Gorilla genome structural variation reveals evolutionary parallelisms with chimpanzee

Mario Ventura, Claudia R. Catacchio, Can Alkan, et al.

*Genome Res.* 2011 21: 1640-1649 originally published online June 17, 2011
Access the most recent version at doi:10.1101/gr.124461.111

<table>
<thead>
<tr>
<th>Supplemental Material</th>
<th><a href="http://genome.cshlp.org/content/suppl/2011/06/15/gr.124461.111.DC1.html">http://genome.cshlp.org/content/suppl/2011/06/15/gr.124461.111.DC1.html</a></th>
</tr>
</thead>
</table>
| Related Content       | Genome-based analysis of the nonhuman primate *Macaca fascicularis* as a model for drug safety assessment  
Martin Ebeling, Erich Küng, Angela See, et al.  
*Genome Res.* October, 2011 21: 1746-1756 |
|                       | Copy number variation analysis in the great apes reveals species-specific patterns of structural variation  
Elodie Gazave, Fleur Darré, Carlos Morcillo-Suarez, et al.  
*Genome Res.* October, 2011 21: 1626-1639 |
| References            | This article cites 53 articles, 20 of which can be accessed free at:  
http://genome.cshlp.org/content/21/10/1640.full.html#ref-list-1 |
|                       | Article cited in:  
http://genome.cshlp.org/content/21/10/1640.full.html#related-urls |
| Creative Commons License | This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see  
http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/. |
| 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:  
http://genome.cshlp.org/subscriptions

© 2011, Published by Cold Spring Harbor Laboratory Press
Gorilla genome structural variation reveals evolutionary parallelisms with chimpanzee

Mario Ventura,1,2 Claudia R. Cataczchio,1,2 Can Alkan,1,3 Tomas Marques-Bonet,1,4 Saba Sajjadian,1 Tina A. Graves,5 Fereydoun Hormozdari,6 Arcadi Navarro,4,7,8 Maika Malig,1 Carl Baker,1 Choli Lee,1 Emily H. Turner,1 Lin Chen,1 Jeffrey M. Kidd,1,9 Nicoletta Archidiacono,2 Jay Shendure,1 Richard K. Wilson,5 and Evan E. Eichler1,3,10

1Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington 98195, USA; 2Department of Genetics and Microbiology, University of Bari, Bari 70126, Italy; 3Howard Hughes Medical Institute, Seattle, Washington 98195, USA; 4IBE, Institut de Biologia Evolutiva (UPF-CSIC), Universitat Pompeu Fabra, Barcelona 08003, Catalonia, Spain; 5Washington University Genome Sequencing Center, School of Medicine, St. Louis, Missouri 63108, USA; 6School of Computing Science, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada; 7Population Genomics Node (GNV8), National Institute for Bioinformatics (INB), Barcelona 08003, Catalonia, Spain; 8Institució Catalana de Recerca i Estudis Avancats (ICREA) and Universitat Pompeu Fabra, Barcelona 08003, Catalonia, Spain; 9Department of Genetics, Stanford University, Stanford, California 94305, USA

Structural variation has played an important role in the evolutionary restructuring of human and great ape genomes. Recent analyses have suggested that the genomes of chimpanzee and human have been particularly enriched for this form of genetic variation. Here, we set out to assess the extent of structural variation in the gorilla lineage by generating 10-fold genomic sequence coverage from a western lowland gorilla and integrating these data into a physical and cytogenetic framework of structural variation. We discovered and validated over 7665 structural changes within the gorilla lineage, including sequence resolution of inversions, deletions, duplications, and mobile element insertions. A comparison with human and other ape genomes shows that the gorilla genome has been subjected to the highest rate of segmental duplication. We show that both the gorilla and chimpanzee genomes have experienced independent yet convergent patterns of structural mutation that have not occurred in humans, including the formation of subtelomeric heterochromatic caps, the hyperexpansion of segmental duplications, and bursts of retroviral integrations. Our analysis suggests that the chimpanzee and gorilla genomes are structurally more derived than either orangutan or human genomes.

[Supplemental material is available for this article.]

The nature of the genetic differences between humans and other great apes has fascinated scientists since the discovery of DNA in the 1950s (Sarich and Wilson 1973; Yunis and Prakash 1982; Goodman et al. 1989). The genetic relationship and phylogeny of humans and great apes is well established, based primarily on studies of single nucleotide variation (Koop et al. 1986; Enard and Paabo 2004). A surprising finding has been the extent of larger forms of structural variation among hominid genomes well below the limit of cytogenetic resolution (Locke et al. 2003; Fortna et al. 2004; Cheng et al. 2005; Bailey and Eichler 2006; Gibbs et al. 2007; Marques-Bonet et al. 2009a). Interestingly, the hominid genomes appear to be enriched with respect to structural variation, but the extent to which this has impacted each of the major lineages is not yet completely known. To date, three hominid genomes have been sequenced and assembled to the working draft stage using capillary-based approaches ([human (The International Human Genome Sequencing Consortium 2001, 2004), chimpanzee (The Chimpanzee Sequencing and Analysis Consortium 2005), and orangutan (Locke et al. 2011)]). Projects are underway to sequence additional apes including the bonobo, gorilla, and gibbon. Many of these remaining ape genomes will be sequenced and assembled using a combination of next-generation sequencing and capillary whole-genome shotgun sequence data sets (Marques-Bonet et al. 2009b). Studies of structural variation, however, are complicated by difficulties in detecting and accurately resolving the sequence structure of these regions. In this study, we set out to systematically investigate the pattern of structural variation in the gorilla genome, combining capillary-based clone sequencing and next-generation genome sequencing in conjunction with detailed cytogenetic characterization and experimental validation. We present a comprehensive overview of inversions, deletions, segmental duplications, and retrotranspositions within the gorilla genome. Comparisons with humans and other apes reveal that parallel and independent mutational processes have more dramatically restructured chimpanzee and gorilla genomes when compared with other hominid genomes.

Results

In order to investigate the gorilla’s pattern of genome structural variation, we undertook a three-pronged approach. First, we tested 788 human BAC clones by fluorescence in situ hybridization (FISH), comparing the probe order on human and gorilla chromosomal metaphases, thus providing a refined cytogenetic framework of large-scale and intermediate-sized rearrangement events (Supplemental Note). Next, we completely end sequenced 176,880 BAC clones (http://www.genome.gov/Pages/Research/Sequencing/BACLibrary/primateProposal.pdf) from a gorilla BAC library (CH277) and mapped them to the human reference genome [NCBI build 35 (NCBI35)].
to generate a clone-based framework of the gorilla genome (Eichler and DeJong 2002). This approach defined potential rearrangements based on discordant end sequence placements. Last, we obtained blood DNA from Kwan, a middle-aged silverback gorilla, and generated 9.6-fold effective sequence coverage using massively parallel Illumina sequencing. While this sequence coverage means that each base is represented on average nine to 10 times, the paired-end sequences flanking a portion of the insert that is not sequenced means a larger fraction of the genome is spanned by anchored mate pairs (34-fold). These data were used to identify regions of copy number variation based on sequence read-depth and paired-end mapping revealing smaller forms of structural variation including mobile element insertions (>300 bp) using end sequence profiling approaches (Tuzun et al. 2005; Hormozdiari et al. 2009; Hormozdiari et al. 2010). The experimental and molecular data were integrated (Table 1; Supplemental Note), allowing us to correctly reclassify events that particularly distinguished translocations from duplicative transposition events and inversions from segmental duplications (SDs). For example, translocations could be distinguished from duplicative transpositions because read-depth and array comparative genomic hybridization (arrayCGH) predicted copy number changes of a segment of DNA but with no evidence of chromosomal rearrangement using cytogenetic markers. In those cases where we were able to completely sequence the corresponding BAC clone, the breakpoints could be resolved at the single base pair level. We summarize the pattern of gorilla genome structural variation from the perspective of size and class, and then compare our findings with human and other great ape genomes.

Large-scale rearrangements and duplicative transpositions

Yunis and Prakash originally reported 11 large-scale cytogenetic differences between human and gorilla (eight pericentric inversions, one paracentric inversion, one translocation, and one fusion) (Yunis and Prakash 1982). Using FISH and data from other studies (Dutrillaux 1980; Yunis and Prakash 1982; Montefalcone et al. 1999; Muller et al. 2000, 2003; Carbone et al. 2002; Eder et al. 2003; Locke et al. 2003; Misceo et al. 2003, 2005; Ventura et al. 2003, 2004; Cardone et al. 2006, 2007; Stanyon et al. 2008), we refined all classical evolutionary breakpoints that distinguish the human and gorilla karyotypes (Supplemental Table S1). It should be noted that cytogenetics typically only resolves megabase pair-level variation, while FISH analyses utilizing overlapping probes can be used to resolve events ~50 kb in size, depending on the complexity of the region. Using gorilla BAC end-sequence (BES) data mapped against the human genome (NCBI35), we identified 424 putative chromosomal rearrangements (Supplemental Table S2; Supplemental Note) including six of the 11 original classical rearrangements. The remaining five were confirmed by FISH but mapped to large and nearly identical SDs that could not be traversed by BES. We selected 14 representative BAC clones corresponding to the classical gorilla-human breakpoints and completely sequenced them using capillary-based sequencing methods (Table 2). Detailed breakpoint analyses (Fig. 1; Supplemental Figs. S1–S5) showed SDs (translocation 5:17 and inversion 7) or common repeat elements (Alus and LINEs) at or in close proximity to all breakpoints. Due to the abundance of these elements in the hominid genome and the small number of sequenced sites, this enrichment is not significant. In most cases, the repetitive sequences were not homologous, suggesting that mechanisms other than nonallelic homologous recombination were responsible for these evolutionary rearrangements. None of these rearrangements disrupted unique genes.

The BAC read-pair analysis predicted an unusually large number of putative inversions and translocations, which is inconsistent with previous chromosomal analyses and our own cytogenetic framework. We selected a subset of these events (six putative translocations and 14 inversions) (Supplemental Table S3) for further investigation. For each rearrangement, if the predicted translocations and inversions were bona fide, we would expect a change in the order of flanking probes (inversions) or a change in chromosomal location (translocation) when comparing human and gorilla. For each of these breakpoints, we selected gorilla BAC clones spanning the putative rearrangement breakpoints, as well as gorilla BAC clones located distally and proximally to each breakpoint, and tested their order between human and gorilla. In all cases, no change in the order of flanking unique sequences was observed. These FISH results suggested the presence of duplicated sequences at new locations in the gorilla genome.

We completely sequenced 20 of the corresponding BAC clones (3.2 Mb of finished capillary-based sequence) to resolve the structure of these selected loci (Supplemental Table S3). In each case, we confirmed duplicative transpositions and gorilla-specific juxtapositions of SDs as opposed to inversions and translocations. For example, BLAST sequence similarity searches of the sequenced gorilla BACs from chromosome 5 (AC23944) (Fig. 1B) indicate that this portion of the gorilla genome consists of a mosaic of four diverse SDs originating from different locations on human chromosome 5. The Miropeats analysis (Parsons 1995) shows the extent of each duplicated segment (ranging in length from 9 to 86 kb). Sequence read-depth analysis among the various species [whole-genome shotgun sequence detection (WSSD) tracks; see Supplemental Note] suggests that different segments have been duplicated at different time points during evolution. We conclude that this particular architecture is unique to gorilla originating from a series of duplicative transpositions to gorilla chromosome 5p13.2. We note that none of the 20 sequenced regions from BACs were collinear with the human genome. These regions carry, on average, three reconfigured or newly interspersed duplications with an average length of 29 kb. We estimate 79% map interspersively within gorilla chromosomes, with another 21% mapping to subtelomeric or pericentromeric regions. These data reveal extensive

<table>
<thead>
<tr>
<th>Table 1. Gorilla sequence and FISH resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>WCS libraries</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>BAC clones</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>BAC end sequence</td>
</tr>
<tr>
<td>BAC clones</td>
</tr>
</tbody>
</table>
Table 2. BAC sequenced chromosomal breakpoints

<table>
<thead>
<tr>
<th>Rearrangement</th>
<th>Breakpoint</th>
<th>Clone</th>
<th>Accession</th>
<th>Size (bp)</th>
<th>Location by BES</th>
<th>Human mapping</th>
<th>Gorilla mapping</th>
<th>Repeat elements at/near breakpoint</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inv chr7</td>
<td>7q arm</td>
<td>CH277-505P01</td>
<td>AC243004.1</td>
<td>36,298</td>
<td>chr7:76489360-102048010</td>
<td>7q15; 7q22</td>
<td>VIIq</td>
<td>Seg Dup</td>
<td>Inversion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH277-325N15*</td>
<td>AC242656.3</td>
<td>95,809</td>
<td>chr7:76558315-102048039</td>
<td>7q15; 7q22</td>
<td>VIIq</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH277-111L17</td>
<td>AC242655.3</td>
<td>180,892</td>
<td>chr8:831006213-86005534</td>
<td>8p12; 8q21</td>
<td>VIIIq</td>
<td></td>
<td>Inversion</td>
</tr>
<tr>
<td>Inv chr8</td>
<td>8q arm</td>
<td>CH277-481C13</td>
<td>AC242627.3</td>
<td>158,706</td>
<td>chr8:31148163-85832627</td>
<td>8q12; 8q21</td>
<td>VIIIq</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH277-402K23</td>
<td>AC242595.2</td>
<td>102,724</td>
<td>chr8:31148078-85914766</td>
<td>8p12; 8q21</td>
<td>VIIIq</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH277-401D02</td>
<td>AC243002.1</td>
<td>55,127</td>
<td>chr8:31287889-86049917</td>
<td>8p12; 8q21</td>
<td>VIIq</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH277-12SA6*</td>
<td>AC241241.3</td>
<td>198,809</td>
<td>chr10:27573884-80780850</td>
<td>10p12</td>
<td>Xp, Xq</td>
<td></td>
<td>Inversion</td>
</tr>
<tr>
<td>Inv chr10</td>
<td>10p arm</td>
<td>CH277-103D21*</td>
<td>AC241522.2</td>
<td>188,246</td>
<td>chr10:27675315-80631965</td>
<td>10p12</td>
<td>Xp, Xq</td>
<td>L1m4 (LINE)/AluSc</td>
<td></td>
</tr>
<tr>
<td>Inv chr12</td>
<td>12q arm</td>
<td>CH277-205P14</td>
<td>AC240968.2</td>
<td>146,531</td>
<td>chr12:21041953-63546183</td>
<td>12p12; 12q14</td>
<td>XIip</td>
<td>AluSx/L1MS (LINE)</td>
<td>Inversion</td>
</tr>
<tr>
<td>Inv chr18</td>
<td>18q arm</td>
<td>CH277-2308</td>
<td>AC239638.4</td>
<td>59,604</td>
<td>chr18:177407-6874356</td>
<td>18p12; 18q11</td>
<td>XVIIIq</td>
<td>AluSx2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH277-492F03</td>
<td>AC243003.2</td>
<td>36,580</td>
<td>chr18:177407-6874356</td>
<td>18p12; 18q11</td>
<td>XVIIIq</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH277-54SD07</td>
<td>AC243178.1</td>
<td>187,650</td>
<td>chr18:287777-168000762</td>
<td>18p12; 18q11</td>
<td>XVIIIq</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t5:17</td>
<td>5q arm</td>
<td>CH277-159N16*</td>
<td>AC240953.2</td>
<td>169,000</td>
<td>chr5:79817957-80044726</td>
<td>Sq14</td>
<td>XVIIIq</td>
<td>Seg Dup</td>
<td>Translocation</td>
</tr>
</tbody>
</table>

All positions are relative to NCB135.

*Clone duplicated in human, chimpanzee, gorilla, and orangutan.
duplicative transposition in the gorilla genome creating a complex pattern of SDs unique to this lineage (see below).

**Deletions**

We initially detected 79 large deletions (>50 kb) compared with human using a combination of interspecies arrayCGH, BES mapping, and depletion in sequence read-depth. Of these events, 89% (70/79) were confirmed experimentally by arrayCGH and/or FISH (Supplemental Table S4). Based on human genome annotation, these regions contained 61 genes that were either completely (38) or partially (23) deleted in Kwan (Supplemental Table S4). We examined 52 of these regions by FISH and found that 62% (32/52) of these apparent deletions correspond to regions of duplication in the human where the gorilla simply showed reduced copy number. Only 16 of the regions contained unique hominid genomic regions that had been completely deleted in the gorilla lineage (Supplemental Fig. S6; Supplemental Table S4). In order to detect smaller deletions in the gorilla genome, we searched for clusters of discordant read pairs based on mapping gorilla sequence (Supplemental Note) against the human reference genome (Tuzun et al. 2005; Hormozdiari et al. 2009). We experimentally validated 1820 deletion intervals (6.7 Mb) using a customized microarray (Supplemental Note). This included 580 partial and 13 complete gene deletions (Supplemental Table S5; Supplemental Note). Many of the completely deleted genes belong to well-known gene families, including olfactory receptors (OR10K1, OR5L2, OR5D16, OR1M1, and OR7G2), keratin-associated proteins (KRTAP13-3 and KRTAP13-4) and HLA genes [HCP5 (HLA complex P5) and HCG26 (HLA complex group 26)]. One particularly intriguing deletion included SELV (selenoprotein V), thought to be important in the metabolism of dietary selenium, implicated in cancer prevention, immune function, aging, male reproduction, and other physiological processes (Kryukov et al. 2003). Among the partial gene deletions, 36 span more than a single exon—the largest being DNAH14 with 45 deleted exons. In total, we discovered and validated 1863 deletion intervals in the gorilla genome corresponding to 12.69 Mb.

**Mobile elements**

We initially excluded regions with >80% repeat content from our deletion analysis. Figure 1. (Legend on next page)
due to difficulties in validating these calls by arrayCGH. However, such deletions likely correspond to mobile element insertions in the human genome that occurred since the two lineages diverged. We specifically searched for the aforementioned events by identifying deletions where the corresponding locus in human was composed largely of a particular common repeat. We predicted 2481 Alu (887 kb), 1861 L1 (5.59 Mb), 663 SVA (1.17 Mb), 524 LTR (1.27 Mb), and 110 HERV (409 kb) insertions in human when compared with gorilla (Hormozdiari et al. 2010). A subset of these were full length, including 2372 Alu (≥275 bp) and 564 L1 (≥5.8 kb) elements. Of these mobile elements, 37% (2066/5639) mapped within 1605 genes (Supplemental Table S6). We performed the reciprocal analysis by searching for retrotranspositions specifically within the gorilla lineage. We detected the insertion breakpoints of various classes of active mobile elements (Alu, L1, SVA, and LTR) as well as nonhuman primate-specific endogenous retroviruses (PTERV1 and PTERV2) (Yohn et al. 2005) and predicted a total of 263 PTERV1, 4272 Alu, 325 SVA, 113 LTR, and 299 full-length L1 insertions (Supplemental Table S7). In order to estimate the rate of false positives, we tested 30 full-length gorilla Alu retrotransposons by PCR analysis of gorilla and human DNA (Supplemental Note). Of the events, 90% (27/30) validated as fixed insertions (Supplemental Fig. S7), with the remaining three being polymorphic. Consistent with recent analyses of human and other ape genomes, these results predict an acceleration of SVA retrotransposition in the chimpanzee-human ancestral lineage with a more recent surge of Alu retrotransposons in the human branch (Table 3). While we found no evidence of PTERV2 integrations in gorilla, we did identify 263 full-length integrations of PTERV1—an endogenous retrovirus initially discovered in chimpanzee (Yohn et al. 2005). A comparison with a previously developed integration map of chimpanzee revealed that 99.6% of these integrations are nonorthologous, mapping to different locations in the gorilla and chimpanzee genomes (Fig. 2). Both experimental and sequence analyses confirm that this mobile element is completely absent from human and orangutan genomes. This provides strong support that PTERV1 arose from an exogenous source that retrotransposed independently in both gorilla and chimpanzee lineages <6 million years ago.

**Segmental duplications**

We developed an SD map of the gorilla genome based on detecting regions with excess sequence read-depth as described previously (Bailey et al. 2002; Alkan et al. 2009; Marques-Bonet et al. 2009a). We detected 99 Mb of SDs (≥20 kb in length and >95% identity) (Cheng et al. 2005; Marques-Bonet et al. 2009a). We validated the duplications by interspecies arrayCGH, discovering 68 complete or partial gene duplications in the gorilla (Supplemental Note). Although most of the duplications are shared with other hominids (Fig. 3; Supplemental Note), we note an apparent excess of gorilla-specific duplications when compared with human, chimpanzee, or orangutan. Comparing the SD maps of five primate genomes, we assigned shared and lineage-specific SDs (Fig. 3B) and computed a genomic duplication rate along each branch under a maximum likelihood model, which assumes 20% homoplasy. The addition of gorilla duplication data into a maximum likelihood framework suggests an excess of SD in the human-African great ape ancestor that is larger than previously reported (Marques-Bonet et al. 2009a), with an estimated rate four- to fivefold higher when compared with the human or chimpanzee branches. Surprisingly, the gorilla-specific branch is also significantly accelerated compared with the human (approximately two to four times), but less so when compared with the common ancestral branch of humans and chimpanzees. This difference becomes more dramatic in the gorilla when adjusting for copy number (Fig. 3A), indicating that several sequences have expanded more prominently. We tested and validated by FISH (Supplemental Table S8) 11 gorilla-specific duplications showing the highest copy number increase (11–45 copies). Nine of these 11 regions contain genes completely and/or partially expanded specifically in the gorilla lineage (e.g., NDUF8—unknown function, LRPA—lowlow-density lipoprotein receptor-related protein, DOK7—downstream of tyrosine kinase 7, HGF—activator propeptide, LETM1—leucine zipper-EF-hand containing transmembrane, and FGR3—fibrobactin growth factor receptor 3 isoform 2). Most of these expansions map to the termini of gorilla chromosome 12 (Egozcue and Chiarelli 1967; Miller et al. 1974). A schematic of the 42.5 Mb inversion (top panel ideogram) is refined by paired-end BES and confirmed by FISH (bottom right panel). A gorilla BAC clone (CH277-205P14) corresponding to the long-arm breakpoint is completely sequenced (AC240968). Miropeats analysis (Parsons 1995) compares two regions on human chromosome 12 (middle panel; blue and orange) with the gorilla sequence. (Repeat elements, UNES, SINES, LTRs, SDA, and genes are annotated based on the human genome.) (Bottom left panel) ClustalW alignment pinpoints an AluSx element (purple) at the precise breakpoint. (B) Sequence characterization of a duplicative transposition region in gorilla. (Top panel) A region on chromosome 5p13 near the retinoic acid-induced 14 gene (RAI1) has acquired at least four SDs in the gorilla lineage. (Middle panel) Sequencing of gorilla BAC clone CH277-50D8 (AC239444.4) and Miropeats analysis show SDs ranging in size from 10 to 28 kb along chromosome 5. Read-depth analyses (WSSD tracks) predict that these gorilla duplicative transpositions are focused on a more ancient duplication carrying the spinal muscular atrophy type 4 gene (PSM4A), with flanking duplications becoming increasingly lineage-specific. (Lower panel) FISH analysis with the clone as a probe shows multiple signals. Unique probes proximal and distal to the region indicate no change in the order between human and gorilla. Duplicative transpositions have occurred without a large-scale chromosomal rearrangement. (C) Duplicative transposition at translocation fusion point. (Top panel) Gorilla chromosome XVII arose as a lineage-specific fusion between chromosome 5 and 17 (Stankiewicz et al. 2001). Cloning and sequencing of the breakpoint (CH277-159N16, AC240953) show the presence of a 385 kb complex gorilla-specific duplication block at the breakpoint. The duplication block is a mosaic composed of at least five distinct SDs originating primarily from chromosome 5 (see Supplemental Note for more detail).

**Table 3. Mobile element comparison among hominid genomes**

<table>
<thead>
<tr>
<th>Humana</th>
<th>Humanb</th>
<th>Chimpanzee</th>
<th>Gorilla</th>
<th>Orangutan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alu</td>
<td>584</td>
<td>7082</td>
<td>2340</td>
<td>4272</td>
</tr>
<tr>
<td>L1</td>
<td>52</td>
<td>1814</td>
<td>1979</td>
<td>299</td>
</tr>
<tr>
<td>SVA</td>
<td>14</td>
<td>970</td>
<td>400</td>
<td>325</td>
</tr>
<tr>
<td>PTERV</td>
<td>ND</td>
<td>ND</td>
<td>275</td>
<td>263</td>
</tr>
</tbody>
</table>

Gorilla-specific mobile elements detected by VariationHunter (see text for details); orangutan-specific elements (Locke et al. 2011). Mobile element prediction is based on comparison with human genome (NCBI35). (ND) not detected

aHuman mobile elements (Venter et al. reference genome) (Xing et al. 2009).

bHuman-specific mobile elements compared with chimpanzee (The Chimpanzee Sequencing and Analysis Consortium 2005).
duplications among humans, chimpanzees, and gorillas searching for regions of hyperexpansion (>500 copies) in one lineage when compared with the other two (Supplemental Table S9). Only in the gorilla and chimpanzee genomes were such hyperexpanded SDs identified with expansions of 1000–1500 copies mapping primarily to acrocentric, pericentromeric, subtelomeric, and subterminal cap regions of African ape chromosomes (Supplemental Table S9).

Subterminal caps

One of the most striking karyotypic differences between humans and African apes is the presence of subterminal heterochromatic caps at the ends of ape chromosomes (Fig. 4A). Evident by G banding (Yunis and Prakash 1982) and post-denaturation DAPI staining, these regions have been classified as subterminal heterochromatin found exclusively among gorillas (80/96 chromosomes), the common chimpanzee (Pan troglodytes) (42/96 chromosomes), and bonobo (Pan paniscus) (42/96 chromosomes). They are thought to be composed primarily of a 32-bp satellite repeat sequence (pCht7/13 sequence) arrayed in tandem (Royle et al. 1994). In addition, it is known that the formation of the subterminal cap in chimpanzee was accompanied by the hyperexpansion of SDs, which map near the human chromosome 2 fusion point (113997859-114024033, NCBI35) (Fan et al. 2002; Cheng et al. 2005). In gorilla, we find no evidence of an association of chromosome 2 sequences with the heterochromatic caps, but rather our copy number and FISH analysis suggests that a segment of chromosome 10 (19557646-19564636, NCBI35) is one of the primary components of the gorilla cap (Fig. 4A). To confirm these results, we selected three large-insert BAC clones corresponding to the cap regions of both gorilla and chimpanzee and subjected these to capillary-based sequence and assembly. The sequence analysis shows dramatic differences in the organization of heterochromatic caps between the two species. While both possess tracts of pCht satellite sequence ranging in size from 10 to 50 kb, chromosome 2 SDs are predominant in the chimpanzee cap, whereas chromosome 10 duplications define the cap organization in gorilla (Fig. 4B). These duplications appear to have expanded in concert with the satellite sequence creating a higher-order tandem array structure of several hundred copies in each species. Since subterminal heterochromatic blocks have also been reported among the lesser apes (Wijayanto et al. 2005), we tested pCht satellite probes on gibbon metaphase chromosomes and observed no hybridization signal above background (data not shown). Combined, these data strongly suggest that the heterochromatic caps have evolved independently in both chimpanzee and gorilla and possibly all ape species.

Discussion

Structural variation has been extensive and episodic during human-great ape evolution. In this study, we identified and validated >7665 (Table 4) structural variant events in the gorilla when compared with human. It is important to note that our analysis is based primarily on a single gorilla genome (Kwan). Consistent with other studies, we expect 10%–30% of these variants to be polymorphic (i.e., not fixed within the gorilla lineage) (Chen and Li 2001; Ebersberger et al. 2002; Marques-Bonet et al. 2009a). Nevertheless,
is possible that extensive differences in the SD architecture facili-

tated genetic isolation of emerging species during evolution (White 1978). We caution, however, that we cannot accurately estimate the time of such SDs with respect to hominid speciation events, so such correlations remain speculative. Thus, a reasonable line of inquiry going forward will be to compare the extent of genetic diversity in an unbiased fashion within each great ape lineage and compare these with divergence estimates between species.

In this study, we document hundreds of duplicative trans-

position differences between human and gorilla that alter the structure of duplication blocks between the two lineages. Most of these structural differences are opaque to standard whole-genome shotgun sequence assembly methods or would be incorrectly classified without integration into a higher-level cytogenetic framework. Our structural variation analysis also suggests that the genomes of chimpanzee and gorilla have experienced several independent genomic rearrangements that did not occur during the evolution of the orangutan and human. The African ape genomes have been bombarded by retroviral integrations that entered the germline after the two lineages diverged. Neither orangutan nor human genomes carry these retroelements (Yohn et al. 2005) and the fact that fine mapping of the integration sites are largely nonorthologous argues for ancient parallel infections (Kaiser et al. 2007). Gorilla and chimpanzee have independently acquired subtelomeric heterochromatin caps, and this chromosome feature has been associated with the hyperexpansion of different SDs in the two lineages. Our molecular analyses suggest that these events occurred independently and in parallel early during the evolution of the Pan and Gorilla lineages, adding many new megabase pairs of DNA that altered the chromosome and chromatin architecture of these two species compared with all other primates. We propose that the orangutan and human genomes represent the hominid archetype, while the African ape genomes are more structurally derived with respect to these properties.

Methods

FISH

BAC and human fosmids clones (n = 1022) were used as probes to develop a comparative cytogenetic framework and to test rearrange-

ments specific to the gorilla lineage. Ancestral state was determined based on comparison with other primate species. Metaphases from nonhuman primates were obtained from lymphoblastoid or fibro-

blast cell lines of the following species: common chimpanzee (Pan troglodytes, PTR); gorilla (Gorilla gorila, GGO), and Borneo orangutan (Pongo pygmaeus pygmaeus, PPY) as representative of great apes; and thersus monkey (Macaca mulatta, MMU, Cercopithecinae) as repre-

sentative of Old World Monkeys. FISH experiments were essentially performed as previously described (Ventura et al. 2003).

Gorilla genome sequencing

Peripheral blood DNA was isolated from a male silverback gorilla, Kwan (Studbook #1107, b. 02/03/1989), housed at the Lincoln Park Zoo. Paired-end whole-genome sequence data were generated on an Illumina Genome Analyzer II using a modified protocol (see
Figure 4. (Legend on previous page)
Large chromosomal rearrangements were visualized using Miropeats (Parsons 1995) and high quality finished sequence within duplicated regions. Rearrangements were visualized using in-house visualization tools (Kidd et al. 2009). A subset of clones was selected for complete insert sequencing: AC243004.1, AC242665.3, AC242655.3, AC242627.3, AC242595.2, AC243002.1, AC241241.3, AC241522.2, AC240968.2, AC240954.2, AC239638.4, AC243003.2, AC243178.1, AC240953.2, AC239379.3, AC239356.2, AC239357.3, AC239280.2, AC239360.3, AC239282.3, AC239362.3, AC239380.3, AC239639.1, AC239363.3, AC239796.3, AC239381.3, AC239444.4, AC239382.3, AC239359.3, AC239358.3, AC239361.3, AC239281.3, AC239393.3, AC239640.2.

**Data access**

Gorilla sequence data (western lowland) have been deposited into the SRA under accession SRP002878.

We designed two oligonucleotide microarrays GEO accession number: GSE27072; samples: GSM665036, GSM665334, GSM665336, GSM665992, GSM665993, GSM667894, and GSM668114; and platforms GGO 2.1 custom: GPL11674 and Human 2.1 standard: GPL9684.


**Acknowledgments**

We thank Peter Sudmant and Tonia Brown for valuable comments in the preparation of this manuscript. We are grateful to Lisa Faust, Mollie Herget, and the staff of the Lincoln Park Zoo for the gorilla material used in this study. We are also indebted to the large genome sequencing centers for early access to BAC end sequence data for targeted analysis of structural rearrangements. This work was supported, in part, by an NIH grant HG002385 to E.E.E. and NIH grant U54 HG003079 to R.K.W. Financial support was provided by the Spanish Ministry of Science and Innovation (Grant BU2009-13409-C02-02) and the Spanish National Institute for Bioinformatics (INB, www.inab.org) to A.N. T.M-B. is supported by a Ramon y Cajal program (MCINN-RYC) and the European FP7 (MC/StGr 2010). Resources for exploring the gorilla genome structural variation online are available at University of Washington Genome Sciences (http://gorillaparalogy.gs.washington.edu). E.E.E. is an investigator of the Howard Hughes Medical Institute.

**Table 4. Summary of gorilla genome structural variation**

<table>
<thead>
<tr>
<th>SV type</th>
<th># Intervals</th>
<th># Base pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duplications</td>
<td>1,258</td>
<td>87,103,094</td>
</tr>
<tr>
<td>GGO specific</td>
<td>88</td>
<td>6,813,344</td>
</tr>
<tr>
<td>Deletions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥50 kb</td>
<td>44</td>
<td>6,298,067</td>
</tr>
<tr>
<td>&lt;50 kb</td>
<td>1,820</td>
<td>6,290,005</td>
</tr>
<tr>
<td>Mobile elements*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alu</td>
<td>4,274</td>
<td>1,325,597</td>
</tr>
<tr>
<td>L1</td>
<td>299</td>
<td>872,642</td>
</tr>
<tr>
<td>LTR</td>
<td>123</td>
<td>125,879</td>
</tr>
<tr>
<td>SVA</td>
<td>325</td>
<td>450,450</td>
</tr>
<tr>
<td>PTRRV1</td>
<td>263</td>
<td>2,033,779</td>
</tr>
<tr>
<td>Large chromosomal rearrangements</td>
<td>1</td>
<td>NA</td>
</tr>
<tr>
<td>Fusion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inversions</td>
<td>9b</td>
<td>NA</td>
</tr>
<tr>
<td>Translocation</td>
<td>1</td>
<td>NA</td>
</tr>
<tr>
<td>Duplicative transpositions</td>
<td>418b</td>
<td>NA</td>
</tr>
</tbody>
</table>

*Alu calls include FRAM element (2); and LTR includes THIE1 (11). The number of base pairs is estimated with the length of the consensus sequence of the predicted retrotransposon subclass.

*Only full-length sequence considered.

*Forty-seven of 424 validated by FISH, with 20/47 validated by sequence (Supplemental Table S3).

Supplemental Note). Sequence data have been deposited into the SRA under accession SRP002878.

**BAC sequencing**

End sequences from a gorilla BAC library (CH277) were retrieved from the NCBI Trace Archive and mapped to the human reference genome (NCBI35) to identify and clone rearrangement breakpoints as described previously (Newman et al. 2005). A subset of clones was selected for complete insert sequencing using capillary sequencing methods (McPherson et al. 2001) in order to obtain high quality finished sequence within duplicated regions. Rearrangements were visualized using Miropeats (Parsons 1995) and previously described in-house visualization tools (Kidd et al. 2010).

**Structural variation discovery**

Gorilla sequence reads were aligned to the human reference genome using the mrFILLD and mrFILLD mapping algorithms (Alkan et al. 2009; Hach et al. 2010). Deletions and mobile element insertions were detected using VariationHunter (Hormozdiari et al. 2009; Hormozdiari et al. 2010), while SDs (>20 kb) were detected and copy number quantified using measures of read-depth (see Supplemental Note; Alkan et al. 2009).

**ArrayCGH**

We designed two oligonucleotide microarrays (n = 385,000) targeted to regions of gorilla deletions and duplications and performed cross-species arrayCGH as previously described (GEO accession numbers: GSE27072; samples: GSM665036, GSM665334, GSM665336, GSM665992, GSM665993, GSM667894, and GSM668114; and platforms GGO 2.1 custom: GPL11674 and Human 2.1 standard: GPL9684).

**PCR**

Thirty PCR assays were designed to test the specificity and polymorphism of predicted Alu insertions in the gorilla genome. We only tested loci not embedded within other repetitive elements or SDs to facilitate reliable primer design.

**References**


