LETTERS

A burst of segmental duplications in the genome of the African great ape ancestor

Tomas Marques-Bonet[1,2], Jeffrey M. Kidd[1], Mario Ventura[3], Tina A. Graves[4], Ze Cheng[1], LaDeana W. Hillier[4], Zhaoshi Jiang[1], Carl Baker[1], Ray Malfavon-Borja[1], Lucinda A. Fulton[4], Can Alkan[1], Gozde Aksay[1], Santhosh Girirajan[1], Priscillia Siswara[1], Lin Chen[1], Maria Francesca Cardone[3], Arcadi Navarro[2,5], Elaine R. Mardis[4], Richard K. Wilson[4] & Evan E. Eichler[1]

It is generally accepted that the extent of phenotypic change between human and great apes is dissonant with the rate of molecular change[1]. Between these two groups, proteins are virtually identical[1,2], cytogenetically there are few rearrangements that distinguish ape–human chromosomes[3], and rates of single-base-pair change[4–7] and retrotransposon activity[8–10] have slowed particularly within hominid lineages when compared to rodents or monkeys. Studies of gene family evolution indicate that gene loss and gain are enriched within the primate lineage[11,12]. Here, we perform a systematic analysis of duplication content of four primate genomes (macaque, orang-utan, chimpanzee and human) in an effort to understand the pattern and rates of genomic duplication during hominid evolution. We find that the ancestral branch leading to human and African great apes shows the most significant increase in duplication activity both in terms of base pairs and in terms of events. This duplication acceleration within the ancestral species is significant when compared to lineage-specific rate estimates even after accounting for copy-number polymorphism and homoplasy. We discover striking examples of recurrent and independent gene-containing duplications within the gorilla and chimpanzee that are absent in the human lineage. Our results suggest that the evolutionary properties of copy-number mutation differ significantly from other forms of genetic mutation and, in contrast to the hominid slowdown of single-base-pair mutations, there has been a genomic burst of duplication activity at this period during human evolution.

We began by developing a segmental duplication map for each of the four primate genomes (macaque, orang-utan, chimpanzee and human; Supplementary Fig. 1). The approach is based on the alignment of whole-genome shotgun (WGS) sequence data against the human reference genome and predicts high-identity segmental duplications based on excess depth of coverage and sequence divergence[13] (Methods). Previous analyses have suggested excellent sensitivity and specificity for computational detection of duplications larger than 20 kilobases (kb) in length[13] (Table 1, Supplementary Table 1 and Supplementary Note Table 2). By this criterion, we characterized 73 megabases (Mb) corresponding to the duplications identified in at least one of the four primate species, correcting for copy number in each primate (Methods). Furthermore, we characterized each duplication as 'lineage specific' or 'shared', depending on whether it was seen in only one or multiple genomes. This comparative map (Supplementary Figs 3 and 4) is available as an interactive UCSC mirror browser (http://humanparalogy.gs.washington.edu), allowing researchers to interrogate the evolutionary history of any duplicated region of interest.

We validated our primate genomic duplication map using two different experimental approaches and, wherever possible, using DNA from the same individuals from which the computational predictions were generated. Using fluorescence *in situ* hybridization (FISH), we found that 86.5% of segmental duplications were concordant with computational predictions when categorized as either lineage specific (50 out of 58) or shared duplications (40 out of 46) (Supplementary Figs 1 and 2; see also Fig. 1 and Supplementary Tables 2–4). As a second approach, we designed a specialized oligonucleotide microarray (1 probe per 585 bp) targeted to primate segmental duplications

Table 1 | Classes of primate segmental duplication

Category	Segmental duplications	Segmental duplications >20 kb	Validation (%)	Copy-number-corrected duplicated base pairs			
				Hsa	Ptr	Ppy	Mmu
Hsa	51,458,805	15,236,422	89–92	17,847,869	–	–	–
Ptr	11,239,390	4,789,874	99	–	16,583,946	–	–
Ppy	30,553,228	6,417,679	98	–	–	23,327,737	–
Mmu	24,962,092	5,360,646	45	–	–	–	45,810,964
Mmu*	35,493,466	7,715,410	85	–	–	–	18,266,656
Hsa/Ptr	32,392,480	21,061,194	NA	21,524,417	26,304,286	–	–
Hsa/Ptr/Ppy	25,450,827	13,402,545	NA	11,259,061	14,012,351	11,541,148	–
Hsa/Ptr/Ppy/Mmu	14,094,156	7,156,616	NA	8,092,997	12,820,607	6,176,876	12,542,691
Total	190,150,978	73,424,976	–	58,724,344	69,721,190	41,045,761	30,809,347

Duplications were divided into eight categories based on the WSSD analysis of each primate genome (subsequent analyses were restricted to segmental duplications >20 kb in length). Lineage-specific and shared duplication content are indicated. Percentage validation indicates the fraction of species-specific duplications confirmed by cross-species array comparative genomic hybridization. Because the human genome was used, we corrected for copy number and examined sequence contigs not aligned to the human genome (see Methods). Segmental duplications assigned to the Y chromosome were not considered.

* Macaque segmental duplications detected in the macaque reference genome using WSSD and WGAC (<94% identity) approaches.

[1]Department of Genome Sciences, University of Washington and the Howard Hughes Medical Institute, Seattle, Washington 98195, USA. [2]Institut de Biologia Evolutiva (UPF-CSIC), 08003 Barcelona, Catalonia, Spain. [3]Sezione di Genetica-Dipartimento di Anatomia Patologica e Genetica, University of Bari, 70125 Bari, Italy. [4]Genome Sequencing Center, Washington University School of Medicine, St Louis, Missouri 63108, USA. [5]Institució Catalana de Recerca i Estudis Avançats (ICREA) and Instituto Nacional de Bioinformática (INB), Dr. Aiguader 88, 08003 Barcelona, Spain.

©2009 Macmillan Publishers Limited. All rights reserved

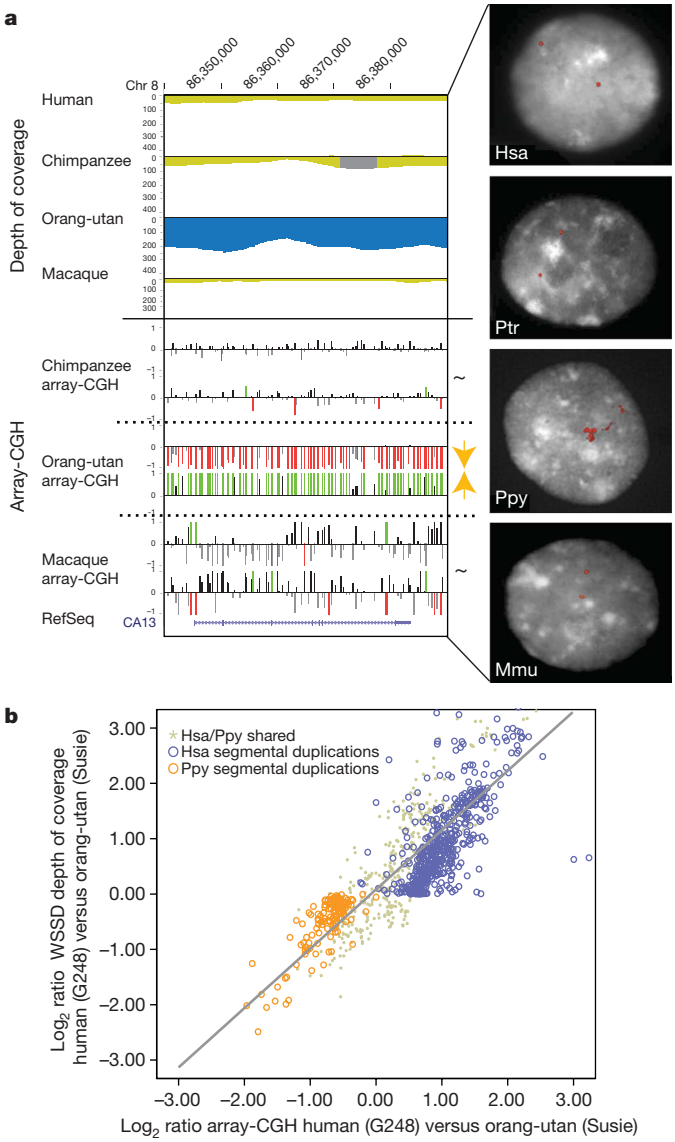

Figure 1 | Experimental validation of duplication map. **a**, A computationally predicted orang-utan-specific duplication (blue, excess depth of coverage of aligned WGS sequence) is confirmed by interspecific FISH and array-CGH (oligonucleotide relative $\log_2$ ratios are depicted as red/green histograms and correspond to an increase and decrease in signal intensity when test/reference are reverse labelled; see Supplementary Note for additional details). Hsa, *Homo sapiens* (human); Mmu, *Macaca mulatta* (macaque); Ptr, *Pan troglodytes* (chimpanzee); Ppy, *Pongo pygmaeus* (orang-utan). **b**, A comparison of duplication copy number as predicted by WSSD sequence analysis versus oligonucleotide array-CGH between nonhuman and human primates showed a good correlation ($r = 0.77$). Relative duplication copy number was computed as the $\log_2$ ratio of the number of aligned nonhuman primate reads against the human reference genome over the number of reads mapping to known single-copy BACs.

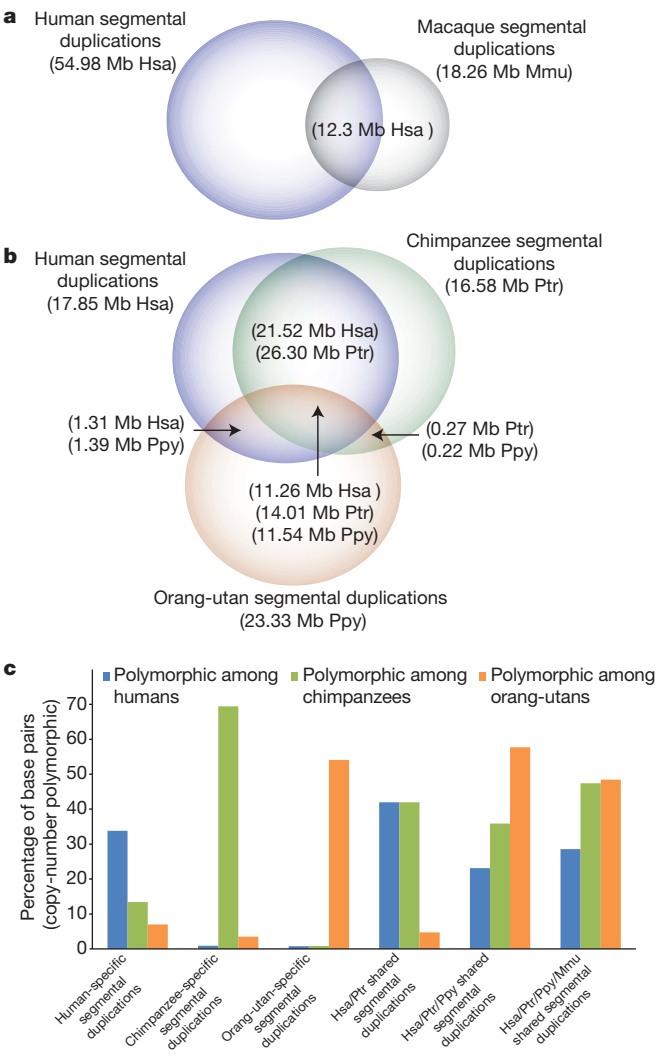

Figure 2 | Shared versus lineage-specific duplications and great-ape polymorphism. Segmental duplications (>20 kb) were classified as lineage specific or shared based on a four-way comparison of human, chimpanzee, orang-utan and macaque genomes. **a**, Human segmental duplications were compared to Old World monkey segmental duplications (based on a separate analysis of the macaque assembly[9]). **b**, As nonhuman great-ape duplications were detected based on alignment of WGS sequence against the human genome, we corrected for copy number based on the depth of coverage of WGS sequence (in brackets with the name of the species for which the correction was performed, see Table 1). **c**, Copy-number polymorphic regions were estimated from the results of array-CGH hybridizations between eight samples each of human, chimpanzee and orang-utan (using the same reference as the computational prediction). The proportion of duplicated bases that showed evidence of copy-number polymorphism (that is, gain or loss for ≥two individuals within each species) is shown for each class of segmental duplication (>20 kb).

(Table 1) and performed array comparative genomic hybridization (array-CGH) between species (Table 1, Fig. 1 and Supplementary Figs 2–4). Among the great-ape genomes, we confirmed 89–99% of the lineage-specific duplications by interspecific array-CGH (Table 1) with a very good correlation between computationally predicted and experimentally validated copy-number differences (Fig. 1b). Because only 45% of macaque-specific duplications could be confirmed by interspecific array-CGH, we performed an independent assessment of the macaque genome assembly and conservatively validated ~85% of macaque-specific duplications[9,14] (Z.J. and E.E.E., unpublished results).

The comparative duplication map reveals several important features of primate segmental duplications. As expected, most (80% or ~55 Mb) high-identity human segmental duplications arose after the divergence of the Old World monkey and hominoid lineages (Fig. 2a). Humans and chimpanzees show significantly more duplications than either macaque or orang-utan (Fig. 2b), with a large fraction being shared between chimpanzee and human. On the basis of our four-way primate genome analysis and leveraging array-CGH data from gorilla and bonobo (*Pan paniscus*), we classify only ~10 Mb of duplication content as human specific (210 duplication intervals with an average length of 53.1 kb). The genomic distribution of great-ape segmental duplications is highly nonrandom (Supplementary Fig. 5), with the presence of ancestral duplications being a strong predictor of 'new', lineage-specific events ($P < 0.001$, randomization test,

©2009 Macmillan Publishers Limited. All rights reserved

Supplementary Note Table 5a, b). For example, 45% of human–chimpanzee shared duplications map within 5 kb of segmental duplications shared among human–chimpanzee–orang-utan, whereas 31% of human–chimpanzee–orang-utan duplications map adjacent to human–chimpanzee–orang-utan–macaque duplications. These observations emphasize that unique sequences flanking more ancient duplications have a much higher probability of segmental duplication[13,15] and the duplication process itself is not random.

Within the human-specific set of duplications, we identified 39 partial and 17 complete human genes (Supplementary Table 7). As expected, we found that full-length hominid genes show greater evidence of positive selection when compared with similarly analysed unique genes (Supplementary Note). Our analysis indicates that several genes associated with human adaptation (amylase (*AMY1*), aquaporin 7 and *NBPF15*) are shared with chimpanzee but humans show a general increase in copy number. Gene models associated with signal transduction, neuronal activities (for example, neurotransmitter release, synaptic transmission) and muscle contraction are significantly enriched in human, chimpanzee and orang-utan lineage-specific duplications (Supplementary Table 7). Human and great-ape shared duplications or those shared with macaque are, in contrast, enriched for biological processes associated with amino acid metabolism ($P = 1.69 \times 10^{-2}$; great-ape shared segmental duplications) or oncogenesis ($P = 5.80 \times 10^{-13}$, 4.64×10^{-6}; ape segmental duplications shared with macaque). Although the number of such duplication events is few, these data suggest a shift in the types of genes that have been duplicated most recently during great-ape and human evolution.

There are two important caveats to the above analysis. First, we have analysed a single individual in each case and it is unclear to what extent that single genome represents the duplication pattern of the species. Second, duplicated sequences shared by two or more species might have potentially been subjected to recurrent mutations (homoplasy) leading to an overestimate of the proportion of ancestral duplications. Both copy-number polymorphism and recurrent duplication, in principle, will complicate classification of segmental duplications as 'ancestral' or 'lineage specific'. We therefore performed a number of additional analyses to address the impact of polymorphism and recurrent events on our assignments.

First, we investigated the extent of copy-number variation for both shared and lineage-specific duplications. Using array-CGH targeted to primate segmental duplications, we assessed the extent of copy-number variation in a set of unrelated DNA samples (Fig. 2c; see Methods). As expected[16,17], lineage-specific segmental duplications are highly copy-number variant, with humans showing 1.5- to two-fold less diversity in copy number when compared to chimpanzees and orang-utans (Fig. 2c; see also Supplementary Note Table 9). Notably, we found that shared segmental duplications are as copy-number variant as lineage-specific duplications and that humans show slightly greater copy-number variation for these (42% versus 34%) when compared with great apes.

It is, however, important to distinguish between duplication copy-number variation versus duplication status. A segmental duplication may show a high level of copy-number variation whereas its status as duplicated remains relatively constant among different individuals within a species. To address this, we performed a series of three-way array-CGH comparisons (Supplementary Note Fig. 7; see also Methods) where we investigated how duplication status (human-specific, chimpanzee-specific and orang-utan-specific segmental duplications) varied as function of copy-number polymorphism within a species. The results from these triangulations indicate that only 1–8% of the segmental duplications change duplication status even though 18–32% of the duplications are copy-number polymorphic between two

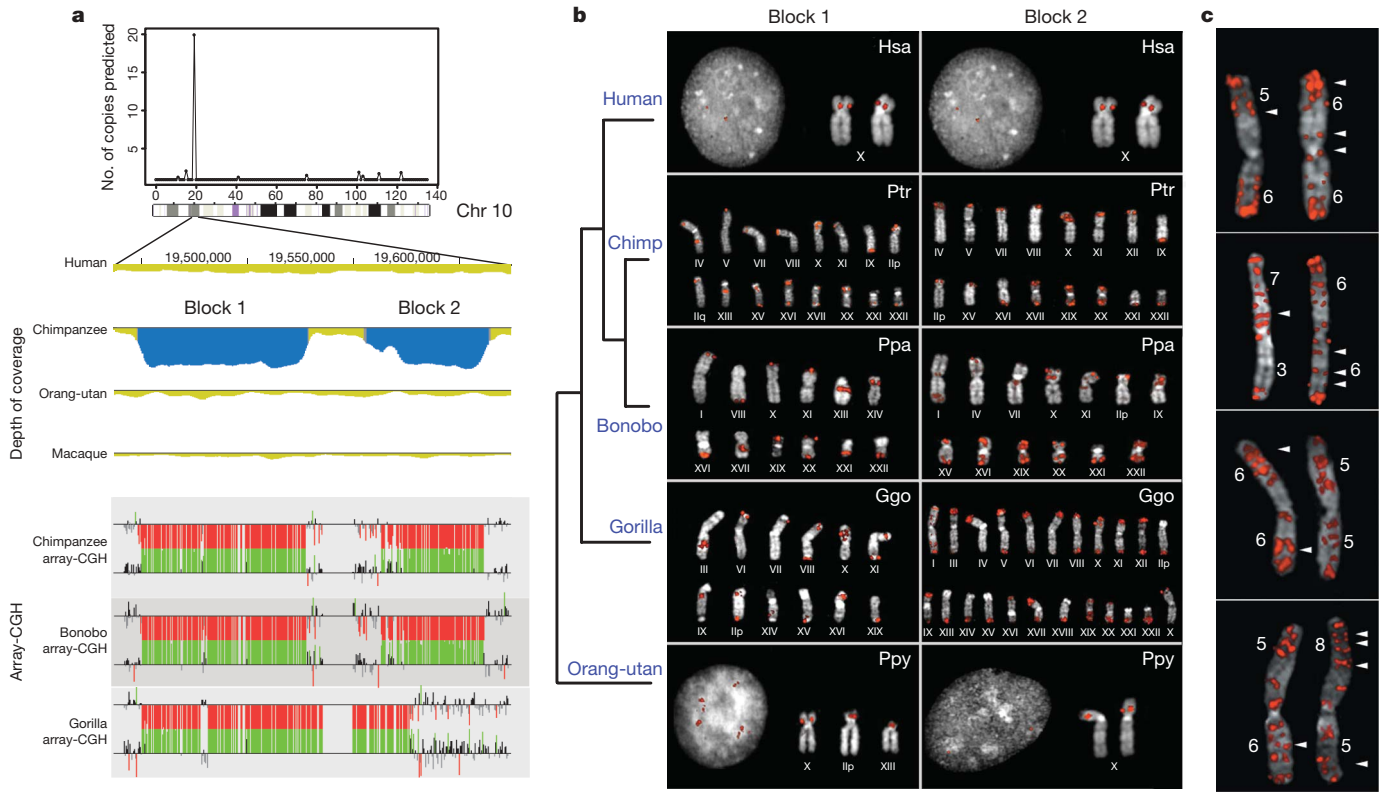

Figure 3 | Convergent gene duplication expansion in African great apes but not humans. **a**, Two regions on chromosome 10 have expanded in chimpanzee, gorilla and bonobo when compared to human based on computational and interspecific array-CGH experiments (see Fig. 1 legend). **b**, FISH confirms 23–50 copies in chimpanzee and bonobo (Ppa, *Pan paniscus*), and >100 copies in gorilla (Ggo, *Gorilla gorilla*) (Methods). End-sequence pair analysis using gorilla and chimpanzee WGS sequences reveals that all but the ancestral location are non-orthologous, indicating independent expansions in both lineages. **c**, Detailed analysis of eight homologues of gorilla chromosome 1 reveals interstitial locations of the block 2 duplication that show variation both in copy number and in terms of location.

©2009 Macmillan Publishers Limited. All rights reserved

individuals within a species (Supplementary Note Fig. 8). As a second independent test, we compared the duplication maps of two human genomes (J. C. Venter or HuRef and J. D. Watson genomes)[18,19] and found that 89% (595 out of 666) of the regions are shared duplications between HuRef and the J. D. Watson genome. Although we predict copy-number differences between these shared duplications, the boundaries of the duplication intervals remain remarkably consistent (Supplementary Fig. 7), suggesting again that duplication status is a relatively constant character state within a species.

To assess the potential impact of recurrent mutations leading to misclassification of ancestral events, we focused on shared duplications between human and chimpanzee that were not identified as duplicated in either orang-utan or macaque. We examined 103 sets of chimpanzee–human shared duplications that mapped to two or more distinct locations in the human genome (Supplementary Note) and determined what fraction of these mapped to two or more orthologous positions between chimpanzee and human. Using a paired end-sequence mapping approach[20,21] (Supplementary Note Fig. 9), we found that 85% (88 out of 103) of the chimpanzee–human shared duplications have two or more copies mapping to the same orthologous position in the two genomes. This implies that most of the shared duplications were already duplicated in the human–chimpanzee common ancestor (Supplementary Note Tables 6 and 7).

As part of our comparative analyses, we identified regions for which duplication patterns were inconsistent with the generally accepted human/great-ape phylogeny (Supplementary Fig. 4 and Supplementary Tables 5 and 6). For example, we identified 43 intervals that are duplicated in human and gorilla but not chimpanzee ($H^+C^-G^+$ duplications). Such a scenario may arise as a result of a deletion event in the chimpanzee lineage, incomplete lineage sorting or, less likely, recurrent duplication events in the human and gorilla lineages. Only the latter possibility would potentially lead to an overestimation of ancestral duplication events. We estimated the frequency of such events by mapping the location of the duplications in each species using paired end-sequence data[21] (see Supplementary Note). If the duplicated sequence mapped to the same location in gorilla and human, we classified it as a chimpanzee-specific deletion event or incomplete lineage sorting. If mapping to different locations in the two genomes, we categorized it as a recurrent event. As expected, most of the informative $H^+C^-G^+$ duplications (80% or 12 out of 15) were the result of chimpanzee-specific deletions.

We investigated the most extreme example of recurrent African ape duplications in more detail (Fig. 3). We identified a region (~150 kb in length) mapping to human chromosome 10 that had expanded in the chimpanzee genome but was largely single copy in human and orang-utan. It consists of two distinct duplication blocks (~86 and 66 kb in length). Both array-CGH and fluorescent *in situ* hybridization (FISH; Fig. 3a, b) confirm that the segments had been duplicated multiple times (~5–100 copies depending on the block and species) in the chimpanzee, bonobo and gorilla genomes but are single copy in all humans tested. Notably, the duplication boundaries (as delimited by array-CGH) differ between the gorilla and chimpanzee lineages. With the exception of the chromosome 10 locus, we found that the map locations between gorilla and chimpanzee are non-orthologous (Supplementary Note and Methods), indicating that this duplication expansion has occurred independently in both lineages.

On the basis of the large number of interstitial sites on gorilla chromosomes, we compared chromosome 1 from four unrelated gorillas for variation in copy number and location of this segmental duplication. Remarkably, we found that both copy number (10–14 copies per homologous chromosome) as well as map location for this segmental duplication vary among these eight gorilla homologues with as many as 50% of the map locations being unoccupied by a duplication in another homologue (Fig. 3c and Supplementary Fig. 13). We conclude that this ancestral region of chromosome 10 has served as a preferred donor of chimpanzee/great-ape duplications and that the chimpanzee and gorilla genomes have been restructured by independent bursts of duplication activity. Interestingly, we detected and confirmed by RT–PCR (reverse transcription PCR) at least one previously uncharacterized gene (14 exons, 141 kb of genomic sequence, 1,311 nucleotides of coding sequence and 437 amino acids) mapping to duplication block 1, which shows significant similarity to endosomal glycoprotein genes (Supplementary Note Figs 14–17). Thus, these duplications, in principle, may have led to African great-ape gene family expansions while remaining conspicuously a single copy in the human lineage. Although the mechanism by which such events have occurred is unclear, our data highlight the rapidity by which segmental duplications have restructured hominid genomes and emphasize their nonrandom nature both temporally and spatially.

Based on our genome-wide assessment of segmental duplications in each of four primate species and our estimate of 20% homoplasy

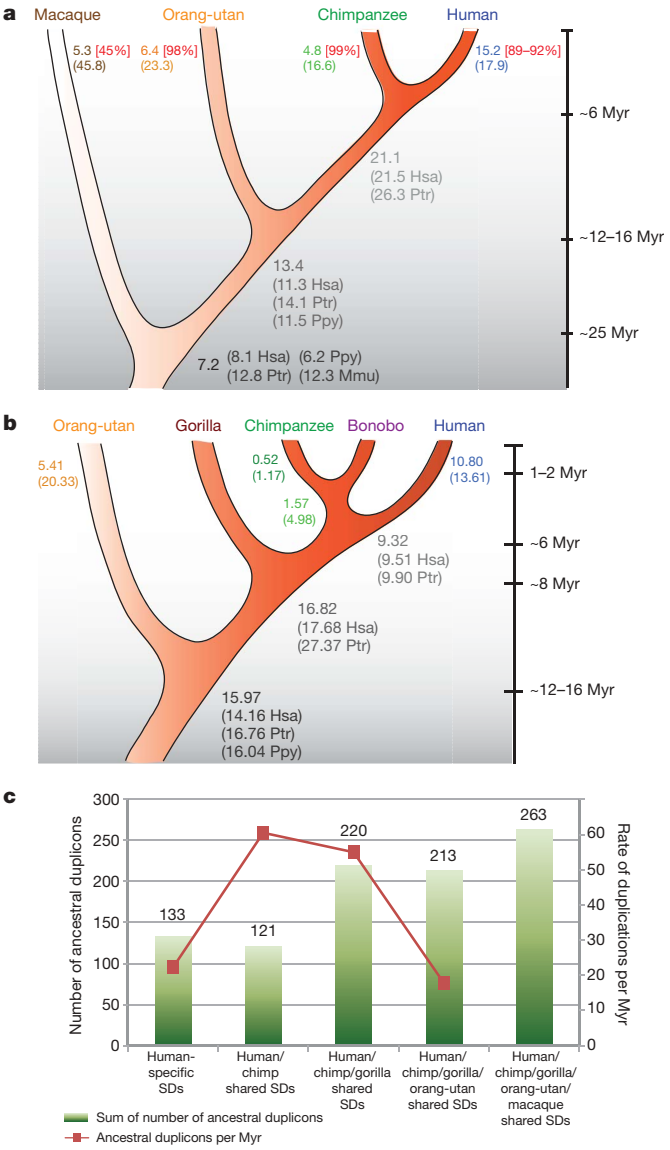

Figure 4 | Rates of segmental duplication. **a**, By base pair: we parsimoniously assigned the number of megabases to different branches, correcting for copy number in each species (shown in brackets). 89–99% of great-ape segmental duplications (SDs) were validated by array-CGH (square brackets). **b**, Further categorization of the segmental duplication data, based on array-CGH from bonobo and gorilla, shows the greatest accumulation in the ancestor of humans and great apes. **c**, By event: we assigned 950 evolutionarily distinct human segmental duplication events[22] to the human/great-ape phylogeny based on array-CGH results. The red line estimates the duplication rate (per million years (Myr)) and suggests an excess of large duplications in the common ancestor of human and chimpanzee but after the separation from gorilla.

©2009 Macmillan Publishers Limited. All rights reserved

(see above), we calculated rates of segmental duplication both in events[22] and base pairs along each lineage and ancestral node (Fig. 4, Supplementary Note Tables 13–16). We developed a maximum likelihood model to test if the rate of accumulation of segmental duplication has remained constant during the course of human/great-ape evolution. We compared the likelihood that the rate of segmental duplication has been uniform versus the likelihood of differential rates within specific lineages (Fig. 4). We find a significant increase (likelihood ratio test (LRT), $P < 1 \times 10^{-10}$) in both the number of events and base pairs in the human/African great-ape lineage when compared to macaque/Old World monkey lineage. Although terminal hominid lineages show an excess of duplications, the most significant burst of activity (4–10-fold, LRT $P = 1 \times 10^{-10}$) occurs in the common ancestor of human/chimpanzee and gorilla and after divergence of gorilla from the human–chimpanzee lineage (Supplementary Note Table 17). Our prediction is in strong agreement with the degree of sequence divergence among human intrachromosomal segmental duplications that shows a mode at 97–99% sequence identity. We note that this burst of duplication activity corresponds to a time when other mutational processes, such as point substitutions and retrotransposon activity, were slowing along the hominoid lineage. This apparent burst of activity may be the result of changes in the effective population size, generation time or imply a genomic destabilization at a period before and perhaps during hominid speciation. In light of the importance of segmental duplications in contributing to copy-number changes associated with neurocognitive disease[23–26] and disease susceptibility[27–29], we predict that this apparent acceleration has had a profound impact on the reproductive success, adaptability and evolution of ancestral hominid populations.

METHODS SUMMARY

We estimated the duplication content of human, chimpanzee, orang-utan and macaque by the whole-genome shotgun sequence detection (WSSD) method[13,30]. We mapped high-quality whole-genome shotgun (WGS) sequence reads for all species against the human reference assembly (NCBI build35) and identified regions of excess depth of coverage and divergence (see Supplementary Note). We also mapped macaque WGS reads to the macaque assembly (v 1.0). In this analysis, we considered segmental duplications >20 kb and >94% of identity (88% of identity for macaque reads against the human genome). We used read depth to estimate the number of copies for each duplication due to the excellent correlation ($r^2 = 0.953$)[13] between probes of known copy number and WGS depth of coverage.

We constructed an oligonucleotide microarray ($n = 385{,}000$) targeted to regions of primate segmental duplication (~180 Mb) and performed cross-species array-CGH (with human as a reference; GEO accession number GSE13884). With the exception of human, we used DNA derived from the same genome that was sequenced as part of primate genome sequencing projects. The same microarray was used to assess copy-number polymorphism in DNA samples from eight humans, eight chimpanzees and eight orang-utans (GSE13885). We also used FISH to validate further a subset of our duplications among the great apes.

We used end-sequence pair data from fosmid clones from a single human and a single chimpanzee as well as plasmid clones from a gorilla to map the location of segmental duplications within great-ape genomes (sequence data available from NIH trace repository). To estimate rates of segmental duplication along the hominoid phylogeny, we modelled the accumulation of segmental duplications in each branch as a pure birth process within a maximum likelihood framework. Nested models of segmental duplication were tested against each other by means of likelihood ratio tests (Supplementary Note).

Received 29 August; accepted 18 December 2008.

1. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188,** 107–116 (1975).
2. Goodman, M. The role of immunochemical differences in the phyletic development of human behavior. *Hum. Biol.* **33,** 131–162 (1961).
3. Yunis, J. J. & Prakash, O. The origin of man: a chromosomal pictorial legacy. *Science* **215,** 1525–1530 (1982).
4. Wu, C. I. & Li, W. H. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl Acad. Sci. USA* **82,** 1741–1745 (1985).
5. Li, W. H. & Tanimura, M. The molecular clock runs more slowly in man than in apes and monkeys. *Nature* **326,** 93–96 (1987).
6. Elango, N., Thomas, J. W. & Yi, S. V. Variable molecular clocks in hominoids. *Proc. Natl Acad. Sci. USA* **103,** 1370–1375 (2006).
7. Steiper, M. E., Young, N. M. & Sukarna, T. Y. Genomic data support the hominoid slowdown and an Early Oligocene estimate for the hominoid-cercopithecoid divergence. *Proc. Natl Acad. Sci. USA* **101,** 17021–17026 (2004).
8. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).
9. Rhesus Macaque Genome Sequencing and Analysis Consortium. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316,** 222–234 (2007).
10. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437,** 69–87 (2005).
11. Hahn, M. W., Demuth, J. P. & Han, S. G. Accelerated rate of gene gain and loss in primates. *Genetics* **177,** 1941–1949 (2007).
12. Dumas, L. *et al.* Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res.* **17,** 1266–1277 (2007).
13. Cheng, Z. *et al.* A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **437,** 88–93 (2005).
14. Jiang, Z., Hubley, R., Smit, A. & Eichler, E. E. DupMasker: A tool for annotating primate segmental duplications. *Genome Res.* **18,** 1362–1368 (2008).
15. Stankiewicz, P., Shaw, C. J., Withers, M., Inoue, K. & Lupski, J. R. Serial segmental duplications during primate evolution result in complex human genome architecture. *Genome Res.* **14,** 2209–2220 (2004).
16. Perry, G. H. *et al.* Hotspots for copy number variation in chimpanzees and humans. *Proc. Natl Acad. Sci. USA* **103,** 8006–8011 (2006).
17. Lee, A. S. *et al.* Analysis of copy number variation in the rhesus macaque genome identifies candidate loci for evolutionary and human disease studies. *Hum. Mol. Genet.* **17,** 1127–1136 (2008).
18. Levy, S. *et al.* The diploid genome sequence of an individual human. *PLoS Biol.* **5,** e254 (2007).
19. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452,** 872–876 (2008).
20. Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nature Genet.* **37,** 727–732 (2005).
21. Newman, T. L. *et al.* A genome-wide survey of structural variation between human and chimpanzee. *Genome Res.* **15,** 1344–1356 (2005).
22. Jiang, Z. *et al.* Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. *Nature Genet.* **39,** 1361–1368 (2007).
23. Lee, J. A. & Lupski, J. R. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron* **52,** 103–121 (2006).
24. Sharp, A. J. *et al.* Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nature Genet.* **38,** 1038–1042 (2006).
25. The International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455,** 237–241 (2008).
26. Sebat, J. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316,** 445–449 (2007).
27. Aitman, T. J. *et al.* Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* **439,** 851–855 (2006).
28. Hollox, E. J. *et al.* Psoriasis is associated with increased β-defensin genomic copy number. *Nature Genet.* **40,** 23–25 (2008).
29. Gonzalez, E. *et al.* The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307,** 1434–1440 (2005).
30. Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297,** 1003–1007 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Mefford, A. Itsara, G. Cooper, T. Brown and G. McVicker for comments during the preparation of this manuscript. The authors are also grateful to J. Sikela and L. Dumas for assistance with the comparison to cDNA microarray data sets. We are grateful to L. Faust, J. Rogers, Southwest National Primate Research Center (P51-RR013986) and P. Parham for providing some of the primate material used in this study and to M. Adams for providing the alignments for the positive selection analysis. We also thank the large genome sequencing centres for early access to the whole genome sequence data for targeted analysis of segmental duplications. This work was supported, in part, by an NIH grant HG002385 to E.E.E. and NIH grant U54 HG003079 to R.K.W. and E.R.M. INB is a platform of Genoma España. T.M.-B. is supported by a Marie Curie fellowship and by Departament d'Educació i Universitats de la Generalitat de Catalunya. E.E.E. is an investigator of the Howard Hughes Medical Institute.

Author Contributions E.E.E. planned the project. M.V. and M.F.C. performed the FISH experiments. T.A.G., L.W.H., L.A.F., E.R.M. and R.K.W. generated the orang-utan WGS sequences. T.M.-B., J.M.K., Z.C., Z.J., L.C., E.E.E. and S.G. analysed the data. C.B. performed the array-CGH experiments. T.M.-B., R.M.-B. and P.S. characterized the chromosome 10 expansion. C.A. and G.A. generated the Venter/Watson comparative duplication maps. A.N. developed the maximum likelihood evolutionary model. T.M.-B., J.M.K. and E.E.E. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.E.E. (eee@gs.washington.edu).

©2009 Macmillan Publishers Limited. All rights reserved