Hominoid lineage specific amplification of low-copy repeats on 22q11.2 (LCR22s) associated with velo-cardio-facial/digeorge syndrome

Melanie Babcock¹, Svetlana Yatsenko², Janet Hopkins³, Matthew Brenton³, Qing Cao⁴, Pieter de Jong⁴, Pawel Stankiewicz², James R. Lupski², James M Sikela³ and Bernice E. Morrow¹,*

¹Department of Molecular Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA, ²Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA, ³Department of Pharmacology and Human Medical Genetics Program, University of Colorado Health Sciences Center, Campus Box 8303, Denver, CO 80217, USA and ⁴Children’s Hospital of Oakland Research Institute, 747 52nd St, Oakland, CA 94609, USA

Received April 6, 2007; Revised and Accepted July 16, 2007

Segmental duplications or low-copy repeats (LCRs) constitute ~5% of the sequenced portion of the human genome and are associated with many human congenital anomaly disorders. The low-copy repeats on chromosome 22q11.2 (LCR22s) mediate chromosomal rearrangements resulting in deletions, duplications and translocations. The evolutionary mechanisms leading to LCR22 formation is unknown. Four genes, USP18, BCR, GGTLA and GGT, map adjacent to the LCR22s and pseudogene copies are located within them. It has been hypothesized that gene duplication occurred during primate evolution, followed by recombin- bination events, forming pseudogene copies. We investigated whether gene duplication could be detected in non-human hominoid species. FISH mapping was performed using probes to the four functional gene loci. There was evidence for a single copy in humans but additional copies in hominoid species. We then compared LCR22 copy number using LCR22 FISH probes. Lineage specific LCR22 variation was detected in the hominoid species supporting the hypothesis. To independently validate initial findings, real time PCR, and screening of gorilla BAC library filters were performed. This was compared to array comparative genome hybridization data available. The most striking finding was a dramatic amplification of LCR22s in the gorilla. The LCR22s localized to the telomeric or subtelomeric bands of gorilla chromosomes. The most parsimonious explanation is that the LCR22s became amplified by inter-chromosomal recombination between telomeric bands. In summary, our results are consistent with a lineage specific coupling between gene and LCR22 duplication events. The LCR22s thus serve as an important model for evolution of genome variation.

INTRODUCTION

It has been estimated that ~5% of the sequenced human genome is comprised of segmental duplications or low-copy repeats (LCRs), that range in size from 10 to 250 kb with >95% sequence identity (1). They represent mosaic structures consisting of genes, gene segments, pseudogenes and repetitive sequence elements (Fig. 1) (2). In pericentromeric or subtelomeric intervals, it has been estimated that 11.6–45% of the genome sequence is comprised of LCRs (2–5). A significant subset is associated with human genomic disorders as mediated by non-allelic homologous recombination (NAHR) events between them (Reviewed in 2,6,7). In addition to deletion disorders, reciprocal products of meiotic inter-chromosomal
recombination events can lead to a duplication of the same interval that is deleted leading to newly recognized disorders (8–12).

Chromosome 22 contains their own LCRs, termed LCR22, comprising roughly 11% of the 22q11.2 interval. Meiotic inter-chromosomal recombination events involving predominantly two of them, LCR22-2 and LCR22–4 [also known as LCR22-A and D; (13)] lead to both deletions and reciprocal duplications, including velocardiofacial syndrome/DiGeorge syndrome (VCFS/DGS OMIM 192430/OMIM 188400; 22q11 deletion syndrome, 22q11DS), dup(22)(q11.2;q11.2) syndrome and cat-eye syndrome (CES OMIM 114570; 14). The LCR22s are a complex mosaic of genes and pseudogenes, which formed partially by Alu-mediated recombination events during primate evolution (4). Four functional genes, USP18, BCR, GGT and GGTLA, reside immediately adjacent to LCR22-2, 6, 7 and 8, respectively (4,20,21), and pseudogenes lie within the LCR22s.

Although the mouse genome does not contain LCR22s (22), they have been present in all primate species examined (13). In fact, LCR22s predate the divergence of New World from Old World monkeys (13). To further delineate the organization of hominoid LCR22s, we focused our study on the genomic loci harboring the four genes and their pseudogene copies within. FISH mapping, real time PCR and screening of gorilla BAC library filters were performed. This was compared to array comparative genome hybridization (aCGH) data (23). When taken together, we found evidence for lineage specific evolution of LCR22s. Our findings underscore the instability and dynamic nature of LCRs in the genome and provide a model to study genome evolution and plasticity.

RESULTS

Duplications of genes on 22q11.2

Four functional genes, BCR, GGT, GGTLA and USP18, flank the LCR22s on 22q11.2 (Fig. 1). In addition, the LCR22s contain partial copies of each of the four genes, forming a complex mosaic (Fig. 1). Furthermore, part of the IGSF3 genomic locus from 1p13.1 had become duplicated during evolution and transposed to LCR22-2 and LCR22-4 (4). Previous work, comparing the structure of the functional genes and the pseudogene copies in the human LCR22s, has suggested that the formation of the LCR22s has occurred via gene duplication and Alu-mediated recombination events (4). The duplicated sequences then might act as substrates for additional NAHR events (4). To ascertain whether duplication of the genomic locus could be detected, FISH mapping was performed using human BAC clones containing the functional, full-length gene and non-LCR22 genomic sequences, avoiding parts comprising LCR22 repeats (Figs 2 and 3). The order in which experimental findings are presented is BCR, GGT, GGTLA and USP18, in order of increasing complexity.

The BCR genomic locus comprises a region of 135 586 bp (chr22:21,847,105-21,982,690; May 2004 Assembly), mapping adjacent to LCR22-6 (Fig. 1). We chose the BAC clone, CTD-3139116, which contained the functional BCR locus but was devoid of other LCR22 sequences (Fig. 2; 14 104 bp of the BCR locus; May 2004 Assembly). The last six exons of BCR, an interval of 14 993 bp, are present in the other LCR22s (4). Part of the BCR gene became duplicated onto other LCR22s, but LCR22 sequences from other LCR22 copies were not present in the BAC. Thus, this clone contains 85.6% of non-LCR22 sequences and 14.4% of sequences mapping to other LCR22s, derived from BCR. To determine

Figure 1. Organization of the LCR22 genes and pseudogenes. Each LCR22 contains genes (given different color designations) and pseudogenes (keeping the same color designations). The four functional genes USP18, BCR, GGTLA and GGT are located in LCR22-2, LCR22-6, LCR22-7 and LCR22-8, respectively. USP18-red, IGSF3-orange, BCR-blue, GGTLA-green, GGT-yellow; the filled boxes represent the gene locus or the portion of the gene that is duplicated with the LCR22, whereas the colored frames represent the surrounding sequence of that particular locus. The two diagonal lines in LCR22-3a depict the region that is uncloned. The map positions of the LCR22s including surrounding sequences are (May 2004 Assembly): LCR22-2, chr22:16,979,908-17,390,952; LCR22-3a, chr22:18,568,140-19,124,855; LCR22-3b, chr22:19,344,450-19,406,306; LCR22-4, chr22:19,776,448-20,053,017; LCR22-5, chr22:21,282,984-21,333,713; LCR22-6, chr22:21,948,521-22,080,710; LCR22-7, chr22:22,933,333-23,065,522; LCR22-8, chr22:23,303,466-23,435,655.
86% of interphase nuclei (BCR, probe, suggesting that the chimpanzee lineage had a specific
chimpanzee, but only one signal from the control telomere and orangutan (green; Fig. 2). There were two signals in
one signal for the analyzed BAC specific in human, gorilla
signal detected on interphase chromosomes. We detected
of the genomic locus would produce a unique hybridization
specific for human chromosome 22 (Fig. 2). A single copy
signals on chromosome 22 in human and syntenic regions in
human and syntenic chromosome 22. Results were identical in the other
three hominoid species tested (data not shown). Interphase FISH mapping (bottom) was performed on lymphocytes (human) and fibroblasts (apes) are
shown using the same probes. Human, gorilla and orangutan cells have 1
signal per chromosome in 81, 84 and 86% cells, respectively. Chimpanzee
cells have 2 signals per chromosome in 86% cells (n = 100).

Figure 2. FISH mapping of BCR in hominoid species. The map of LCR22-6
containing the full-length BCR gene as in Figure 1 is shown. The map position
of the BAC clone, CTD-3193I16 of 130,942 bp, used for FISH analysis is
shown. The metaphase chromosome spread (DAPI stained, blue) from a
human lymphoblast is next to the map of LCR22-6. The BAC clone,
CTD-3193I16, is labeled in red (spectrum orange) and a control chromosome 22
probe (RP11-316L10) is labeled in green (spectrum green). Both signals
were present on human chromosome 22. Results were identical in the other
two hominoid species tested (data not shown). Interphase FISH mapping
was performed on lymphocytes (human) and fibroblasts (apes) are
shown using the same probes. Human, gorilla and orangutan cells have 1
signal per chromosome in 81, 84 and 86% cells, respectively. Chimpanzee
cells have 2 signals per chromosome in 86% cells (n = 100).

The copy number of the interval in hominoid species that is
contained within the BAC, we performed FISH mapping on
cell lines. Metaphase FISH mapping revealed fluorescence
signals on chromosome 22 in human and syntenic regions in
gorilla and orangutan, as confirmed using a control probe
specific for human chromosome 22 (Fig. 2). A single copy
of the genomic locus would produce a unique hybridization
signal detected on interphase chromosomes. We detected
one signal for the analyzed BAC specific in human, gorilla
and orangutan (green; Fig. 2). There were two signals in
chimpanzee, but only one signal from the control telomere
probe, suggesting that the chimpanzee lineage had a specific
duplication of the BCR locus. Two signals were found in
86% of interphase nuclei (n = 100) when compared with one
signal in human interphase nuclei (81%; n = 100).
Gibbon cells had five or greater signals per chromosome in 81%
cells (data not shown).
The LCR22-8 is 90,016 bp in size and harbors the
functional locus of GGT (25, 794 bp; exons 1–17), as well
as partial pseudogene copies of GGTLA (exon 1) and BCR [terminal 6 exons; (4)]. Two-thirds of the genomic BAC
cloned, CTD-2348F3, chosen for FISH mapping contains
non-LCR22 sequences and one-third contains sequences
present in other LCR22s (Fig. 3). Despite this, FISH
mapping showed that there is only one hybridization signal
in human cells, consistent with the presence of a single copy
of GGT on chromosome 22q11.2 (Fig. 3). Thus, even
though the BAC contained LCR22 sequences, the probe
itself was specific for this locus. As for BCR, the chimpanzee
had two signals per chromosome 23 for GGT, showing that
there are two copies of this region present. In gorilla, there
were three to four sets of signals and for the orangutan,
there were three copies on each chromosome (gorilla and
orangutan chromosome 18 and 19, respectively). Two copies
of the GGT and surrounding locus were present in gibbon.
These results indicate that GGT and/or surrounding sequences
was duplicated in non-human hominoids but not in human,
or that the human copy(ies) was lost during evolution.

The full-length functional genomic locus GGTLA is
25, 391 bp in size and it is juxtaposed to LCR22-7 (Figs 1
and 3). LCR22-7 is 31, 711 bp in size (4). A total of 75.2%
of the sequence in the BAC, CTD-2266C8, chosen for
FISH mapping contains non-LCR22 sequence and 24.8%
is present in LCR22s (Fig. 3). FISH mapping revealed signals
on metaphase chromosome 22 in human and syntenic
regions in chimpanzee, gorilla, orangutan and gibbon as
confirmed using a control probe specific for the human chromo-
some 22 telomere (data not shown). We detected one
hybridization signal by interphase FISH, for the telomere or
centromere control probe in all species (green; Figs 2–4).
However, two signals per chromosome 22, or four total, on
interphase chromosomes were present in the human cells.
Since there is only one copy of GGTLA on 22q11.2, the
cloned CTD-2266C8 is detecting another interval in addition
to the GGTLA locus. To determine whether the LCR22-7
probe was also detecting LCR22-8, the LCR22-8 probe was
labeled in green (spectrum green) and the LCR22-7 probe in
red (spectrum orange) and one yellow signal and one red
signal was identified (Fig. 3). The yellow signal derives
from the co-localization of the green and red signals. This
shows that the LCR22-7 probe hybridizes to both
LCR22-8 and LCR22-7 in humans. FISH mapping using
CTD-2266C8 in the gorilla and orangutan cell lines showed
three hybridization signals on each chromosome, identical to
that for LCR22-8. Thus, there is an additional copy of this
locus in these species, representing a duplication of the sequences
present in the BAC. The gibbon cell line had
greater than 10 copies (92%; n = 100) of this probe on
chromosome 8, in the region of synten to chromosome 22
(Fig. 3). The most parsimonious explanation for the large
increase in the gibbon is that there was a lineage specific
duplication in copy number of sequences recognized by
the LCR22-7 probe.

The LCR22-2 is 241, 823 bp in size (4). The last exon of
USP18 is located in LCR22-2, at the most proximal end
(Fig. 1). The BAC clone, CTC-771117, containing USP18
also harbors the pseudogene of KIAA0649, which is present
in all the LCR22s and is also dispersed on multiple chromo-
somes. Using the probe CTC-771117, multiple hybridization
signals were detected and the chromosome 22 signals could
not be discerned (data not shown).

For three of the four genes, BCR, GGTLA and GGT, humans
have one copy of each (two signals in Fig. 3 for GGTLA
represents hybridization of the LCR22-7 probe to both LCR22-7 and
LCR22-8 in humans), but other hominoid species showed
evidence for gene or BAC sequence duplication. We next set
out to determine whether there was variation in the copy
number of LCR22s in the hominoid species.

Lineage specific variation in copy number of LCR22s
To determine if the non-human hominoids contain multiple
copies of LCR22s, when compared with functional gene loci,
we performed FISH mapping with probes from genomic clones mapping within the human LCR22s, completely avoiding non-LCR22 sequences (Fig. 4). Two partially overlapping genomic clones, CTD-2280L11 (Fig. 4A) and PAC-996O6 (Fig. 4B), were chosen because they map to the largest LCR, LCR22-2 (Figs 1 and 4). We found three sets of signals present on interphase nuclei in human cells, using probe CTD-2280L11 from LCR22-2 (Fig. 4A). These three copies likely correspond to LCR22-2, -3a and -4 (Fig. 1). Bonobo, chimpanzee and orangutan had approximately four to six copies on each chromosome. Orangutan cells had three signals per chromosome in 79% of cells, respectively. The human, chimpanzee and gorilla cells had three signals per chromosome in 87% of the cells, as well as two signals on chromosome 19 (data not shown). The copy number of the PAC-995O6 interval is lower in orangutan and more in gibbon when compared with CTD-2280L11 (Fig. 4). This demonstrates structural differences in the LCR22s in the other hominoid species, which was confirmed in Figure 5, by the visualization of separate and shared hybridization signals in the different species in interphase FISH mapping. The FISH data using probes CTD-2280L10 and PAC-995O6 is also internally consistent with lineage specific LCR22 amplification in the gorilla (Figs 4 and 5).

Figure 3. FISH mapping of GGT and GGTLA in hominoid species. The maps of LCR22-8 and LCR22-7 containing the GGT and GGTLA genes, respectively, are shown above the FISH mapping data. LCR22-8 contains the functional copy of GGT and portions of GGTLA, and BCR. DAPI stained metaphase (top panel) and interphase nuclei (bottom two panels) are shown for five hominoid species using the LCR22-8 and LCR22-7 as well as control probes. The LCR22-8 probe derives from the BAC clone CTD-2348F3 (red) and the LCR22-7 probe derives from the BAC clone CTD-2266C8 (red). The control probe (RP5-925J7) was labeled in green. Metaphase FISH mapping using the LCR22-8 probe is shown. Results were identical using the LCR22-7 probe (data not shown). For the LCR22-8 probe, human cells have one signal per chromosome in 91% cells, chimpanzee and gibbon cells have two signals per chromosome in 79 and 88% cells, respectively. Gorilla cells have between three to four signals per chromosome in 97% cells and orangutan cells have three signals per chromosome in 87% of the cells, as well as, two signals on chromosome 19 (data not shown). For the LCR22-7 probe, human and chimpanzee cells have two signals per chromosome in 87 and 92% cells, respectively. Gorilla and orangutan cells have three signals per chromosome in 80 and 77% cells, respectively, and gibbon cells have ten or greater signals in 92% cells (n = 100). FISH mapping was performed with the LCR22-7 (red) and LCR22-8 (green) probes together.

Exonic structure of genes in LCR22s
To determine the exonic organization of genes in LCR22s, we examined previously published aCGH data (23) and performed quantitative-PCR (QPCR), on the genomic DNA from hominoid species (except for Gibbon). For whole genome aCGH, we analyzed data from a genome-wide study of gene copy number variation (CNV) (23) to assess
alterations in copy number of USP18, BCR, GGT and GGTLA (Figs. 6A). Each DNA was compared to that of human genomic DNA. Because the probes for aCGH consisted of only the 3’ most exon of each of the four genes, QPCR was performed as well to assess copy number for additional exons in each gene. For USP18, using an exon 11 probe for aCGH, there were no significant differences among the species, which was consistent with QPCR results (Fig. 6A). For BCR, using the terminal region of the gene as a probe for aCGH, a significant amplification was observed in

Figure 4. FISH mapping of LCR22-2 in hominoid species. (A) The map of LCR22-2, as in Figure 1, is shown with the genomic clone, CTD-2280L11 which was used to generate a probe for FISH mapping. Metaphase (enlarged metaphase chromosome with green and red signals taken from the image and enlarged for better visualization) and interphase FISH data (LCR22-2 clone, red; chromosome 22 centromere probe, green) on DAPI stained metaphase and interphase lymphoblasts and fibroblasts are shown for five hominoids: human, bonobo, gorilla, orangutan, gibbon and chimpanzee. As shown gorilla had the greatest number of red signals, many mapping to the telomeric regions of many chromosomes. All images examined were internally consistent \((n = 100)\). (B) The genomic clone, PAC 995O6 was used to generate a probe for FISH mapping. The copy number of the interval hybridizing to PAC-995O6 in bonobo, chimpanzee and gorilla is similar to that hybridizing to CTD-2280L11. The copy number of the same interval in orangutan is less and human and gibbon is more, compared to the interval hybridizing to CTD-2280L11. The yellow signal represents the overlap of the red and green probes.
gorilla. QPCR data showed that there was an expansion in both gorilla and orangutan (Fig. 6A). The terminal exons of BCR are expanded in human LCR22s, indicating that this region is more unstable than the 5’ end of the gene (Fig. 6A). The aCGH probe for GGTLA was constructed from the last exon, which is not amplified in humans and other Hominoid species, except for orangutan (Fig. 6A). The first exon of GGTLA is amplified in humans and other Hominoid species, except orangutan as shown by QPCR. The terminal exons of GGT were used as a probe for aCGH and showed a dramatic amplification of copies in the gorilla, confirmed by QPCR (Fig. 6A). In fact, all the exons of GGT were expanded in the gorilla, as compared to other species.

All three approaches, FISH, aCGH and QPCR, have differences in the probe content and alterations in sensitivity, with FISH proving the strongest insights at the lowest resolution and QPCR providing the highest resolution and potential for false positive signals (Fig. 6B). For USP18, we found QPCR and aCGH probes did not detect the amplification of the surrounding interval in the LCR22s because the FISH probe contains more sequences than the probes for QPCR and aCGH. We found one copy of the BCR and GGT genomic loci by FISH mapping of human interphase chromosomes and two copies in the chimpanzee, but different locations in the genome could not be detected by QPCR and aCGH approaches. Furthermore, neither QPCR nor aCGH can also distinguish between the full-length gene loci and partial pseudogene copies comprising the LCR22s, but FISH mapping with different probes can. Of note, the amplification of copy number was not consistent with the previous FISH study of primate LCR22s (13). We suggest that this may reflect differences in the probes used. The probe used in the previous study (BAC CTA-48M11) did not overlap with our probes (Fig. 1). Thus, we were able to take advantage of the different experimental approaches to distinguish the ancestral loci from the partial duplicated pseudogene copies in the LCR22s, and resolve inconsistencies. In order to further characterize the genome architecture of the LCR22s in gorilla, screening of BAC library filters was performed.

### Structural organization of LCR22s in gorilla

The genes, GGTLA (exon 1), BCR (3’ exons) and GGT (exons 1–17) showed an amplification in gorilla genomic DNA. To obtain further organizational details, we screened BAC library filters to find clones containing gorilla LCR22s. Half of the gorilla BAC library (CHORI 255; 10-fold redundancy) was screened with PCR probes for the last exon of GGT and the first exon of GGTLA to determine the organization of the LCR22s. The positive clones were analyzed for the presence of our LCR22 genes, USP18, BCR, GGT and GGTLA (Fig. 7). As expected from the real time PCR results, most of the BAC genomic clones contained the entire GGT gene. The major difference between LCR22s in human and the gorilla clones was that the entire GGT gene was amplified in the gorilla and only exons 3–17 were amplified in humans. This is consistent with the QPCR findings in Figure 6. The clones containing the full length GGT gene also had the other LCR22 genes with a similar organization to that of LCR22-2, again similar to QPCR findings (Fig. 7). Thus, the structural organization is partially conserved between humans and gorilla with the caveat that the probes used for screening the BAC filters were from GGT and GGTLA but did not include USP18 or BCR. Some of the gorilla specific variation detected in Figure 5 could have been missed.

### Evolution of the LCR22s

In the mouse genome, there is only one copy of USP18, BCR, GGTLA and GGT (22). Three of these genes, BCR, GGT and GGTLA, are found on mouse chromosome 10, while USP18 is on mouse chromosome 6 (24). In humans, LCR22-2, and 4 have almost the same block structure, showing the highest conservation overall, while LCR22-5, 6, 7 and 8 have a similar block structure, with lower levels of conservation (4). The sequence homologies between all four LCR22 genes were compared by pairwise alignment to give further insight into their evolutionary history and sequence relationships (Fig. 8). Part of BCR and GGT were amplified on 22q11.2. Examination of these two genes showed two branches of homology. One branch contained LCR22-2 and LCR22-4 and the other contained, LCR22-5, LCR22-7 and LCR22-8. All copies showed similar % identity to each other, suggesting either/both a similar evolutionary history or homogenization via gene conversion. The GGTLA gene showed a more complicated branch structure, with more distant evolutionary origin, but again, the overall branch structure matched the LCR22 organization. The percent identity between the USP18 pseudogenes, present on LCR22-2, LCR22-3a and LCR22-4, is 98.41–99.65%, suggesting that
the pseudogenes evolved more recently than the other genes or that there has been more extensive gene conversion (25).

Based upon these data, the LCR22 blocks cluster into two groups LCR22-2 and -4 and LCR22-5, 6, 7 and 8. Both the four genes as well as their pseudogene copies have shown to be dynamic and a recent evolutionary event.

Figure 6. (A) Real time PCR (QPCR) and aCGH analysis for USP18, BCR, GGTLA and GGT. Intron/exon (lines; boxes, respectively) structures of the four main genes of the LCR22s are shown. The red frame surrounding the terminal exon(s) represents the portion of the gene that is present in the LCR22s. QPCR was performed in different species (H = human; C = chimpanzee; B = bonobo; G = gorilla; O = orangutan) using primers from specific exons (boxes below the exons). Some of the exons were tested with multiple primer pairs as indicated. The copy number as determined by QPCR is indicated by color coding (white 0–8 copies; light grey 8–16 copies; dark grey 16–30 copies; black 30 or more copies). Underneath the QPCR results are the aCGH data extracted from Fortina et al (23). The aCGH was performed using a human cDNA array hybridized with human genomic DNA (green) and hominoid DNA (red). The cDNAs on the arrays for the aCGH experiments were all derived from the 3' end of the genes (blue boxes). Green signal signifies more copies are present in humans compared to the particular ape species examined and red is the opposite. Similar copy number in the two species is shown as black color. In contrast to FISH mapping, the aCGH and QPCR experiments will detect duplicated genomic loci independent as to whether they are part of LCR22s. (B) Summary of FISH, QPCR and array CGH (aCGH) data. The combined data for the four genes, USP18, BCR, GGTLA and GGT, were compared. Each method provides different types of data, making careful interpretation before drawing conclusions necessary. This figure attempts to clarify the differences among the approaches. The probe sizes for each assay with respect to the cDNA of each gene are shown above the grid, color coded as depicted in Figure 1. The grid depicts the data as presented in the previous Figures 2–6 (NI, not interpretable due to presence of repeats in probe; yellow box, discrepancies in data between different approaches). For the QPCR results, L (low), represents 0–8 copies, M (medium) is 8–16 and H (high) is 16 copies. Open boxes reflect the fact that a particular assay was not performed in that individual species.

Figure 7. Compilation of gorilla BAC clones to identify structure of LCR22s. Gorilla BAC library filters were screened with PCR products from the LCR22 genes that were amplified from gorilla DNA (3' exon for GGT and the most 5' exon of GGTLA). The DNAs from the clones were tested for the presence or absence of the genes mapping to all the LCR22s (same color code as Fig. 1). GGT-yellow; USP18-red; GGTLA-green; BCR-blue; IGSF3-orange. The BAC clones are grouped (A–G, on left) according their structure and the number of group members are shown on the right of each group. Group A has three clones (174G10; 174O22; 195N5); B, four clones (178F24; 178G19; 178H19; 182C18); C, two clones (180K7; 205A9); D, seven clones (241C14; 197H4; 199A15; 199J24; 202C20; 203N20; 208A8); E, four clones (204B19; 205D15; 209B5; 233I16); F, thirteen clones (165I5; 165I6; 186D19; 170I7; 189C2; 189E12; 196B20; 206C19; 206D8; 207J2; 229G14; 230B8; 239E18); G, twelve clones (192D2; 194P9; 195L4; 196F10; 196O24; 216J12; 230J8; 232J24; 233I14; 234P4; 235F2; 239L8); H, one clone (214O2).

For the QPCR results, L (low), represents 0–8 copies, M (medium) is 8–16 and H (high) is 16 copies. Open boxes reflect the fact that a particular assay was not performed in that individual species.
**DISCUSSION**

**Lineage specific evolution of LCR22s in hominoid species**

The 22q11.2 LCR22s comprise some of the most complex segmental duplications in the human genome. Their presence creates substrates for unequal recombination during meiosis leading to gene dosage imbalance and genomic disorders. To understand how LCR22s have evolved during evolution, we performed FISH mapping and found surprising evidence for very significant lineage specific variation of LCR22 structure in hominoid species. The most significant finding was in the gorilla, where there was a dramatic amplification of LCR22s that were located in the telomeric (or subtelomeric) bands, implicating interchromosomal recombination as being responsible.

Evolutionary studies of human LCRs associated with human genomic disorders

To determine whether lineage specific variation is a common feature, we compared our findings with LCRs on 7q11.23 (Williams-Beuren syndrome, WBS), 15q11–q13 (Prader-Willi/Angelman syndromes) and 17p11.2–p12 (Charcot-Marie-Tooth type 1A, CMT1A/hereditary neuropathy with liability to pressure palsies, HNPP and Smith-Magenis syndrome, SMS) (26–28).

The WBS deleted region is flanked by region specific LCRs of ~100 kb (29). The LCRs were analyzed by a combination of FISH mapping and molecular studies to determine the percentage of sequence divergence and phylogenetic relationships (27). Those studies showed that the duplication of the LCRs occurred recently, after the divergence of the hominoids from the Old World monkeys and were consistent with a rapid evolution coupled with both sequential and lineage specific duplications (27). Although duplications were present, there was no dramatic amplification of the chromosome 7 LCRs as observed for the LCR22s. The differences could be due to the fact that the LCRs on chromosome 7 are not as complex as those on 22q11.

The LCRs on chromosome 15q11–q13, associated with Prader-Willi and Angelman syndrome (PWS/AS) also are present and arose at least before the divergence between Old World monkeys and Great Apes (28). Multiple signals were present in primate species analyzed including the chimpanzee, gorilla and Rhesus monkey, suggesting that the LCRs were present throughout primate evolution. Similar to the WBS region, dramatic lineage specific amplification might not be present in these species, however, the LCRs on 15 have not been fully defined, and further analysis is needed to conclude that this is the case.

Two 24 kb, CMT1A-REP flank the CMT1A/HNPP duplication and deletion (30). The duplication of this relatively small LCR (24 kb) occurred after the divergence.
between the common ancestor of the gorilla and the chimpanzee (30). Thus, the evolution of the CMT1A-REP is more recent than the evolution of the other LCRs. In addition to the CMT1A-REP, there are other disease-associated LCRs on 17p11.2–12 that are more complex in structure. There are three copies of ~200 kb SMS-REP as well as additional copies on chromosome 17 (31). Evolutionary studies of the SMS-REPs suggest that three copies are present in gorilla, orangutan, baboon, rhesus monkey and the New World, squirrel monkey, suggesting a more distant evolution than for the CMT1A-REPs (31). Additional components of the SMS-REPs lie along chromosome 17 and these have a more complicated evolutionary history than the three main SMS-REPs (32). In addition to these two classes of repeats, there are different LCRs termed LCR17p that are interspersed between CMT1A-REPs and SMS-REPs (3,26,31). The data for all three are consistent with serial duplications and data may be consistent with other evolutionary mechanisms (26), but again, more work is needed in evolutionary studies of the complex LCRs to determine if lineage specific amplification might also be important on chromosome 17.

Most recently, the LCRs on chromosome 16p were analyzed by physical mapping approaches. More than 10% of 16p contains LCR16, in particular a single gene family, Morpheus, within LCR16a has been particularly prolific, due to reasons of positive evolutionary selection (33). They found evidence similar to ours for 22q11, lineage specific duplication in chimpanzee and orangutan as compared to humans. The LCRs did not become duplicated and transposed to telomeric loci as we have found, but moved to other regions of the genome, suggesting that there are different mechanisms responsible for LCR instability.

One explanation for the increase in LCR16 copy number is positive evolutionary selection. There is a gene family in LCR16 termed Morpheus that is amplified (32). In our case, we found the entire GGT gene became amplified in Gorilla, suggesting, like Morpheus, there could be positive selection in this species, based upon environmental needs.

Alu-recombination events shaped LCRs during primate evolution

One major question to address is the molecular mechanism responsible for lineage specific LCR amplification from the gene loci or from LCR progenitors. One possible mechanism is via Alu-mediated recombination. It has been shown that Alu SINE elements are at the edge of most of the segmental duplications in the 7q11.23 region associated with Williams-Beuren syndrome (27). Examination of the junctions of evolutionary duplication of LCR16 revealed a threefold enrichment of Alu content, supporting this idea (33). It is likely that similar Alu-mediated mechanisms are responsible for the lineage specific duplications observed herein. In addition to Alu-mediated mechanisms for duplications, entire LCR22s have also become duplicated. The fact that the most dramatic duplications appear to be linked to the telomeric (or subtelomeric) regions of the genome in gorilla, suggests specific molecular mechanisms, perhaps involving positive selection.

LCR22s in subtelomeric bands of gorilla chromosomes

We found a dramatic amplification of LCR22s in the telomeric bands of gorilla chromosomes. This indicates that a lineage specific interchromosomal duplication process must have taken place. A model for subtelomere rearrangements has been proposed recently in which subtelomeric segmental duplications formed by healing of double strand breaks by translocations and non-homologous end joining repair mechanisms (36). This is followed by duplication by NAHR (36). Extensive independent events were demonstrated in different primate lineages by FISH mapping (36). Another example of lineage specific duplication is the olfactory receptor genes that are found on multiple chromosomes on the subtelomeric ends in humans, but are present only once in the non-human hominoids, suggesting that their distribution occurred after the divergence of humans with its common ancestor (37,38). The amplification of the olfactory receptor genes is a product of positive selection. Based partly upon these previous studies and our data, we present a model for lineage specific LCR22 amplification in the gorilla (Fig. 9). In this model, instead of a translocation, a gene conversion could have occurred as an attempt to repair a subtelomeric double strand break and a break in an LCR22. We propose a gene conversion instead of a translocation because there is no evidence for such a translocation occurring during evolution and there is clear evidence that sequence transfer can take place without an ensuing translocation (39). Once the LCR22 was seeded onto the subtelomeric region of one chromosome, exchange could have led to amplification, as provided in the model that has been reported (36).

CONCLUSIONS

In conclusion, we describe the complex evolutionary history of LCR22s deduced by examining the genome of different hominoid species. Our results are most consistent with lineage specific duplication of gene loci and LCR22s. We found a dramatic amplification of LCR22 copies in both gibbon on the region of synteny to chromosome 22q11 and in gorilla, in the telomeric (likely the subtelomere) regions of many chromosomes. The genomic sequences, once becoming duplicated serve to seed pericentromeric, as well as 22q11.2 in this case, subtelomeric regions, which are both supported by the genome wide studies (29,40). The fact that LCRs are particularly unstable, combined with possible positive
evolutionary selection for such rearrangements provides new possibilities for understanding the basis of genome instability for creating new genes and for mechanisms leading to human disease. By delineating genome organization of the LCRs in primates, we can understand the basis for species-specific differences and potential DNA structural or functional consequences. Furthermore, there is evidence for normal intra-species CNV in the genome, detected as deletions and duplications >5 kb. More work in the future will address intra-species variation, including possible insertion and deletion polymorphisms within LCR22s. The data presented here will serve as a foundation for future investigations.

MATERIAL AND METHODS

Fluorescence in situ hybridization (fish) mapping

Epstein-Barr virus transformed lymphoblastoid cell lines (LCLs) from humans were grown and harvested using standard methods. Primate fibroblast samples: Pan troglodytes (chimpanzee-AG06939), Pan paniscus (bonobo-AG05253), Gorilla gorilla (western lowland gorilla-AG05251) and Pongo pygmaeus abelii (Sumatran orangutan-AG12256) were purchased through Coriell cell repositories and maintained under the conditions provided. The human placenta DNA (Sigma) was used for control purposes. BAC clone PAC-99506 was obtained from BAC-PAC resources (P. de Jong). DNA was prepared from each BAC using the Nucleobond kit (BD Biosciences) according to the manufacturer’s instructions.

Purification of BAC DNA for fish mapping

The BAC clones for FISH mapping were chosen based upon their position in the human genome sequence [UCSC browser, May 2004 (build 35)]. The BAC clones: CTD-2280L11, CTD-2266C8 and CTD-2348F3 were purchased from Open Biosystems. Clone PAC-99506 was obtained from BAC-PAC resources (P. de Jong). DNA was prepared from each BAC using the Nucleobond kit (BD Biosciences) according to the manufacturer’s instructions.

Real time PCR

DNAs used in the real time PCR experiments were obtained through Coriell Cell Repository: Pan troglodytes (chimpanzee-NG06939), Pan paniscus (bonobo-NG05253), Gorilla gorilla (western lowland gorilla-NG05251) and Pongo pygmaeus abelii (orangutan-NG12256). Human placenta DNA (Sigma) was also used in the real time PCR experiments. The primers (conserved between human and mouse) were designed in the exons (Supplementary Material, Table S1) of the genes. Real time PCR was performed using SYBR Green mix (Applied Biosystems). The PCR conditions were 0.5 μM of each primer with 4 μl of the SYBR Green mix, and the DNAs were set at varying concentrations (1.25, 5, and 20 ng/μl). All samples were PCR amplified on the real time thermocycler ABI 7900HT at 50°C for 2 min, 95°C for 10 min, (95°C for 10 s, 60°C for 20 s, 72°C for 30 s for 40 cycles), 95°C for 15 s, 60°C for 15 s and 95°C for 15 s. Data were analyzed using SDS 2.1 software.

Array comparative genome hybridization

aCGH was performed in Fortna et al. (23). The data were extracted for the LCR22s from Fortna et al., 2004. The cDNAs used on the array for USP18, BCR, GGTLA and GGT were from the 3’ end of the genes GenBank ID AA626356 (USP18), A1968010 (BCR), AA419342 (BCR), AA150687 (GGTLA), A1342464 (GGT) and A1345015 (GGT).

Identification of BAC clones for gorilla LCR22s

Half (~5-fold redundancy) of the gorilla BAC library (CHORI-255) was screened using overgo probes, which are two anti-parallel 24-mers overlapping 8 nucleotides at the 3’ ends, thus generating a 40-mer duplex after filling in with radioactive dNTPs. The primers were designed using John McPherson’s approach and modified algorithm. The probes were designed for human chromosome 22 sequences, where the gorilla cDNA showed an increase compared to human DNA. The primers were then pooled and hybridized to the filters. After obtaining the clones that were positive in the screening, they were screened by PCR using the following primers: GGT_J04131_3’ Exon 1 -CTCTATGCCTGGG TGCAA, GGT_J04131_3’ Exon 1 -TGCTAGTATCTCG TCCCTGG; GGTLA Exon 1 -CTGTGACATTTT CGTGCTC, GGTLA Exon 1 -R-GCA GCCAGAGGAGGA GG. Those clones were then screened by the same primers.

Figure 9. Model of lineage specific LCR22 amplification in gorilla. The GGT gene and LCR22s are present in gorilla chromosome 23 (dark gray band in white chromosome). We propose that during meiosis, a double strand break occurred in the LCR22 and the telomeric band of another chromosome (light gray). This led to a gene conversion event (strand invasion and extension) in which the LCR22 became duplicated (rejoined) onto the telomeric band of the damaged chromosome. Telomeric bands are susceptible to breakage. The ends must be repaired or else the chromosome will be lost. To prevent this, there are frequent exchanges of genetic material between telomeres (B), resulting in a dramatic amplification of LCR22s, in this case containing the full-length GGT gene, in the telomeric bands in the gorilla. If the copies in the telomeric bands do not have mutations and are expressed as functional proteins, it may be possible to conclude that positive selection could have occurred thereby fixing these amplified copies in this species.
used in the real time PCR analysis, as well as, primers from
the one exon of IGSF3 found in the LCR22s (Supplementary
Material, Table S1). The PCR conditions were 94°C for
10 min, followed by 30 cycles of 94°C for 30 s, 58°C for
30 s and 72°C for 30 s, and extension of 72°C for 10 min in
2.5 mM Mg, 0.2 mM dNTPs, 0.2 μM of each primer and
0.04 units of Fast Start High Fidelity Taq polymerase (Roche).

ACKNOWLEDGEMENTS
We would like to acknowledge Dr Adam Pavlicek at GIRI for
help generating Figure 8 and for intellectual guidance. We
would like to thank Tomas Stopka for all his insight on how
to calculate copy number by using real time PCR. We also
appreciate discussions with Drs. David Nelson and Lisa
Shaffer during the course of this project.

Conflict of Interest statement. None declared.

FUNDING
J.M.S. is supported by National Institute of Health AA11853
and a grant from the Butcher Foundation. J.R.L. and B.E.M.
are supported by P01 HD39420.

REFERENCES
1. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin,
J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001) Initial
3. Stankiewicz, P., Shaw, C.J., Dapper, J.D., Wakui, K., Shaffer, L.G.,
architecture catalyzes nonrecurrent chromosomal rearrangements.
Am. J. Hum. Genet., 72, 1101–1116.
4. Babcock, M., Pavlicek, A., Spiteri, E., Kashork, C.D., Joshikhes, L.,
low-copy repeats on 22q11 (LCR22) by Alu-mediated recombination
Genet., 34, 246–247.
can lead to DNA rearrangements and human disease traits. Trends Genet.,
14, 417–422.
Kimonis, V., Summers, A.M., Meschino, W.S., Anyone-Yeboa, K.,
Kashork, C.D. et al. (2000) Molecular mechanism for duplication
17p11.2: the homologous recombination reciprocal of the Smith-Magenis
(2004) A new genomic duplication syndrome complementary to the
10. Yobb, T.M., Somerville, M.J., Willatt, L., Firth, H.V., Harrison, K.,
(2005) Microduplication and triplication of 22q11.2: a highly variable
chromosome 15q in a boy with the Prader-Willi syndrome; further
mediate the common 3-Mb deletion in patients with velo-cardio-facial
(2004) Analysis of segmental duplications and genome assembly in the
14. Fortna, A., Kim, Y., MacLean, F., Marshall, K., Hahn, G., Meltlesen, L.,
Bromley, M., Hink, R., Burgers, S., Hernandez-Boussard, T. et al.
(2004) Lineage-specific gene duplication and loss in human and great ape
15. Puech, A., Saint-Jore, B., Funke, B., Gilbert, D.J., Sirothin, K., Copeland,
N.G., Jenkins, N.A., Kucherlapati, R., Morrow, B. and Skoulthench, A.
(1997) Comparative mapping of the human 22q11 chromosomal region
and the orthologous region in mice reveals complex changes in gene
Traffic of genetic information between segmental duplications flanking
the typical 22q11.2 deletion in velo-cardio-facial syndrome/DeGeorge
(2004) Serial segmental duplications during primate evolution result in
(2005) Evolutionary mechanisms shaping the genomic structure of
the Williams-Beuren syndrome chromosomal region at human 7q11.23.
D.H. (1999) Large genomic duplicons map to sites of instability in the
Prader-Willi/Angelman syndrome chromosome region (15q11–q13).
Fine-scale comparative mapping of the human 7q11.23 region and the
orthologous region on mouse chromosome 5G: the low-copy repeats that
flank the Williams-Beuren syndrome deletion arose at breakpoint sites of
an evolutionary inversion(s). Genomics, 69, 1–13.
(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.
22. Park, S.S., Stankiewicz, P., Bi, W., Shaw, C., Lehoczky, J., Dewar, K.,
Smith-Magenis syndrome repeat gene clusters, SMS-REPs. Genome Res.,
12, 729–738.
23. Ou, Z., Jarmuz, M., Sparagana, S.P., Michaud, J., Decarre, J.C., Yatsenko,
(2006)


