have estimated that the divergence between human and chimpanzee genomes may be as high as 5%, with the majority of differences resulting not from single-base-pair changes but from indel events (Britten 2002; Britten et al. 2003; Frazer et al. 2003).

Interestingly, molecular characterization of some primate evolutionary genomic rearrangements showed the association of chromosome breakpoints with LCRs (Nickerson 2000; Stankiewicz et al. 2001; Eder et al. 2003; Locke et al. 2003; Goidts et al. 2004; Tseng-Ayush et al. 2004). Also, comparison between human and mouse genomes revealed enrichment of segmental duplications in regions of breaks of synteny, suggesting their involvement in evolutionary rearrangements (Probst et al. 1999; Bi et al. 2002; Armengol et al. 2003; Kent et al. 2003; Pennacchio 2003; Pevzner and Tesler

2003). Recently, however, Bailey et al. (2004) proposed that the association between small-scale segmental duplication and large-scale genomic rearrangements is not causative.

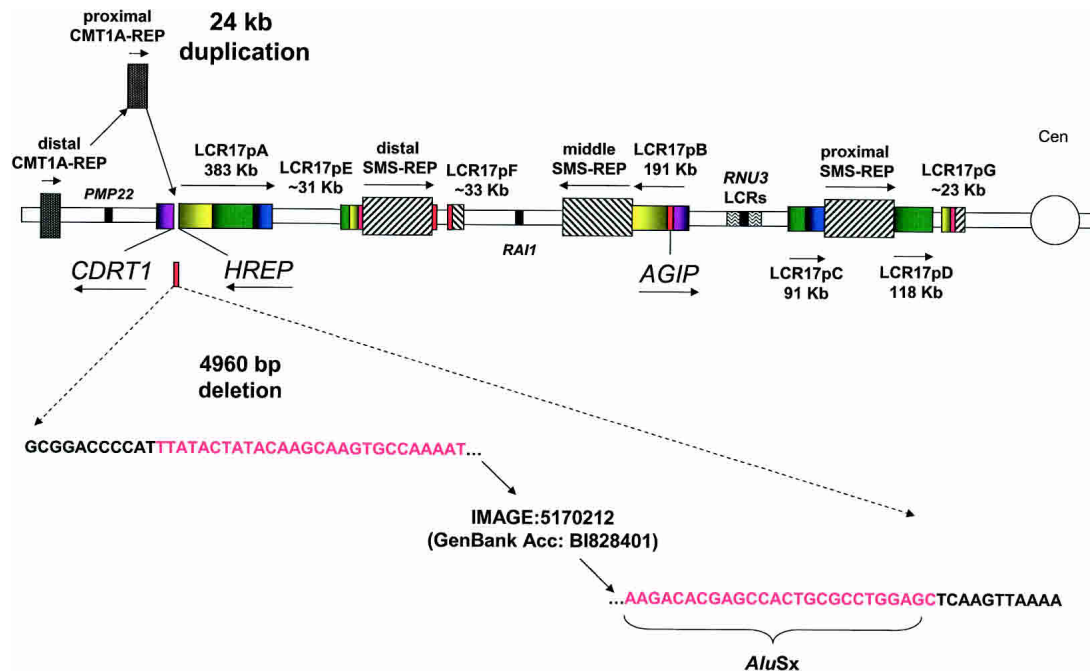


Figure 5. Schematic representation of genomic events accompanying the origin of proximal CMT1A-REP. Between divergence of gorilla and chimpanzee, ~3–7 Mya, an insertional duplication of the distal CMT1A-REP into LCR17pA resulted in a proximal CMT1A-REP copy. A 4960-bp genomic DNA fragment was deleted at the junction of the insertion site. The deleted fragment and corresponding DNA sequence are shown in red. Note that this 4960-bp fragment is still present in the human LCR17pB copy, which was duplicated earlier from LCR17pA, between orangutan and gorilla, 7–12 Mya, thus reflecting the structure of the ancestral LCR17pA copy.

Genetic recombination and segmental duplications

In a classical chromosome evolution model, because of a partial sterility of the heterozygous individuals (underdominance), chromosomal changes act as a barrier to gene flow between species (Rieseberg 2001). Recently, Navarro and Barton (2003) and Rieseberg and Livingstone (2003) proposed a novel perspective of the role of chromosomal rearrangements in primate evolution. According to this model, evolutionary genomic inversions have led to a decreased recombination rate in the involved chromosome regions that, in turn, enabled faster divergence of genes contained in such non-colinear chromosomal segments. This model provides an explanation for continuous interbreeding between separating, but not yet geographically isolated lineages. Zhang et al. (2004), however, could not confirm this hypothesis for humans and chimpanzees.

Rozen et al. (2003) and Hawley (2003) showed that intra-

chromosomal recombination activity (resulting in gene conversion events) involving large, complex, and nearly identical (99.96%) palindromic LCRs on human chromosome Y (Kuroda-Kawaguchi et al. 2001) is much more frequent than initially anticipated (Delbridge and Graves 1999). It has been thus suggested that to protect the important genetic information from evolutionary divergence, because of an inability to recombine with chromosome X, the Y chromosome contains 45% of its structure as LCR copies that serve as targets for gene conversion events.

We propose that a similar “evolutionary pressure” mechanism might have caused the selection and retention of many autosomal LCRs. The genomic regions that are now LCR rich might have been recombination poor in earlier evolutionary time (when compared with the average genome rate). Similar to LCR duplications on the Y chromosome, the “necessity” for retaining the genetic information might have placed evolutionary pressure on such regions to generate genomic “back-ups” in the

Figure 4. (A) A proposed model of the gorilla evolutionary translocation t(4;19). FISH results showed that the GGO 19 breakpoint maps between BAC clones RP11–640I15 and CTD-3157E16, at the border between LCR17pA/B and LCR17pA/D. The BLAST analysis of this junction in human (indicated with solid arrows) revealed several fission fragments scattered within proximal 17p (border between yellow and green rectangles depicted with dotted arrows), suggesting evolutionary unequal crossing-over constraint on this junction. Such genomic misalignments might have led to chromosome breakage and the origination of the evolutionary gorilla translocation t(4;19). Note the presence of a Charlie3 transposon at the edge of the SMS-REP fragment identified by sequencing the GGO 19 breakpoint and the sequence IMAGE:5265056 likely expressed in the human brain (hippocampus). (B) PCR confirmation analysis of the breakpoint of the gorilla translocation t(4;19). PCR primers spanning the t(4;19) were designed from gorilla BAC genomic clone CHORI-255–413A16 sequence. Genomic DNA from *Homo sapiens* (HSA), *Pan troglodytes* (PTR), *Gorilla gorilla* (GGO), and *Pongo pygmaeus* (PPY) was subjected to PCR amplification. A 444-bp product was obtained only in gorilla genomic DNA, indicating the translocation is unique to this genus. (C) Schematic diagram of human chromosomes 5 and 17 and gorilla translocation chromosomes 4 “der(HSA17)” and 19 “der(HSA5)”. Vertical bars represent the breakpoint spanning clones. (D,E). FISH analysis of gorilla BACs. (D) Human metaphase spread with human BAC probe RP11–90A9 (red), which hybridizes to chromosome 5 (this clone spans the evolutionary breakpoint but doesn’t show a signal on chromosome 17 because the breakpoint is very close to the end of the BAC), and gorilla BAC probe CHORI 255–413A16 (green), which hybridizes to chromosomes 5 and 17, suggesting this clone spans the t(4;19) breakpoint in gorilla. (E) Gorilla metaphase spread with human BAC probe RP11–90A9 (red), which hybridizes to chromosome 19, and gorilla BAC probe CHORI 255–413A16 (green), which hybridizes to chromosomes 4 and 19. (F) The BLAST analysis of the gorilla BAC clone 413A16 spanning the t(4;19) breakpoint revealed a complex rearrangement on the HSA17 portion of this clone. Note the presence of DNA sequences from HSA 17p11.2, 17p13.2, and 17q23.2.

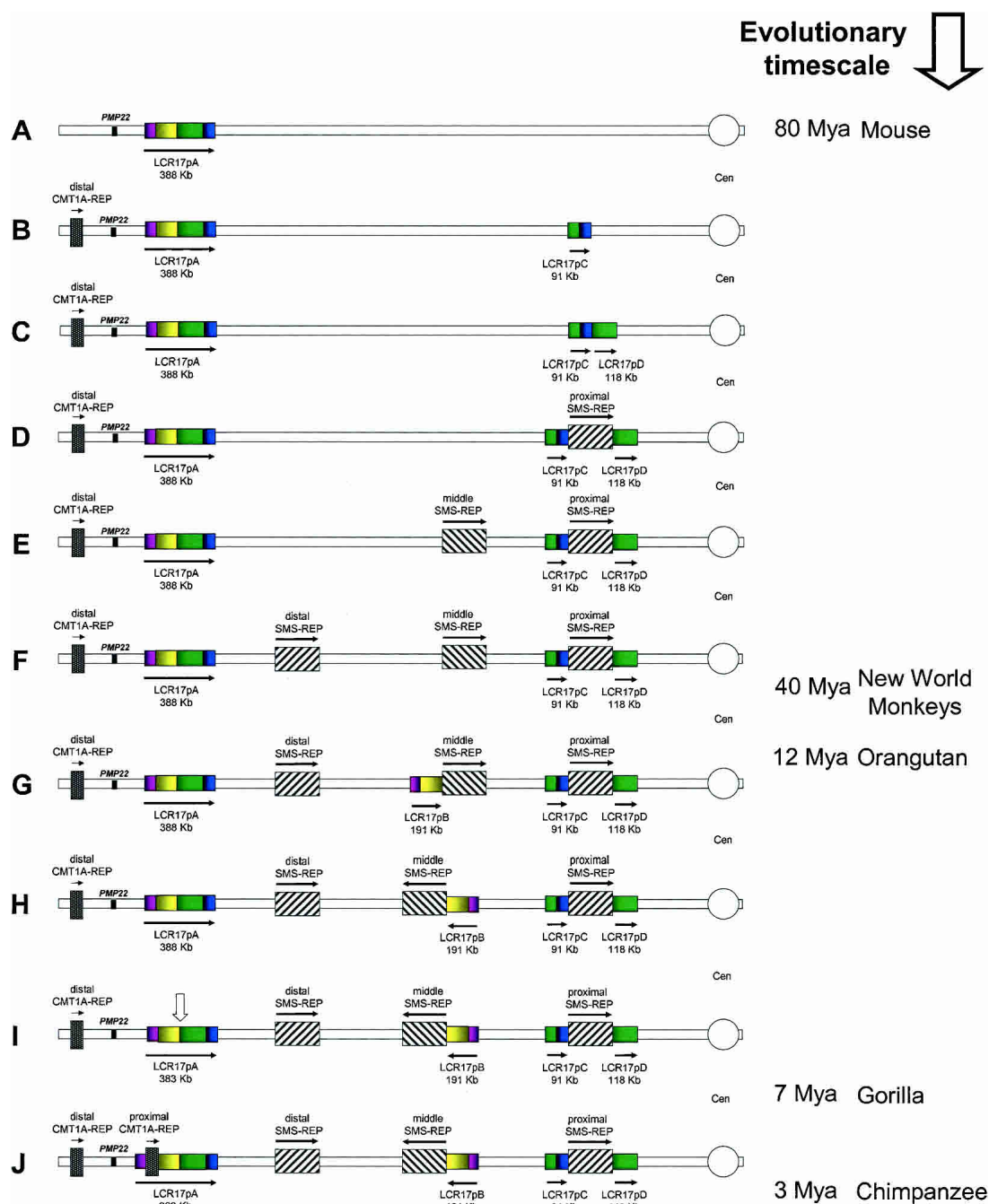


Figure 6. Rearrangements of proximal 17p LCR during primate evolution. Proximal chromosome 17p is depicted by two thin horizontal lines with the centromere (○) shown to the *right*. LCRs are shown as horizontal rectangles with the same color or black-and-white graphic, representing highly homologous sequence. (A) In the mouse genome, only LCR17pA is present. (B,C) The proximal portion of LCR17pA was duplicated >25 Mya. Note that LCR17pC and LCR17pD represent two overlapping portions of LCR17pA. (D) The proximal SMS-REP split the LCR17pC and LCR17pD copies resulting in three directly adjacent large LCRs. Two tandem duplications resulted in directly oriented middle (E) and distal (F) SMS-REPs (Park et al. 2002). (G) After the divergence of orangutan and gorilla 7–12 Mya, the distal portion of LCR17p was tandemly duplicated creating a directly oriented LCR17pB copy. (H) Both middle SMS-REP and LCR17pB, adjacent to it, were inverted. (I) Following that, at the junction between LCR17pA/B and LCR17pD copies, the evolutionary translocation t(4;19) occurred in a pre-gorilla individual, 7–12 Mya (vertical open arrow). (J) Lastly, between gorilla and chimpanzee, 3–7 Mya the proximal CMT1A-REP, present only in human and chimpanzee, resulted from the insertional duplication of the distal copy (Kiyosawa and Chance 1996; Reiter et al. 1997; Inoue et al. 2001). (Right) Time line of mammalian, mainly primate, evolution with million of years (Mya), as indicated.

form of segmental duplications. Such LCRs have then stimulated unequal crossing-over events (NAHRs). When the duplicated DNA fragment was inserted inversely with respect to the progenitor copy, the NAHR between these copies might have resulted in

chromosome inversion (Lupski 1998). Such inverted segments, in turn, served as a nest for increased evolutionary gene divergence (Navarro and Barton 2003). Thus, in addition to segmental duplications generating novel fusion/fission genes, the genomic

regions of decreased recombination might have been another source of gene evolution. Interestingly, Inoue et al. (2001) and Bi et al. (2002) reported decreased recombination over the ~1.4-Mb CMT1A and ~4-Mb SMS genomic regions in proximal 17p, potentially representing the evolutionary remnant of low evolutionary recombination activity. It will be interesting to analyze the recombination activity in other LCR-rich genomic regions.

It is remarkable that despite different sizes, locations, orientations, and most importantly times of origin, proximal and distal CMT1A-REPs; proximal, middle, and distal SMS-REPs; LCR17pA/B and LCR17pB; and LCR17pA/D and LCR17pD have retained very similar (~98.1–98.7%) nucleotide identity. Our studies show that in contrast to LCR17pA/D (~115 Kb)- and LCR17pD (~118 Kb)-mediated NAHR, resulting in recurrent SMS chromosome deletions (Shaw et al. 2004b), no deletions with breakpoint within similarly sized LCR17pA/C (~79 Kb) and LCR17pC (~91 Kb) have been found (Stankiewicz et al. 2003; Shaw et al. 2004a). We propose that during primate evolution, the DNA sequence homology between LCR17pA/C and LCR17pC copies must have dropped below a minimal misalignment/recombination stimulating threshold (<95%) that in turn resulted in a lack of LCR/NAHR gene conversion events and DNA homogenization and a subsequent steady decrease of nucleotide identity to the current ~88% identity (Table 1).

DNA rearrangements generate new genes

To investigate the evolutionary mechanism of the proximal CMT1A-REP duplication, we took advantage of the fact that this event occurred after the divergence of gorilla from chimpanzee, ~3–7 Mya (Fig. 5) (Kiyosawa and Chance 1996; Reiter et al. 1997; Boerkoel et al. 1999; Keller et al. 1999; Inoue et al. 2001). Thus, the human LCR17pB copy reflects the sequence of the progenitor pre-chimpanzee LCR17pA copy before the insertional duplication of proximal CMT1A-REP.

Interestingly, at two breakpoints of rearrangements that occurred during primate evolution, both the proximal CMT1A-REP insertion and the gorilla t(4;19) breakpoint, we found evidence for the sequences being expressed in brain, IMAGE:5170212 (Fig. 5) and IMAGE:5265056 (Fig. 4), respectively. It is tempting to speculate that the genes at the breakpoints of chromosome rearrangements may be responsible for higher-order cognitive and behavioral functions accompanying human evolution (Khaitovich et al. 2004; Marquès-Bonet et al. 2004).

Unfortunately, because of the lack of multi-tissue RNA samples from the nonhuman primates, we could not perform Northern hybridizations to further study the effects of this evolutionary genomic rearrangement. Computational analyses of human genome sequence suggest the generation of fusion/fission transcripts accompanying segmental duplication in many genomic regions (Courseaux and Nahon 2001; Edelman et al. 2001; Eichler 2001; Bailey et al. 2002). These findings illustrate the potential role of DNA rearrangements during mammalian genome evolution in the generation of new genes (Inoue et al. 2001; Inoue and Lupski 2002).

Transposable elements and large genomic rearrangements

Gray (2000) described an “alternative DNA transposition” model for the formation of reciprocal chromosome translocations; however, transposon-mediated gross genomic rearrangements have been reported only in *Drosophila* and maize (Lim and Simmons

1994; Caceres et al. 1999; Zhang and Peterson 1999). Among mammals, a recombination hotspot within a retrotransposon was identified in mice (Edelman et al. 1989), and recently, the breakpoint of the evolutionary pericentric inversion of chimpanzee chromosome 19 was found in a small single-copy segment flanked by a medium reiteration-frequency repeat MER117 and a MLT1K element, a member of the mammalian retrotransposon-like superfamily (Kehrer-Sawatzki et al. 2002).

We found that the formation of the gorilla t(4;19) was possibly mediated by the DNA transposon Charlie3 (MER1 group) on GGO4 and LTR retrotransposon ERVL-MER74A on GGO19. The vast majority of DNA sequences homologous to the MER1-Charlie3 fragment map to HSAs 5 and 17, suggesting potential colocalization of these chromosomes in cell nuclei territories. Interestingly, Tanabe et al. (2002) has shown an evolutionary conservation of chromosome territories in higher primate cell nuclei; however, no studies about HSAs 5 and 17 or their syntenic primate chromosomes have been reported to date.

Models of chromosome evolution

Recently, Nahon (2003) proposed a two-step process for the origination of “human-specific” genes during primate evolution. First, between the divergence of *Platyrrhini* and *Catarrhini*, 25–30 Mya, transposition events mediated by RNA and DNA (gene nurseries, genic tinkering) resulted in gene duplications. In a second step, at the time of divergence of *Hominidae*, 5–10 Mya, intra-chromosomal duplication events were occurring. Our data imply, however, a more complex mechanistic model.

Intriguingly, computer analyses of genomic sequence data show that despite different times of formation, most of the analyzed large LCRs in proximal 17p are directly adjacent to one another. Many of the analyzed junction fragments revealed the presence of the same DNA segment (red-yellow junction on Fig. 4A). These data imply that the primary or secondary DNA structure found at edges/junction of LCRs, may be particularly prone to both segmental duplications and large genomic rearrangements.

Serial segmental duplication model for the origin of complex genomic architecture

By analyzing the genome architecture of the primate genomic regions syntenic to the human proximal chromosome 17p, we derived a multi-step model for reconstruction of the evolution of the ~7.5-Mb LCR-rich genome architecture (Fig. 6) (Lupski 2003).

In summary, our molecular and computational analyses of the ~7.5-Mb proximal 17p region revealed that complex genome architecture has evolved in a multi-step manner involving serial segmental duplications. Segmental duplications appear to influence genome evolution by a number of different mechanisms, including the creation of novel fusion/fission genes at the LCR insertion site, as well as potentially enabling an increased mutation rate because of LCR-mediated genomic inversion associated with reduced recombination. In essence, such genome architecture enables both shuffling of the genome and generating new gene functions providing ample diversity for selection to act on during evolution.

Methods

DNA sequence analysis

The analyzed clones were identified from the existing physical maps of this region (Inoue et al. 2001; Bi et al. 2002; <http://>

www.ncbi.nlm.nih.gov/; <http://genome.ucsc.edu/>). Sequences were downloaded from the NCBI web site (<http://www.ncbi.nlm.nih.gov/>). Regional assemblies for known LCR17ps and searches for additional LCRs were constructed following NCBI BLAST searches against the high-throughput and the nonredundant sequence database (<http://www.ncbi.nlm.nih.gov/blast/>) and were assembled using the Sequencher software (Gene Codes) and NCBI BLAST2.

Molecular clock analysis was performed as described elsewhere (Li 1997). In short, we determined the sequence homology between each unit of LCRs using BLAST2 and RepeatMasker. The sequence identity between unique segments and those between repetitive segments between paralogs were not significantly different, thus the use of the RepeatMasker did not affect the overall sequence homology calculation. We eliminated small indels from the calculation and took only base pair changes into account.

Cell lines

The human and primate lymphoblastoid cell lines and fibroblasts were grown and harvested as described previously (Park et al. 2002).

Fluorescence in situ hybridization

The selected BAC and PAC clones were purchased from the BACPAC Resource Center. DNA was isolated from liquid cultures using PSI ϕ Clone BAC DNA Kit (Princeton Separations, Inc.). FISH was performed on metaphase and interphase cells of peripheral blood lymphocytes, Epstein-Barr virus-transformed peripheral blood lymphoblasts, and skin fibroblasts as described (Shaffer et al. 1997). Chromosomal duplication was distinguished from replication with a control probe consisting of a unique sequence BAC clone from the same chromosome region labeled with different fluorochrome.

Screening the gorilla BAC library

We screened the gorilla BAC library CHORI255 (BACPAC Resource Center, The Children's Hospital Oakland Research Institute, Oakland, CA) using "overgo" oligonucleotides from the LCR17pA (subunit B)-derived BAC clone RP11-640I15: F:CAATTTCTTCAATGGGAAGTATGA, R:GCCATTGGCTTTTATCACCTTCATACAG; F:CCATGGCCAGTTGGTAAAGGTA, R:AGCCCACTGCCCACCTTACCTTTA; F:GCTCTCTAATGAAGAGAGAACATGGATGGTC, R:ATCCAATCACTTCAAGACCATCC; CTD-3157E16 (subunit D): F:AGTGAACCTGTAAACA TACTAACCTCC, R:TATGGTAATGTAAAAAGGCAGGAGGGTT; F:AGATAACAGGAGCAGTTTGTAGAACAAA, R:TTTAGATCCTTCATTATAGTTTGTCT; F:GGATCCTTTTCTTTCTCCAAC TGGGAGATT, R:TTTCCAAACCAGACACAATCTCCC; and CTD 2327L5: F:TATCACTGTTTGTTCAGGGGGATG, R:GCCAGCCA ATTATGGCCCATCCCC; F:TTTAAAAGTCAGGGGCACACTTCT GCT, R:GAACCATGCAGTCTCCCAAGCAGAAG.

Hybridizations were performed according to the methods described previously (www.chori.org/bacpac; Cai et al. 1998).

DNA and whole BAC clone sequencing

The PCR products were sequenced in the MRRC DNA sequencing core at Baylor College of Medicine.

The entire gorilla BAC clone CHORI-255-413A16/AC146897 was sequenced at the Whitehead Institute for Biomedical Research/MIT, Center for Genome Research, Cambridge, MA.

PCR

PCR was performed with extracted BAC DNA using each of the following primer sets: Distal junction F: TCCCTATTCTCCTTG CAAACTC, R: TCATGACTGGGATGTCCCTT; proximal junction F: CTGCAGTTACCATCCCACCT, R: GGTGTCTGTGTGGAG GGTCT. PCR reactions were performed as described (Shaw et al. 2002). For long-range PCR identifying the LCR17pA-derived clones, the following primers were used: F: GTCAAAATCTTC CCTATTCTCCTTGCAAACCTC, R: GACTGTGCTTTATATATGCG TAGACTCAGGGT. The products were visualized on 2% agarose gels.

To confirm the computational results, the DNA junction fragment of the gorilla translocation breakpoint was amplified (F: TAGATGCCTGATGTGGCTTG, R: TGGATGCCTCTATGTGC TGT) and the PCR product was sequenced.

Brain expression of the identified human cDNA clone IMAGE:5170212 (GenBank accession no. BI828401) was validated using PCR with primers: F: TGGCAAGATCCGAATTTACA, R: GCGCTAACAGTCAGGAGGAA on a human multi-tissue cDNA library panel (BD Biosciences).

Acknowledgments

We appreciate the critical reviews of Drs. B. Morrow and S. Rosenberg. We thank Dr. B. Birren and the Whitehead Institute for genomic sequence of the gorilla BAC clone 413A16. This study was supported in part by grants from the Muscular Dystrophy Association, the National Institute of Neurological Disorders and Stroke (R01 NS27042), and the National Institute of Child Health and Human Development (PO1 HD39420).

References

- Armengol, L., Pujana, M.A., Cheung, J., Scherer, S.W., and Estivill, X. 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.
- Babcock, M., Pavlicek, A., Spiteri, E., Kashork, C.D., Ioshikhes, I., Shaffer, L.G., Jurka, J., and Morrow, B.E. 2003. Shuffling of genes within low-copy repeats on 22q11 (LCR22) by *Alu*-mediated recombination events during evolution. *Genome Res.* **13**: 2519–2532.
- Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J., and Eichler, E.E. 2001. Segmental duplications: Organization and impact within the current human genome project assembly. *Genome Res.* **11**: 1005–1017.
- Bailey, J.A., Yavor, A.M., Viggiano, L., Misceo, D., Horvath, J.E., Archidiacono, N., Schwartz, S., Rocchi, M., and Eichler, E.E. 2002. Human-specific duplication and mosaic transcripts: The recent paralogous structure of chromosome 22. *Am. J. Hum. Genet.* **70**: 83–100.
- Bailey, J.A., Liu, G., and Eichler, E.E. 2003. An *Alu* transposition model for the origin and expansion of human segmental duplications. *Am. J. Hum. Genet.* **73**: 823–834.
- Bailey, J.A., Baertsch, R., Kent, W.J., Haussler, D., and Eichler, E.E. 2004. Hotspots of mammalian chromosomal evolution. *Genome Biol.* **5**: R23.
- Barboudi, A., Stankiewicz, P., Nusbaum, C., Cuomo, C., Cook, A., Hoglund, M., Johansson, B., Hagemeijer, A., Park, S.-S., Mitelman, F., et al. 2004. The breakpoint region of the most common isochromosome, i(17q), in human neoplasia is characterized by a complex genomic architecture with large, palindromic, low-copy repeats. *Am. J. Hum. Genet.* **74**: 1–10.
- Bi, W., Yan, J., Stankiewicz, P., Park, S.-S., Walz, K., Boerkoel, C.F., Potocki, L., Shaffer, L.G., Devriendt, K., Nowaczyk, M.J.M., et al. 2002. Genes in a refined Smith-Magenis syndrome critical deletion interval on chromosome 17p11.2 and the syntenic region of the mouse. *Genome Res.* **12**: 713–728.
- Bi, W., Park, S.-S., Shaw, C.J., Withers, M.A., Patel, P.I., and Lupski, J.R. 2003. Reciprocal crossovers and a positional preference for strand exchange in recombination events resulting in deletion or duplication of chromosome 17p11.2. *Am. J. Hum. Genet.*

- 73:** 1302–1315.
- Boerkoel, C.F., Inoue, K., Reiter, L.T., Warner, L.E., and Lupski, J.R. 1999. Molecular mechanisms for CMT1A duplication and HNPP deletion. *Ann. N.Y. Acad. Sci.* **883**: 22–35.
- Britten, R.J. 2002. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc. Natl. Acad. Sci.* **99**: 13633–13635.
- Britten, R.J., Rowen, L., Williams, J., and Cameron, R.A. 2003. Majority of divergence between closely related DNA samples is due to indels. *Proc. Natl. Acad. Sci.* **100**: 4661–4665.
- Caceres, M., Ranz, J.M., Barbado, A., Long, M., and Ruiz, A. 1999. Generation of a widespread *Drosophila* inversion by a transposable element. *Science* **285**: 415–418.
- Cai, W.W., Reneker, J., Chow, C.-W., Vaishnav, M., and Bradley, A. 1998. An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. *Genomics* **54**: 387–397.
- Chance, P.F., Abbas, N., Lensch, M.W., Pentao, L., Roa, B.B., Patel, P.I., and Lupski, J.R. 1994. Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. *Hum. Mol. Genet.* **3**: 223–228.
- Chen, K.-S., Manian, P., Koeuth, T., Potocki, L., Zhao, Q., Chinault, A.C., Lee, C.C., and Lupski, J.R. 1997. Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nat. Genet.* **17**: 154–163.
- Cheung, V.G., Nowak, N., Jang, W., Kirsch, I.R., Zhao, S., Chen, X.-N., Furey, T.S., Kim, U.-J., Kuo, W.-L., Olivier, M. et al. 2001. Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* **409**: 953–958.
- Courseaux, A. and Nahon, J.-L. 2001. Birth of two chimeric genes in the *Hominidae* lineage. *Science* **291**: 1293–1297.
- Courseaux, A., Richard, F., Grosgeorge, J., Ortolà, C., Viale, A., Turc-Carel, C., Dutrillaux, B., Gaudray, P., and Nahon, J.-L. 2003. Segmental duplications in euchromatic regions of human chromosome 5: A source of evolutionary instability and transcriptional innovation. *Genome Res.* **13**: 369–381.
- Delbridge, M.L. and Graves, J.A.M. 1999. Mammalian Y chromosome evolution and the male-specific functions of Y chromosome-borne genes. *Rev. Reprod.* **4**: 101–109.
- Edelmann, W., Kroger, B., Goller, M., and Horak, I. 1989. A recombination hotspot in the LTR of a mouse retrotransposon identified in an in vitro system. *Cell* **57**: 937–946.
- Edelmann, L., Stankiewicz, P., Spiteri, E., Pandita, R.K., Shaffer, L., Lupski, J.R., and Morrow, B.E. 2001. Two functional copies of the *DGCR6* gene are present on human chromosome 22q11 due to a duplication of an ancestral locus. *Genome Res.* **11**: 208–217.
- Eder, V., Ventura, M., Ianigro, M., Teti, M., Rocchi, M., and Archidiacono, N. 2003. Chromosome 6 phylogeny in primates and centromere repositioning. *Mol. Biol. Evol.* **20**: 1506–1512.
- Eichler, E.E. 2001. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet.* **17**: 661–669.
- Eichler, E.E., Clark, R.A., and She, X. 2004. An assessment of the sequence gaps: unfinished business in a finished human genome. *Nat. Rev. Genet.* **5**: 345–354.
- Emanuel, B.S. and Shaikh, T.H. 2001. Segmental duplications: An “expanding” role in genomic instability and disease. *Nat. Rev. Genet.* **2**: 791–800.
- Frazer, K.A., Chen, X., Hinds, D.A., Pant, P.V.K., Patil, N., and Cox, D.R. 2003. Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates. *Genome Res.* **13**: 341–346.
- Goidts, V., Szamalek, J.M., Hameister, H., Kehrer-Sawatzki, H. 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.
- Gray, Y.H.M. 2000. It takes two transposons to tango: Transposable-element-mediated chromosomal rearrangements. *Trends Genet.* **16**: 461–468.
- Hawley, R.S. 2003. The human Y chromosome: Rumors of its death have been greatly exaggerated. *Cell* **113**: 825–828.
- Hurles, M.E. and Jobling, M.A. 2003. A singular chromosome. *Nat. Genet.* **34**: 246.
- Inoue, K. and Lupski, J.R. 2002. Molecular mechanisms for genomic disorders. *Annu. Rev. Genomics Hum. Genet.* **3**: 199–242.
- Inoue, K., Dewar, K., Katsanis, N., Reiter, L.T., Lander, E.S., Devon, K.L., Wyman, D.W., Lupski, J.R., and Birren, B. 2001. The 1.4-Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes. *Genome Res.* **11**: 1018–1033.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Kehrer-Sawatzki, H., Schreiner, B., Tanzer, S., Platzer, M., Müller, S., and Hameister, H. 2002. Molecular characterization of the pericentric inversion that causes differences between chimpanzee chromosome 19 and human chromosome 17. *Am. J. Hum. Genet.* **71**: 375–388.
- Keller, M.P., Seifried, B.A., and Chance, P.F. 1999. Molecular evolution of the CMT1A-REP region: A human- and chimpanzee-specific repeat. *Mol. Biol. Evol.* **16**: 1019–1026.
- Kennerson, M.L., Nassif, N.T., and Nicholson, G.A. 1998. Genomic structure and physical mapping of *C17orf1*: A gene associated with the proximal element of the CMT1A-REP binary repeat. *Genomics* **53**: 110–112.
- Kent, W.J., Baertsch, R., Hinrichs, A., Miller, W., and Haussler, D. 2003. Evolution’s cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl. Acad. Sci.* **100**: 11484–11489.
- Khaitovich, P., Muetzel, B., She, X., Lachmann, M., Hellmann, I., Dietzsch, J., Steigle, S., Do, H.-H., Weiss, G., Enard, T.M., et al. 2004. Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* **14**: 1462–1473.
- Kiyosawa, H. and Chance, P.F. 1996. Primate origin of the CMT1A-REP repeat and analysis of a putative transposon-associated recombinational hotspot. *Hum. Mol. Genet.* **5**: 745–753.
- Kumar, S. and Hedges, S.B. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**: 917–920.
- Kuroda-Kawaguchi, T., Skaletsky, H., Brown, L.G., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Silber, S., Oates, R., Rozen, S., et al. 2001. The *AZFc* region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat. Genet.* **29**: 279–286.
- Li, W.-H. 1997. Molecular clocks. In *Molecular evolution*, pp. 215–235. Sinauer Associates, Inc. Sunderland, MA.
- Lim, J.K. and Simmons, M.J. 1994. Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*. *Bioessays* **16**: 269–275.
- Locke, D.P., Archidiacono, N., Misceo, D., Cardone, M.F., Deschamps, S., Roe, B., Rocchi, M., and Eichler, E.E. 2003. Refinement of a chimpanzee pericentric inversion breakpoint to a segmental duplication cluster. *Genome Biol.* **4**: R50.
- Lupski, J.R. 1998. Genomic disorders: Structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* **14**: 417–422.
- . 2003. Genomic disorders: Recombination-based disease resulting from genome architecture. *Am. J. Hum. Genet.* **72**: 246–252.
- Marquès-Bonet, T., Cáceres, M., Bertranpetit, J., Preuss, T.M., Thomas, J.W., and Navarro, A. 2004. Chromosomal rearrangements and the genomic distribution of gene-expression divergence in humans and chimpanzees. *Trends in Genet.* (in press).
- Murakami, T., Reiter, L.T., and Lupski, J.R. 1997. Genomic structure and expression of the human heme A:farnesyltransferase (*COX10*) gene. *Genomics* **42**: 161–164.
- Nahon, J.-L. 2003. Birth of ‘human-specific’ genes during primate evolution. *Genetica* **118**: 193–208.
- Navarro, A. and Barton, N.H. 2003. Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes. *Science* **300**: 321–324.
- Nickerson, E. 2000. Molecular definition of pericentric inversions distinguishing the genomes of humans and chimpanzees since divergence from a common ancestor. Ph.D. dissertation, Baylor College of Medicine, Houston, TX.
- Park, S.-S., Stankiewicz, P., Bi, W., Shaw, C., Lehoczy, J., Dewar, K., Birren, B., and Lupski, J.R. 2002. Structure and evolution of the Smith-Magenis syndrome repeat gene clusters, SMS-REPs. *Genome Res.* **12**: 729–738.
- Pennacchio, L.A. 2003. Insights from human/mouse genome comparisons. *Mamm. Genome* **14**: 429–436.
- Pentao, L., Wise, C.A., Chinault, A.C., Patel, P.I., and Lupski, J.R. 1992. Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nat. Genet.* **2**: 292–300.
- Pevzner, P. and Tesler, G. 2003. Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proc. Natl. Acad. Sci.* **100**: 7672–7677.
- Potocki, L., Chen, K.-S., Park, S.-S., Osterholm, D.E., Withers, M.A., Kimonis, V., Summers, A.M., Meschino, W.S., Anyane-Yeboah, K., Kashork, C.D., et al. 2000. Molecular mechanism for duplication 17p11.2- the homologous recombination reciprocal of the Smith-Magenis microdeletion. *Nat. Genet.* **24**: 84–87.
- Probst, F.J., Chen, K.-S., Zhao, Q., Wang, A., Friedman, T.B., Lupski, J.R.,

- and Camper, S.A. 1999. A physical map of the mouse *shaker-2* region contains many of the genes commonly deleted in Smith-Magenis syndrome (del17p11.2p11.2). *Genomics* **55**: 348–352.
- Reiter, L.T., Murakami, T., Koeuth, T., Gibbs, R.A., and Lupski, J.R. 1997. The human *COX10* gene is disrupted during homologous recombination between the 24 kb proximal and distal CMT1A-REPs. *Hum. Mol. Genet.* **6**: 1595–1603.
- Reiter, L.T., Hastings, P.J., Nelis, E., De Jonghe, P., Van Broeckhoven, C., and Lupski, J.R. 1998. Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. *Am. J. Hum. Genet.* **62**: 1023–1033.
- Rieseberg, L.H. 2001. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**: 351–358.
- Rieseberg, L.H. and Livingstone, K. 2003. Chromosomal speciation in primates. *Science* **300**: 267–268.
- Rozen, S., Skaletsky, H., Marszalek, J.D., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., and Page, D.C. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**: 873–876.
- Samonte, R.V. and Eichler, E.E. 2002. Segmental duplications and the evolution of the primate genome. *Nat. Rev. Genet.* **3**: 65–72.
- Shaffer, L.G., Kennedy, G.M., Spikes, A.S., and Lupski, J.R. 1997. Diagnosis of CMT1A duplications and HNPP deletions by interphase FISH: Implications for testing in the cytogenetics laboratory. *Am. J. Med. Genet.* **69**: 325–331.
- Shaw, C.J. and Lupski, J.R. 2004. Implications of human genome architecture for rearrangement based disorders: The genomic basis of disease. *Hum. Mol. Genet.* **13**: R57–R64.
- Shaw, C.J., Bi, W., and Lupski, J.R. 2002. Genetic proof of unequal meiotic crossovers in reciprocal deletion and duplication of 17p11.2. *Am. J. Hum. Genet.* **71**: 1072–1081.
- Shaw, C.J., Shaw, C.A., Yu, W., Stankiewicz, P., White, L.D., Beaudet, A.L., and Lupski, J.R. 2004a. Comparative genomic hybridisation using a proximal 17p BAC/PAC array detects rearrangements responsible for four genomic disorders. *J. Med. Genet.* **41**: 113–119.
- Shaw, C.J., Withers, M.A., and Lupski, J.R. 2004b. Uncommon Smith-Magenis syndrome deletions can be recurrent by utilizing alternate LCRs as homologous recombination substrates. *Am. J. Hum. Genet.* **75**: 75–81.
- Stankiewicz, P. and Lupski, J.R. 2002a. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**: 74–82.
- . 2002b. Molecular-evolutionary mechanisms for genomic disorders. *Curr. Opin. Genet. Dev.* **12**: 312–319.
- Stankiewicz, P., Park, S.-S., Inoue, K., and Lupski, J.R. 2001. The evolutionary chromosome translocation 4;19 in *Gorilla gorilla* is associated with microduplication of the chromosome fragment syntenic to sequences surrounding the human proximal CMT1A-REP. *Genome Res.* **11**: 1205–1210.
- Stankiewicz, P., Shaw, C.J., Dapper, J.D., Wakui, K., Shaffer, L.G., Withers, M., Elizondo, L., Park, S.-S., and Lupski, J.R. 2003. Genome architecture catalyzes nonrecurrent chromosomal rearrangements. *Am. J. Hum. Genet.* **72**: 1101–1116.
- Tanabe, H., Müller, S., Neusser, M., von Hase, J., Calcagno, E., Cremer, M., Solovei, I., Cremer, C., and Cremer, T. 2002. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc. Natl. Acad. Sci.* **99**: 4424–4429.
- Tsend-Ayush, E., Grützner, F., Yue, Y., Grossmann, B., Hansel, U., Sudbrak, R., and Haaf, T. 2004. Plasticity of human chromosome 3 during primate evolution. *Genomics* **83**: 193–202.
- van Geel, M., Eichler, E.E., Beck, A.F., Shan, Z., Haaf, T., van der Maarel, S.M., Frants, R.R., and de Jong, P.J. 2002. A cascade of complex subtelomeric duplications during the evolution of the hominoid and Old World monkey genomes. *Am J Hum. Genet.* **70**: 269–278.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. The sequence of the human genome. *Science* **291**: 1304–1351.
- Versalovic, J., Koeuth, T., Britton, R., Geszvain, K., and Lupski, J.R. 1993. Conservation and evolution of the *rpsU-dnaG-rpoD* macromolecular synthesis operon in bacteria. *Mol. Microbiol.* **8**: 343–355.
- Weinstock, G.M. 1994. Bacterial genomes: Mapping and stability. *ASM News* **60**: 73–78.
- Wildman, D.E., Uddin, M., Liu, G., Grossman, L.I., and Goodman, M. 2003. Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: Enlarging genus *Homo*. *Proc. Natl. Acad. Sci.* **100**: 7181–7188.
- Zhang, J. and Peterson, T. 1999. Genome rearrangements by nonlinear transposons in maize. *Genetics* **153**: 1403–1410.
- Zhang, J., Wang, X., and Podlaha, O. 2004. Testing the chromosomal speciation hypothesis for humans and chimpanzees. *Genome Res.* **14**: 845–851.

Web site references

- www.chori.org/bacpac; Children's Hospital Oakland—BAC-PAC Resources.
- <http://www.ncbi.nlm.nih.gov/>; National Center for Biotechnology Information (NCBI).
- <http://genome.ucsc.edu/>; The UCSC genome bioinformatics site.

Received May 3, 2004; accepted in revised form August 14, 2004.



Serial segmental duplications during primate evolution result in complex human genome architecture

Pawel Stankiewicz, Christine J. Shaw, Marjorie Withers, et al.

Genome Res. 2004 14: 2209-2220

Access the most recent version at doi:[10.1101/gr.2746604](https://doi.org/10.1101/gr.2746604)

References

This article cites 77 articles, 25 of which can be accessed free at:
<http://genome.cshlp.org/content/14/11/2209.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>
