Great ape genetic diversity and population history

Javier Prado-Martinez1, Peter H. Sudmant2, Jeffrey M. Kidd1–4, Heng Li5, Joanna L. Kelley3, Belen Lorente-Galdos4, Krishna R. Veeramah6, August E. Woerner6, Timothy D. O’Connor2, Gabriel Santpere1, Alexander Cagan3, Christoph Theunert1, Ferran Casals1, Hafid Laayouni1, Kasper Munch1, Anders E. Halager5, Maika Malig2, Jessica Hernandez-Rodriguez1, Irene Hernando-Herraez2, Kay Prüfer2, Marc Pybus2, Laurel Johnstone6, Michael Lachmann7, Tomas Marques-Bonet1,3, Laurel Johnstone2, Michael Lachmann7, David Reich4, 1Institut de Biologia Evolutiva, (CSIC-Universitat Pompeu Fabra), PRBB, Doctor Aiguader 88, Barcelona, Catalonia 08003, Spain. 2Department of Human Genetics, University of Michigan, 1241 E. Catherine Street, Ann Arbor, Michigan 48109, USA. 3Department of Genetics, Stanford University, 300 Pasteur Drive, Lane L301, Stanford, California 94305, USA. 4Department of Genetics, Harvard Medical School, Boston, 77 Avenue Louis Pasteur, Massachusetts 02115, USA. 5Arizona Research Laboratories, Division of Biotechnology, University of Arizona, 1041 E. Lowell Street, Tucson, Arizona 85721, USA. 6Department of Evolutionary Genomics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig, 04103, Germany. 7Bioinformatics Research Centre, Aarhus University, DK-8000 Aarhus C, Denmark. 8Institut de Biologia Evolutiva, (CSIC-Universitat Pompeu Fabra), Barcelona, Catalonia 08003, Spain. 9Centro Nacional de Análisis Genómico (CNAG), PCB, Barcelona, Catalonia 08028, Spain. 10Department of Medical Genetics and Biotechnology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. 11National Institutes of Health Intramural Sequencing Center (NISC), Bethesda, Maryland 20892, USA. 12Department of Cell and Molecular Medicine, University of California San Diego, San Diego, California 92093, USA. 13Department of Bioinformatics, University of Copenhagen, Copenhagen 2200, Denmark. 14Centro Nacional de Análisis Genómico (CNAG), PDB, Barcelona, Catalonia 08028, Spain. 15Genome Sequencing Center, Washington University School of Medicine, St. Louis, Missouri 63108, USA. 16Department of Evolutionary Anthropology, Duke University, Durham, North Carolina 27708, USA. 17Limbo Wildlife Centre, BP 878, Limbo, Cameroon. 18Paul G. Allen School for Global Animal Health, Washington State University, Washington 99164, USA. 19North Carolina Zoological Park, Asheboro, North Carolina 27205, USA. 20Department of Psychology, Franklin and Marshall College, Lancaster, Pennsylvania 17604, USA. 21Department of Statistics, Oxford University, 1 South Parks Road, Oxford OX1 3TG, UK. 22Department of Genetics and Microbiology, University of Bar, Bari 70126, Italy. 23Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, California 92093, USA. 24Department of Biology, Bioinformatics, University of Copenhagen, Copenhagen 2200, Denmark. 25Centro Nacional de Análisis Genómico (CNAG), PDB, Barcelona, Catalonia 08028, Spain. 26Department for Bioscience, Aarhus University, DK-8000 Aarhus C, Denmark. 27Copenhagen Zoo, DK 2000 Frederiksberg, Denmark. 28Howard Hughes Medical Institute, Stanford 15790 Avenue NE, Seattle, Washington 98195, USA. *Present address: Centre for Genomic Regulation (CRG), C/D’Aiguader, 88, 08003 Barcelona, Spain.

©2013 Macmillan Publishers Limited. All rights reserved
and both orangutan species show the greatest genetic diversity (1.6 × 10^−3−2.4 × 10^−3 heterozygotes per bp). These differences are also reflected by measures of inbreeding from runs of homozygosity13 (Fig. 1c and Supplementary Information). Bonobos and western lowland gorillas, for example, have similar distributions of tracts of homozygosity as human populations that have experienced strong genetic bottlenecks (Karitiana and Papuan). Eastern lowland gorillas appear to represent the most inbred population, with evidence that they have been subjected to both recent and ancient inbreeding.

To examine the level of genetic differentiation between individuals we performed a principal component analysis (PCA) of SNP genotypes (Supplementary Information). Chimpanzees were stratified between subspecies with PC1 separating western and Nigeria–Cameroon chimpanzees from the eastern and central chimpanzees and PC2 separating western and Nigeria–Cameroon chimpanzees. In gorillas, PC1 clearly separates eastern and western gorillas, whereas the western lowland gorillas are distributed along a gradient of PC2, with individuals from the Congo and western Cameroon positioning in opposite directions along the axis. The isolated Cross River gorilla is genetically more differentiated from all gorilla groups, for example, have similar distributions of tracts of homozygosity as human populations that have experienced strong genetic bottlenecks (Karitiana and Papuan). Eastern lowland gorillas appear to represent the most inbred population, with evidence that they have been subjected to both recent and ancient inbreeding.

To examine the level of genetic differentiation between individuals we performed a principal component analysis (PCA) of SNP genotypes (Supplementary Information). Chimpanzees were stratified between subspecies with PC1 separating western and Nigeria–Cameroon chimpanzees from the eastern and central chimpanzees and PC2 separating western and Nigeria–Cameroon chimpanzees. In gorillas, PC1 clearly separates eastern and western gorillas, whereas the western lowland gorillas are distributed along a gradient of PC2, with individuals from the Congo and western Cameroon positioning in opposite directions along the axis. The isolated Cross River gorilla is genetically more differentiated from all gorilla groups, for example, have similar distributions of tracts of homozygosity as human populations that have experienced strong genetic bottlenecks (Karitiana and Papuan). Eastern lowland gorillas appear to represent the most inbred population, with evidence that they have been subjected to both recent and ancient inbreeding.

To examine the level of genetic differentiation between individuals we performed a principal component analysis (PCA) of SNP genotypes (Supplementary Information). Chimpanzees were stratified between subspecies with PC1 separating western and Nigeria–Cameroon chimpanzees from the eastern and central chimpanzees and PC2 separating western and Nigeria–Cameroon chimpanzees. In gorillas, PC1 clearly separates eastern and western gorillas, whereas the western lowland gorillas are distributed along a gradient of PC2, with individuals from the Congo and western Cameroon positioning in opposite directions along the axis. The isolated Cross River gorilla is genetically more differentiated from all gorilla groups, for example, have similar distributions of tracts of homozygosity as human populations that have experienced strong genetic bottlenecks (Karitiana and Papuan). Eastern lowland gorillas appear to represent the most inbred population, with evidence that they have been subjected to both recent and ancient inbreeding.

To examine the level of genetic differentiation between individuals we performed a principal component analysis (PCA) of SNP genotypes (Supplementary Information). Chimpanzees were stratified between subspecies with PC1 separating western and Nigeria–Cameroon chimpanzees from the eastern and central chimpanzees and PC2 separating western and Nigeria–Cameroon chimpanzees. In gorillas, PC1 clearly separates eastern and western gorillas, whereas the western lowland gorillas are distributed along a gradient of PC2, with individuals from the Congo and western Cameroon positioning in opposite directions along the axis. The isolated Cross River gorilla is genetically more differentiated from all gorilla groups, for example, have similar distributions of tracts of homozygosity as human populations that have experienced strong genetic bottlenecks (Karitiana and Papuan). Eastern lowland gorillas appear to represent the most inbred population, with evidence that they have been subjected to both recent and ancient inbreeding.

To examine the level of genetic differentiation between individuals we performed a principal component analysis (PCA) of SNP genotypes (Supplementary Information). Chimpanzees were stratified between subspecies with PC1 separating western and Nigeria–Cameroon chimpanzees from the eastern and central chimpanzees and PC2 separating western and Nigeria–Cameroon chimpanzees. In gorillas, PC1 clearly separates eastern and western gorillas, whereas the western lowland gorillas are distributed along a gradient of PC2, with individuals from the Congo and western Cameroon positioning in opposite directions along the axis. The isolated Cross River gorilla is genetically more differentiated from all gorilla groups, for example, have similar distributions of tracts of homozygosity as human populations that have experienced strong genetic bottlenecks (Karitiana and Papuan). Eastern lowland gorillas appear to represent the most inbred population, with evidence that they have been subjected to both recent and ancient inbreeding.

To examine the level of genetic differentiation between individuals we performed a principal component analysis (PCA) of SNP genotypes (Supplementary Information). Chimpanzees were stratified between subspecies with PC1 separating western and Nigeria–Cameroon chimpanzees from the eastern and central chimpanzees and PC2 separating western and Nigeria–Cameroon chimpanzees. In gorillas, PC1 clearly separates eastern and western gorillas, whereas the western lowland gorillas are distributed along a gradient of PC2, with individuals from the Congo and western Cameroon positioning in opposite directions along the axis. The isolated Cross River gorilla is genetically more differentiated from all gorilla groups, for example, have similar distributions of tracts of homozygosity as human populations that have experienced strong genetic bottlenecks (Karitiana and Papuan). Eastern lowland gorillas appear to represent the most inbred population, with evidence that they have been subjected to both recent and ancient inbreeding.
For instance, migration has been identified from western into eastern plex demographic history has been previously reported for chimpanzees that Nigeria–Cameroon and western chimpanzees form a clade. A maximum-likelihood tree estimated from allele frequencies both show acted more efficiently in populations with higher \( N_e \) consistent with neutral theory (Supplementary Information).

Although the phylogeny of bonobos and western, central and eastern common chimpanzees has been well established based on genetic data\(^1\), there is still uncertainty regarding their relationship to Nigeria–Cameroon chimpanzees\(^2,19\). Regional neighbour-joining trees and a maximum-likelihood tree estimated from allele frequencies both show that Nigeria–Cameroon and western chimpanzees form a clade. A complex demographic history has been previously reported\(^6\) for western and central chimpanzee populations. c. Runs of homozygosity among great apes. The relationship between the coefficient of inbreeding (\( F_{\text{ROH}} \)) and the number of autozygous >1 megabase segments is shown. Bonobos and eastern lowland gorillas show an excess of inbreeding compared to the other great apes, suggesting small population sizes or a fragmented population. d. Genetic structure based on clustering of great apes. All individuals (columns) are grouped into different clusters (\( K = 2 \) to \( K = 6 \), rows) coloured by species and according to their common genetic structure. Most captive individuals, labelled on top, show a complex admixture from different wild populations. A signature of admixture, for example, is clearly observed in the known hybrid Donald, a second-generation captive predicted to be a 15% admixture of central chimpanzee on a western background consistent with its pedigree. A grey line at the bottom denotes new groups at \( K = 6 \) in agreement with the location of origin or ancestral admixture.

Genetic diversity is depressed at or close to genes in almost all species (Supplementary Fig. 11.1) with the effect less pronounced in subspecies with lower estimated \( N_e \) consistent with population genetic theory. When we compare the relative level of X chromosome and autosomal (X/A) diversity across great apes as a function of genetic distance from genes, the eastern lowland gorillas and Bornean orangutans are outliers, with substantially reduced X/A diversity compared to the neutral expectation of 0.75, regardless of the distance to genes. This pattern is consistent with a recent reduction in effective population size\(^20\), clearly visible in the PSMC analysis for both species (Fig. 3). However, bonobos also demonstrate a relatively constant level of X/A diversity regardless of distance from genes, with values very much in line with neutral expectations. All other subspecies demonstrate a pattern consistent with previous studies in humans\(^21\) where X/A diversity is lower than 0.75 close to genes and higher farther away from genes.

It has been proposed that loss of gene function may represent a common evolutionary mechanism to facilitate adaptation to changes in an environment\(^22\). There has been speculation that the success of humans may have, in part, been catalysed by an excess of beneficial

---

**Figure 1** | Samples, heterozygosity and genetic diversity. a. Geographical distribution of great ape populations across Indonesia and Africa sequenced in this study. The formation of the islands of Borneo and Sumatra resulted in the speciation of the two corresponding orangutan populations. The Sanaga River forms a natural boundary between Nigeria–Cameroon and central chimpanzee populations whereas the Congo River separates the bonobo population from the central and eastern chimpanzees. Eastern lowland and western lowland gorillas are both separated by a large geographical distance. b. Heterozygosity estimates of each of the individual species and subspecies are superimposed onto a neighbour-joining tree from genome-wide genetic distance estimates (branch lengths in units of substitutions per bp). Arrows indicate heterozygositess previously reported\(^6\) for western and central chimpanzee populations. c. Runs of homozygosity among great apes. The relationship between the coefficient of inbreeding (\( F_{\text{ROH}} \)) and the number of autozygous >1 megabase segments is shown. Bonobos and eastern lowland gorillas show an excess of inbreeding compared to the other great apes, suggesting small population sizes or a fragmented population. d. Genetic structure based on clustering of great apes. All individuals (columns) are grouped into different clusters (\( K = 2 \) to \( K = 6 \), rows) coloured by species and according to their common genetic structure. Most captive individuals, labelled on top, show a complex admixture from different wild populations. A signature of admixture, for example, is clearly observed in the known hybrid Donald, a second-generation captive predicted to be a 15% admixture of central chimpanzee on a western background consistent with its pedigree. A grey line at the bottom denotes new groups at \( K = 6 \) in agreement with the location of origin or ancestral admixture.
Our analysis provides one of the first genome-wide views of the major patterns of evolutionary diversification among great apes. We have generated the most comprehensive catalogue of SNPs for chimpanzees (27.2 million), bonobos (9.0 million), gorillas (19.2 million) and orangutans (24.3 million) (Table 1) to date and identified several thousand AIMs, which provides a useful resource for future analyses of ape populations. Humans, western chimpanzees and eastern gorillas all show a remarkable dearth of genetic diversity when compared to other great apes. It is striking, for example, that sequencing of 79 great ape genomes identifies more than double the number of SNPs obtained from the recent sequencing of more than a thousand diverse humans—a reflection of the unique out-of-Africa origin and nested phylogeny of our species.

We provide strong genetic support for distinct populations and subpopulations of great apes with evidence of additional substructure. The common chimpanzee shows the greatest population stratification when compared to all other lineages with multiple lines of evidence supporting two major groups: the western and Nigeria–Cameroon and the central and eastern chimpanzees. The PSMC analysis indicates a temporal order to changes in ancestral effective population sizes over the last two million years, previous to which the Pan genus suffered a dramatic population collapse. Eastern chimpanzee populations reached their maximum size first, followed by the central and western chimpanzee. The Nigeria–Cameroon chimpanzee population size appears much more constant.

Despite their rich evolutionary history, great apes have experienced drastic declines in suitable habitat in recent years, along with declines in local population sizes of up to 75% (ref. 27). These observations highlight the urgency to sample from wild ape populations to more fully understand reservoirs of genetic diversity across the range of each species and to illuminate how basic demographic processes have affected it. The >80 million SNPs we identified in this study may now be used to characterize patterns of genetic differentiation among great apes in sanctuaries and zoos and, thus, are of great importance for the conservation of these endangered species with regard to their original range. These efforts will greatly enhance conservation planning and management of apes by providing important information on how to maintain genetic diversity in wild populations for future generations.
METHODS SUMMARY

We sequenced to a mean coverage of 25× (Illumina HiSeq 2000) a total of 79 great ape individuals, representing 10 subspecies and four genera of great apes from a variety of populations across the African continent and Southeast Asia. SNPs were called using GATK\(^1\) after BWA\(^2\) mapping to the human genome (NCBI Build 36) using relaxed mapping parameters. Samples combined by species were realigned around putative indels. SNP calling was then performed on the combined individuals for each species. For indels, we used the GATK Unified Genotyper to produce an initial set of indel candidates applying several quality filters and removing variants overlapping segmental duplications and tandem repeats. We also removed groups of indels clustering within 10 bp to eliminate possible artefacts in problematic regions. Conservative allelic imbalance filters were used to eliminate false heterozygotes that may affect demographic analyses, some of which are sensitive to low levels of contamination. We estimate that the application of this filter resulted in a 14% false negative rate for heterozygotes. Our multispecies study design facilitated this assessment of contamination, which may remain undetected in studies focused on assessing diversity within a single species. The amount of cross-species contamination was estimated from the amount of non-endogenous mitochondrial sequence present in an individual. Because we wished to compare patterns of variation between and within species, we report all variants with respect to coordinates of the human genome reference. For FRAPPE analyses, we used MAF0.06 (human, orangutan and bonobo) and 0.05 (chimpanzee and gorilla) to compare patterns of variation between and within species, we report all variants.

We sequenced to a mean coverage of 25× (Illumina HiSeq 2000) a total of 79 great ape individuals, representing 10 subspecies and four genera of great apes from a variety of populations across the African continent and Southeast Asia. SNPs were called using GATK\(^1\) after BWA\(^2\) mapping to the human genome (NCBI Build 36) using relaxed mapping parameters. Samples combined by species were realigned around putative indels. SNP calling was then performed on the combined individuals for each species. For indels, we used the GATK Unified Genotyper to produce an initial set of indel candidates applying several quality filters and removing variants overlapping segmental duplications and tandem repeats. We also removed groups of indels clustering within 10 bp to eliminate possible artefacts in problematic regions. Conservative allelic imbalance filters were used to eliminate false heterozygotes that may affect demographic analyses, some of which are sensitive to low levels of contamination. We estimate that the application of this filter resulted in a 14% false negative rate for heterozygotes. Our multispecies study design facilitated this assessment of contamination, which may remain undetected in studies focused on assessing diversity within a single species. The amount of cross-species contamination was estimated from the amount of non-endogenous mitochondrial sequence present in an individual. Because we wished to compare patterns of variation between and within species, we report all variants with respect to coordinates of the human genome reference. For FRAPPE analyses, we used MAF0.06 (human, orangutan and bonobo) and 0.05 (chimpanzee and gorilla) to compare patterns of variation between and within species, we report all variants.