

RESEARCH

Open Access

The genetic impact of an Ebola outbreak on a wild gorilla population



Claudia Fontserè^{1*†}, Peter Frandsen^{2,3†}, Jessica Hernandez-Rodriguez^{1,4}, Jonas Niemann⁵, Camilla Hjorth Scharff-Olsen⁵, Dominique Vallet⁶, Pascaline Le Gouar⁶, Nelly Ménard⁶, Arcadi Navarro^{1,7,8,9}, Hans R. Siegismund³, Christina Hvilsom², M. Thomas P. Gilbert^{5,10}, Martin Kuhlwilms^{1,11*†}, David Hughes^{1,12,13†} and Tomas Marques-Bonet^{1,7,14,15*†}

Abstract

Background: Numerous Ebola virus outbreaks have occurred in Equatorial Africa over the past decades. Besides human fatalities, gorillas and chimpanzees have also succumbed to the fatal virus. The 2004 outbreak at the Odzala-Kokoua National Park (Republic of Congo) alone caused a severe decline in the resident western lowland gorilla (*Gorilla gorilla gorilla*) population, with a 95% mortality rate. Here, we explore the immediate genetic impact of the Ebola outbreak in the western lowland gorilla population.

Results: Associations with survivorship were evaluated by utilizing DNA obtained from fecal samples from 16 gorilla individuals declared missing after the outbreak (non-survivors) and 15 individuals observed before and after the epidemic (survivors). We used a target enrichment approach to capture the sequences of 123 genes previously associated with immunology and Ebola virus resistance and additionally analyzed the gut microbiome which could influence the survival after an infection. Our results indicate no changes in the population genetic diversity before and after the Ebola outbreak, and no significant differences in microbial community composition between survivors and non-survivors. However, and despite the low power for an association analysis, we do detect six nominally significant missense mutations in four genes that might be candidate variants associated with an increased chance of survival.

Conclusion: This study offers the first insight to the genetics of a wild great ape population before and after an Ebola outbreak using target capture experiments from fecal samples, and presents a list of candidate loci that may have facilitated their survival.

Keywords: Ebola, gorilla, non-invasive samples, candidate genes

Background

The Ebola virus (EBOV), discovered in 1976, causes a severe disease and often fatal hemorrhagic fever for which numerous human outbreaks have been reported throughout Africa [1]. The most virulent outbreak reported to date was in West Africa in December 2013 and lasted until 2016 with more than 28,000 confirmed or suspected human cases and more than 11,000 human deaths [2]. Since then, other outbreaks of Ebola have been observed. In June 2020, when the 2018 outbreak

* Correspondence: claudia.fontserè@upf.edu; martin.kuhlwilms@univie.ac.at; tomas.marques@upf.edu

†Claudia Fontserè and Peter Frandsen share co-first authorship.

†Martin Kuhlwilms, David Hughes and Tomas Marques-Bonet share co-last authorship.

¹Institut de Biologia Evolutiva, (CSIC-Universitat Pompeu Fabra), PRBB, Doctor Aiguader 88, 08003 Barcelona, Catalonia, Spain

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

that occurred in Ituri, North Kivu and South Kivu provinces of Democratic Republic of Congo was declared over by the World Health Organization (WHO), 3470 cases had been reported with 2287 deaths (fatality rate of 66%) [3].

EBOV belongs to the single-stranded RNA virus family *Filoviridae* [4] with five distinct strains in the Ebola genus: *Zaire*, *Sudan*, *Bundibugyo*, *Tai Forest* and *Reston*. The first three are responsible for the majority of human infections [5, 6]. The virus is highly infectious and can enter the body through direct contact of broken skin or mucous membranes with infected blood or body fluids, causing symptoms including fever, vomiting, diarrhea, internal and external bleeding. Ebola hemorrhagic fever or Ebola virus disease (EVD) is an acute and severe disease with a fatality rate in humans around 50% [5–7].

Infectious diseases such as Ebola are considered to be a threat to the survival of African great apes [8], together with other threats such as habitat loss, climate change and poaching [9]. In some cases, the human outbreaks have been linked to contact with infected bushmeat from gorillas or chimpanzees [10] and several surveys have reported dramatic declines in populations of great apes in parallel with human EVD outbreaks with laboratory confirmation of Ebola virus infection in some carcasses [10–12]. Gorilla populations from the Republic of Congo suffered severe die-offs during a human EVD outbreak near the Lossi sanctuary in 2002–2003 [12] and Odzala-Kokoua National Park in 2004 [13] with reported mortality rates as high as 95%. In Lossi sanctuary alone, it was estimated that the Ebola virus killed 5000 wild gorillas [12]. The severe population decline, as a result of illegal hunting, disease and habitat loss, has contributed to the 2007 shift of the conservation status of western gorillas from “endangered” to “critically endangered” by the International Union for Conservation of Nature (IUCN) [14]. Furthermore, the recent outbreak of human EVD in North Kivu (Democratic Republic Congo) [3] was in close proximity to the remnant populations of eastern gorilla species, hence a human to ape transmission of the virus could potentially mark the end of existence for this critically endangered species.

Threats such as infectious diseases are relevant for conservation efforts, and efficient strategies are needed to reduce the effects of EVD on wild great ape populations [15]. Understanding any genetic impact that EBOV outbreaks might have on wild populations is vital, as EVD contributes to the fragmentation of gorilla populations due to a heterogeneous spatial influence of the outbreak [16]. Social dynamics in gorillas are rapidly affected by Ebola through a decrease in social cohesion, although recovery after the outbreak has been observed [17, 18]. One study reported that solitary individuals were less affected than individuals living in groups,

marking the relevance of social dynamics for transmission [13]. A previous study using 17 microsatellites found no loss of genetic diversity after one EBOV outbreak in Lossi sanctuary and Odzala-Kokoua National Park, which could be explained by post-epidemic immigration, sufficiently large remnant effective population size or a short period of time after the decline [18]. The present study represents a continuation of this aforementioned research since many aspects of EBOV infection in wild gorilla populations are not yet explored, such as genotypes of surviving individuals that might contribute to resistance or higher chance of survival to EVD.

Furthermore, microbial organisms inhabiting the gut also play a potentially crucial role in training and maintaining the immune system [19–21]. While some commensal microbes are associated with priming the immune response or activating antiviral responses, others facilitate the development of the infection or suppress the immune response [22]. For instance, a recent study reported a link between the gastrointestinal microbiome of healthy humans and a predisposition to severe COVID-19 [23], with an abundance of *Klebsiella*, *Streptococcus*, and *Ruminococcus* being correlated with elevated levels of proinflammatory cytokine. The association between infectious diseases and the gut microbiome of nonhuman primates is less well understood. While the gut microbiome of SIVgor-infected (gorilla Simian Immunodeficiency virus) wild gorillas seems to be more robust to dysbiosis than those of chimpanzees and humans [24], it is unclear how shifts in the gorilla gut microbiome can impact the severity of viral infections, and in particular the immune response to Ebola virus infection.

Long-term monitoring of a western lowland gorilla (*Gorilla gorilla gorilla*) population in Odzala-Kokoua National Park (Republic of Congo) has been ongoing since 2001 until 2017 encompassing the Ebola outbreak of 2004 [16–18, 25] which resulted in a mortality rate of 95% [13]. Population monitoring involved the recording of individual histories of hundreds of identified individuals, the determination of sex, age and social status and the collection of fecal samples in different time periods. During the Ebola outbreak, which lasted six months in the monitored population, about 340 individuals disappeared and entire groups were missing. Knowing the occurrence of Ebola epidemics in humans, the occurrences of gorilla carcasses infected by Ebola [10], and the detection of Ebola virus antibodies in great ape fecal samples in the areas of gorilla population decline [26], the scientific community has attributed the abnormal rate of gorilla disappearances to massive death caused by EVD. Individuals without any sign of injury, handicap or senescence that went missing after the Ebola period were

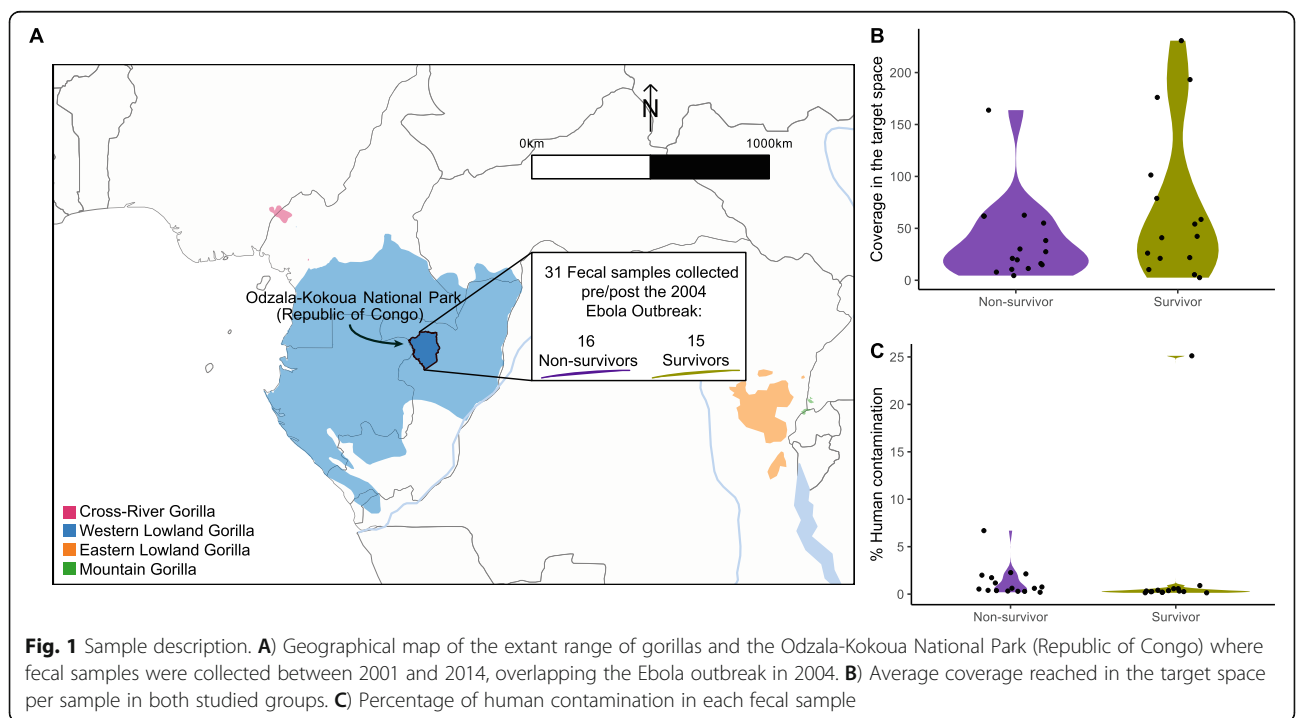
considered as dead because of EVD, and individuals that were seen before and after the outbreak were considered as surviving the epidemic period.

Here, focusing on a well described western lowland gorilla population [18, 27], we have obtained genetic data from gorilla fecal samples pre- and post-outbreak to explore potential associations to survivorship. Particularly, we compare the genetic variation at the single nucleotide level in 123 autosomal target-captured genes with putative roles in virus immune response and the gut microbiome composition in a reduced panel of survivors and non-survivors of this Zaire EBOV outbreak. With that we show that targeted capture on non-invasive fecal samples and next-generation sequencing can be used to study the impact of this severe disease in a natural population.

Results

A total of 31 non-invasive fecal samples from identified western lowland gorillas were collected between 2001 and 2014 in Odzala-Kokoua National Park, Congo [18] (Fig. 1A). Sixteen of these were from individuals declared missing after the Zaire Ebola virus outbreak in 2004. These individuals did not show any sign of injury, handicap or senescence that could lead to natural death. Also, they were observed shortly before the epidemic period and were thus unlikely to disperse. Therefore, they were suspected to have died because of the infection, here termed ‘non-survivors’. These samples were collected between 2001 and 2004. The remaining 15

samples were from gorillas identified before the epidemic and still observed after the epidemic, and are identified henceforth as ‘survivors’ [17]. Although no immunological test could be performed on the dried fecal samples, the survivors were known to live in the area where the epidemic emerged and thus they were suspected to have been in contact with the virus. These 15 samples were collected between 2005 and 2014 (Supplementary Table S1). We used target capture enrichment to sequence the genomic regions of 123 genes, which had previous evidence of putative roles in immune response to EBOV or other viruses (Supplementary Table S2). In addition, 15 neutral regions previously studied in other human and non-human primate studies were also targeted [28, 29] (Supplementary Table S2). Target design and all analyses were performed using the human reference genome due to the higher quality of annotations compared to the gorilla reference genome. We sequenced an average of 73 million paired reads per individual, 12% of which were unique (Supplementary Table S3). On average, 0.38% of the data mapped to the target space, representing an on-target effective coverage of 53.89-fold (range: 2.52-fold to 230.70-fold; Fig. 1B, Supplementary Table S3), with 72% of the target space covered by at least 4 reads per individual. Samples G282, G1392 and G638 performed poorly, with <50% of the target space covered at a minimum depth of 4 reads (Supplementary Fig. S1). Overall performance can be assessed by calculating how well the capture resequencing experiment went relative to expectations had we



performed random shotgun sequencing. In that regard, we observe an average enrichment of 125-fold (88–346-fold) (Supplementary Table S3). Individuals with extremely low proportions of target space covered by at least 4 reads (<30%) (G638 and G282) and high heterozygosity and high levels of human contamination were removed from further analysis (individuals G374 and G1392, Fig. 1C and Supplementary Fig. S2, Table S3).

With the final dataset of 13 survivors and 14 non-survivors, we validated that the genotype information obtained was in concordance with previously published gorilla whole-genomes, as determined by a principal component analysis (PCA) (Supplementary Fig. S3). Since the sampling of the gorilla population was done to avoid related individuals by sampling from different groups (a total of 45 groups and 31 solitary males were present before the outbreak), we identified a single case of close relatedness (1st degree) among them (non-survivor G374 and survivor G739; kinship coefficient = 0.4) (Supplementary Fig. S4). In addition, we observed no stratification correlating with survivor/non-survivor classification. Individuals appear to be dispersed randomly across a dendrogram derived from shared genotype likelihood dosage states (Fig. 2A), and a univariate linear regression of survivorship on the top 5 PCs identified no significant structure associated to survivorship (P -value >0.1; Supplementary Fig. S5). Hence, we determined no genome-wide group structure differences between survivors and non-survivors. The overall level of genetic diversity within the target space of the studied gorillas was, on average, lower than that of western lowland gorillas obtained from whole-genome sequencing (Supplementary Fig. S6) [30, 31], an expected outcome following the target capture procedure. Moreover, there are no statistically significant differences in heterozygosity between survivors and non-survivors (Student's t -test, p -value = 0.34; Fig. 2B).

In order to determine genetic differences between the groups, we calculated three summary statistics on a dataset of 6852 high-quality variants: (1) the difference in allele frequency (Δ Frequency), (2) the fixation index (F_{ST}), and (3) the significance level (α) of each variant for its association with the binary trait survivor/non-survivor (Fig. 3). We found 118 SNPs within the target space that surpassed our α threshold in the association test, and we also reported their Δ Frequency and F_{ST} values. However, after controlling for type I error none of these remained significant. Out of these, seven genes have multiple nominally significant SNPs (*CD1B*, *IGKV4-1*, *HLA-A*, *ACTB*, *LYN*, *CD68* and *MX1*), while 10 neutral regions (~10 kb each) have at least 1 nominally significant variant (Supplementary Table S4). For comparative validation, we repeated the association analysis using ANGSD [32], a software explicitly built to

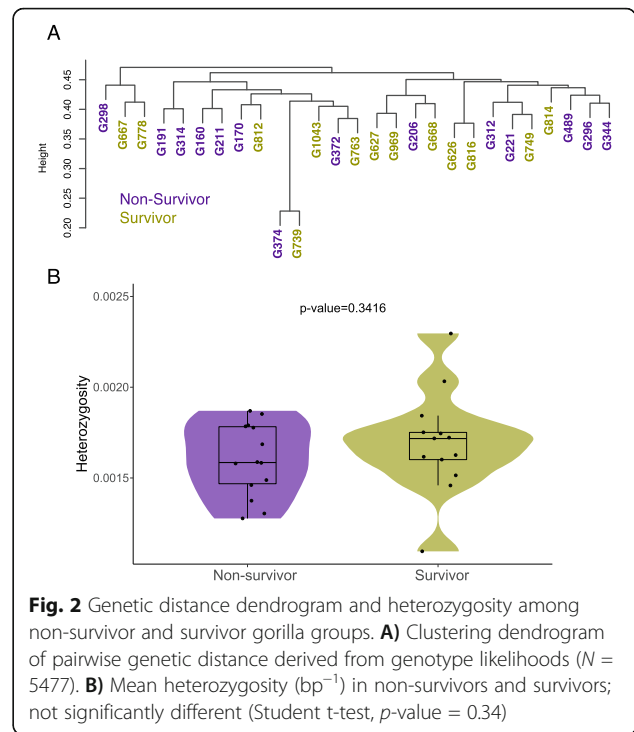
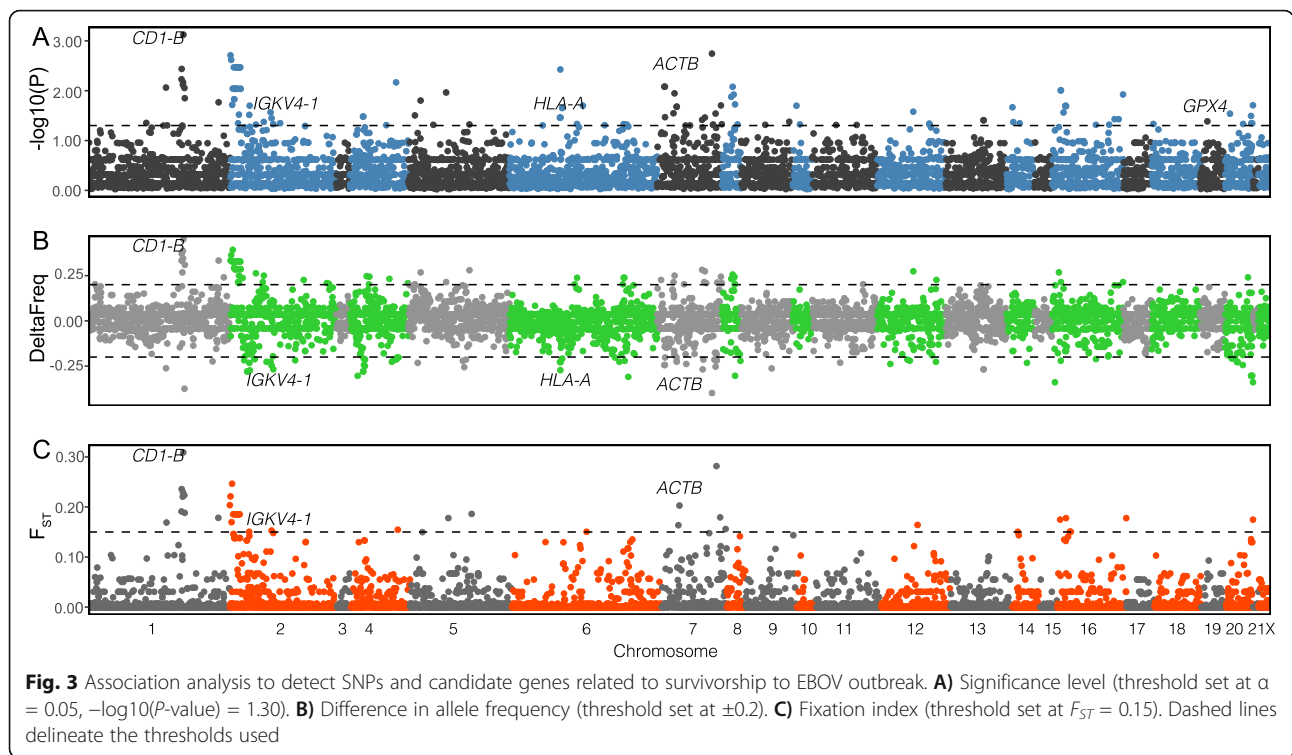


Fig. 2 Genetic distance dendrogram and heterozygosity among non-survivor and survivor gorilla groups. **A**) Clustering dendrogram of pairwise genetic distance derived from genotype likelihoods ($N = 5477$). **B**) Mean heterozygosity (bp^{-1}) in non-survivors and survivors; not significantly different (Student t -test, p -value = 0.34)

work with low coverage data that relies on genotype likelihoods. With this method, we were able to recover the majority of the genes found above (30 out of 36). However, ANGSD returned more hits and thus more genes (Supplementary Figs. S7, S8 and Table S5), rendering the above approach more conservative.

Next, we explored the potential functional impact of the variants in the nominally significant candidate loci that differentiate survivors from non-survivors. While we found no significant associations with gene ontology categories, the analysis of predictions of functional consequences pinpointed six missense mutations which differed in frequency between the two groups (Supplementary Table S6), one each in the *ATM*, *IGKV4-1* and *RNF167* genes (all lower in survivors) and three in the *ACTB* (Actin Beta) gene (one unique to survivors, two lower in survivors, Supplementary Table S4). All three missense variants in *ACTB* are predicted as deleterious by both the PolyPhen [33] and SIFT [34] algorithms. Furthermore, the derived variant in survivors in the immunoglobulin-encoding *IGKV4-1* might be deleterious (C-score > 20 [35]), hence potentially functionally relevant. Since we used the human genome for target design, mapping and variant calling, we caution that differences in exon usage or pseudogenization on the gorilla lineage might confound these inferences of protein-coding changes. In order to confirm the expression of these genes, and specifically the exons of interest, we mapped transcriptome data from six tissues in gorillas [36] to the same reference genome, and quantified expression levels. We confirmed the expression



of these genes and found high transcript abundance (\log_2 -value of counts > 9) for *ACTB*, *RNF167* and *ATM*, while *IGKV4-1* was only detected at low levels in these tissues (\log_2 -value of counts = 5.84). We found ~10,000 RNA sequencing reads overlapping the three loci of interest in *ACTB*, 437 in *RNF167*, and 96 in *IGKV4-1*, supporting the expression of these specific loci in gorilla tissues, while for *ATM* only 10 reads overlapped (Supplementary Table S6).

Among the 118 significant SNPs, we found no direct overlap with loci associated to 2385 traits in genome-wide association studies (GWAS) for humans [37]. However, we found 49 associated loci within close proximity (5000 bp) to GWAS loci (Supplementary Table S7), among which the locus 19:1104078 near *GPX4* stands out for its association with five blood cell traits (counts and percentages of leukocyte cell types). Furthermore, 6:29916885 in the *HLA* is associated to hemoglobin levels. Five loci in *HLA-DRB1* are associated with different autoimmune diseases like systemic lupus erythematosus or multiple sclerosis (Supplementary Table S7). Furthermore, nine loci in this region (*HLA-DRB5* and *HLA-A*) are associated with schizophrenia or autism spectrum disorder (Supplementary Table S7).

It has been reported that intestinal microbiota play a critical role in immune response to infectious diseases [19–21]. Thus, the microbiome might be relevant for the survival of wild gorilla populations experiencing an Ebola outbreak. Taking advantage of

the nature of the samples, we first analyzed the microbiota present in each fecal extraction using a 16S rRNA library (Methods). We obtained a total of 96,928,400 reads with an average sequencing depth of 1,101,459 (SD \pm 418,989) reads per gorilla fecal extraction (Supplementary Table S8), and determined the abundance of taxa (Supplementary Figs. S9 and S10). Firmicutes (53.79%), Bacteroidetes (12.02%) and Chloroflexi (11.11%) were the predominant phyla (Supplementary Fig. S9 and Table S9), that include the following most abundant orders: Clostridiales (39.37%), Bacillales (11.24%), Bacteroidales (10.87%) and Anaerolineales (11.11%) (Supplementary Fig. S10 and Table S10), concordant with previous findings [38, 39]. We found no taxa significantly differing in relative abundance between survivor and non-survivor gorillas (Bonferroni-corrected p -values > 0.05 ; Supplementary Fig. S11 and Table S11), and sample groups were not separated in a clustering analysis (Fig. S12).

Since these results on the gut microbiome diversity did not support differences between both gorilla groups, we decided to perform deep sequencing on the fecal libraries. We generated a total of 801,132,281 sequences from DNA libraries (4,025,054–25,593,317 reads per sample; Supplementary Table S8) and used MALT (MEGAN Alignment Tool) [40] for an alternative characterization of the microbial profile of the samples (Supplementary Fig. S13). We find that the majority of the identified taxa are associated with the gut

microbiome (Supplementary Table S12 and S13). By far the most abundant taxon is the gut bacterium *Escherichia coli*, which could be detected in all samples, and makes up more than 50% of all assigned sequences in nine of the samples. Also in high abundance are species of the Bacteroidales order, such as *Bacteroides cellulosilyticus* and *Prevotella spp.*, as well as members of the Clostridiales, Lactobacillales, and Bacillales orders, corroborating previous reports on the composition of western lowland gorilla gut microbiomes [38, 41]. Furthermore, we detected pathogenic taxa in high abundance in some of the samples, such as *Clostridium botulinum*, *Acinetobacter baumannii*, and *Klebsiella pneumoniae*, which have been previously found in the gorilla gut [42]. However, the microbial profiles of survivors and non-survivors do not differ significantly from each other in this analysis either (Bonferroni-corrected *p*-values >0.05, two-sided t-test, Supplementary Table S11), and the two groups do not form separate clusters in a Principal Coordinate analysis or a Neighbor Joining Tree (Supplementary Figs. S14 and S15).

Discussion

We investigated non-invasive fecal samples from a long-term monitored population of western lowland gorillas in the Republic of Congo, including individuals that most likely succumbed to the *Zaire* Ebola virus outbreak in 2004, as well as surviving individuals [17, 18]. We used targeted capture of 123 autosomal genes with putative roles in immune response to EBOV or other viruses (Supplementary Table S2) from fecal samples. This yielded an enrichment of more than 100-fold across samples, and a medium to high coverage of the target space across most individuals (Supplementary Fig. S1 and Table S3). Although a large proportion of reads were duplicates, the overall performance was high and these results demonstrate the great potential of capture experiments for obtaining genotypes from fecal samples of wild great ape populations [43, 44], for which high-coverage sequencing would be prohibitively expensive. We determined that the studied individuals were not closely related, hence most likely representing a random sampling of the wild gorilla population before and after the outbreak. We also investigated the microbial community composition of survivors and non-survivors, finding no significant differences in taxa abundance, neither using 16S rRNA or deep sequencing data (Supplementary Figs. S11, S13 and Table S11). Hence, we find no evidence that the gut microbiome of individuals has an influence on the survival rate of wild gorillas exposed to Ebola. However, these observations are limited by (1) the sample size, and (2) the broad range of collection dates (Supplementary Table 1).

The latter is particularly true (2a) relative to the timing of any exposure, but also in respect to the (2b) dynamic nature of the gut microbiome [39].

Given the limited sample size, we developed an approach using differences in allele frequency, the fixation index and the effect size to determine variants most strongly associated to survivability in the studied population, generally replicable using an association analysis with ANGSD. While 44 of the 118 nominally significant SNPs (Supplementary Table S4) do fall within 10 of the 15 neutral regions included in the study, some SNPs might be functionally relevant for surviving the EBOV outbreak. The non-synonymous variants in *ACTB*, *RNF167* and *IGKV4-1* genes are obvious candidate loci, and particularly the three deleterious missense mutations in Actin Beta appear to be strong candidates for a higher survival rate. The actin cytoskeleton is important for virus assembly [45], and a disturbed assembly process could have influenced the viral load in individuals with changes in this protein. As expected for a gene encoding a structural protein, *ACTB* is highly expressed in gorilla tissues. Furthermore, the variant in *IGKV4-1* might improve the immune response to viral infection through antigen recognition [46]. The missense mutation in ATM, which belongs to the PI3-kinase family, could interfere with the cellular entry specifically of the Ebola virus [47], although we could not confirm expression of this locus *in vivo* in the available tissues.

We find other potentially relevant non-coding variants, 49 of which are in close proximity to SNPs associated to GWAS traits in humans, suggesting possible regulatory functions (Supplementary Table S7). Among those, the association of *GPX4* with leukocyte cell type count might reflect differences in leukocyte composition after viral infection. Differences in hematocrit or hemoglobin levels might have contributed to the survival of wild gorillas considering that hemorrhage and internal bleeding are symptoms of Ebola infection. Since eight loci are associated to the *HLA-DRB* gene, a direct involvement in the adaptive immune system might cause the signature observed at this locus, particularly given that human survivors of EVD show a lower frequency of *HLA-DR*-positive T cells [48].

Conclusion

By using fecal samples and targeted capture enrichment, non-invasive assessment of numerous individuals from wild populations is possible. Here, we demonstrate that this approach can be used to analyze temporal genetic changes in wild great ape populations in response to environmental factors. Additionally, we present candidate loci that may have facilitated the survival of gorilla individuals or groups after an outbreak of the *Zaire* Ebola

virus. Understanding putative adaptive responses to this pathogen in wild populations can help to advance our knowledge on the natural dynamics of this severe disease. Such a strategy might be useful in a broader context, since these and other primates are susceptible to other infectious diseases such as Covid-19 [49].

Methods

Samples, DNA extraction, library preparation

Non-invasive fecal samples from western lowland gorillas were collected between 2001 and 2014 near the Lokoué clearing in Odzala-Kokoua National Park, Republic of Congo [18], following the standard practices in the field. The population and group structure of this western lowland gorilla population has been described previously [27]. Among collected fecal samples, we selected 31 samples from previously identified individuals each from unique social groups (Supplementary Table S1). Sixteen of those individuals were declared missing after the epidemic in 2004 [13]. Given the high mortality rate (95%) of Ebola virus in gorillas, they were presumed dead during the time span of the epidemic and were identified here as non-survivors. Fifteen individuals were observed before and after the epidemic, and were described here as survivors (Fig. 1A and Supplementary Table S1) [17]. Fresh samples were collected by field investigators wearing masks and gloves, and were dried with silica beads and then stored at room temperature until arrival in a laboratory where they were stored at 4 °C until extraction. Sampling dates are recorded in Supplementary Table S1.

DNA was extracted from 10 mg of dried sample using the 2CTAB/PCI protocol [50] using negative controls that were checked for DNA contamination before subsequent experiment. Three different extractions were carried out, except for samples G778, G374, G344, G498 and G372, where only two extractions were performed (Supplementary Table S1). A DNA library [51–53] and a 16S rRNA library [54–56] were prepared for each extract. Isolated DNA samples were quantified with Qubit with a mean estimated concentration of 13.3 ng/μl (range: 0.90–74.7). Whenever possible, a total of 250 ng of DNA was used to construct DNA libraries, but never more than a total volume of 33 μl was taken from any single sample. DNA was sheared with a Covaris S2 instrument and 88 fecal DNA (fDNA) libraries were prepared following a custom dual-indexing protocol with 25 cycles of amplification [51, 52]. Subsequent to DNA library preparation, 88 16S rRNA libraries were prepared using 1 μl of total DNA. The V3 and V4 regions of 16S rRNA were target amplified using modified 341F and 806Rb primers [54–56], incorporated into the dual-indexing protocol [52]. The forward primer (IS1_P5_16S_341f: ACACTCTTCCCTACACGACGCTCTCCGATCTNNNNCTACGGGNGGCWGCAG),

and reverse primer (IS2_P7_16S_806rB: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGACTACNVGGTWTCTAAT) include the complementary sequences necessary for the final indexing step [52]. Protocols are provided in [53].

Target Design, Capture and Sequencing

RNA baits covering the target space were designed and synthesized by Agilent with a minimum of 3x bait coverage. The target space included specific autosomal genes (123 genes) and 15 neutral regions (~10 kb each) (Supplementary Table S2). For target enrichment, the fDNA libraries were pooled into one equimolar batch and subjected to two consecutive rounds of DNA capture with the RNA baits in 8 hybridizations. Captured fDNA libraries were sequenced on the Illumina system in four HiSeq 2500 2x125 lanes and one HiSeq 2500 rapid run at 2x250 bp. The 16S libraries were sequenced on one HiSeq 2500 rapid run 2x250 bp lane. In addition, we generated paired-end sequences from the fDNA libraries on four HiSeq 4000 lanes (2x150bp) to study the whole microbiome composition (Table S8).

Mapping and Variant Discovery

Prior to mapping, paired-end reads belonging to the same library but sequenced in different lanes were merged into a single FASTQ file. PCR duplicates were directly removed from FASTQ files using FASTuniq (v1.1) [57]. Overlapping reads were merged (minimum overlap of 10 bp, minimum length of final read to 50 bp) using PEAR (v0.9.6) [58]. Reads were mapped using BWA mem (v0.7.12) [59] to the human reference genome Hg19 (GRCh37 from the UCSC database). Assembled reads were mapped considering single-end specifications and unassembled reads considering paired-end specifications. Any remaining PCR duplicates were removed using PicardTools MarkDuplicates (v1.95) (<http://broadinstitute.github.io/picard/>). Non-primary alignments and reads with quality below 30 were filtered from the dataset with samtools (v1.5) [60]. Finally, single-end and paired-end reads were merged into a single BAM file using PicardTools MergeSamFiles (<http://broadinstitute.github.io/picard/>). The percentage of aligned reads for each DNA extraction and sample was calculated by dividing the number of uniquely and high-quality mapped reads (without duplicates) by the total number of sequenced reads. The percentage of on-target aligned reads was calculated for each sample by dividing the number of on-target filtered reads by the number of sequenced reads. The average target effective coverage was calculated dividing the number of aligned bases by the total length of the targeted genomic space. Finally, the enrichment factor (ER) of the capture performance was calculated using the ratio between the on-target reads by the total mapped reads over

the targeted size by genomic size ($ER = (\text{On-Target Reads}/\text{Mapped Reads})/(\text{Target Size}/\text{Genome Size})$). The coverage for each target region was retrieved using SAMtools bedcov [60].

For variant calling, all BAM files belonging to the same sample were merged into a single BAM file using Picard-Tools MergeSamFiles (v1.95) (<http://broadinstitute.github.io/picard/>). Variant discovery was performed using GATK 'Unified Genotyper' [61] for each sample independently with the following parameters `-out_mode EMIT_ALL_SITES -stand_call_conf 5.0 -stand_emit_conf 5.0 -A BaseCounts -A GCCContent -A RMSMappingQuality -A BaseQualityRankSumTest`. Afterwards, we merged each sample gvcf to a single one using GATK 'CombineVariants' [61] with the following parameters `-genotypeMergeOptions UNIQUIFY -exclude-NonVariant`. We also included in the gvcf the genotype information of available whole genome data of six *Gorilla beringei beringei*, eight *Gorilla beringei graueri*, one *Gorilla gorilla dielhi*, and twenty-three *Gorilla gorilla gorilla* samples [30, 31]. The VCF was filtered with VCFtools [62] to keep only biallelic positions with $DP > 3$ and quality > 30 and without indels.

Genotype likelihoods were directly obtained from BAM files with ANGSD [32] including four *Gorilla beringei beringei*, four *Gorilla beringei graueri*, one *Gorilla gorilla dielhi*, and four *Gorilla gorilla gorilla*, with the following parameters and only in the target space: `-uniqueOnly 1 -remove_bads 1 -only_proper_pairs 1 -trim 0 -C 50 -baq 1 -minInd 21 -skipTriallelic 1 -GL 2 -minMapQ 30 -doGlf 2 -doMajorMinor 1 -doMaf 2 -minMaf 0.05 -SNP_pval 1e-6`.

Quality control

We evaluated the amount of human contamination in each fecal library using the HuConTest script [63], as described previously [44]. The majority of samples have less than 2% of human contamination, but samples G348 and G1392 have estimates of human contamination of 6.7 and 25.1%, respectively (Supplementary Table S3). These two samples also show extreme values of heterozygosity (deviating > 1 s.d. from mean heterozygosity; Fig. S2). For the identification of individuals and markers with elevated missing data rates we used the proportion of the target space covered by at least 4 reads. Individuals with less than 30% of covered target space (4 reads) were not used for further analysis (G638 and G282).

A principal component analysis (PCA) was performed to validate that the genotype information obtained for the case study gorillas was in concordance with previously published data. We used PCAnsd [64] with the genotype likelihoods obtained with ANGSD ($N = 6484$), including 13 previously published whole-genomes

representative of each known gorilla subspecies [30, 31]. We also obtained a PCA using the GATK genotype calls after keeping only variants with minor allele frequency of 0.02 with plink `--pca` option ($N = 6051$) [65].

Genetic distance, relatedness and heterozygosity

We used the genotype likelihood information for the studied individuals to obtain the genetic distance by running ngsDist in ANGSD [66] with the following parameters: `--n_sites 5477 --probs TRUE --pairwise_del`. Then, we constructed an Euclidean distance matrix based on the genotypes and performed a hierarchical clustering using the R package *ape* [67]. We also run PCAnsd [64] considering only the study gorillas to discard any possible intra-group structure.

The theta coefficients of kinship (probability of a pair of randomly sampled homologous alleles are identical by descent) were calculated using the NgsrelateV2 [68, 69] on the genotype likelihood obtained with ANGSD [32]. Note that all possible genotype likelihoods, even outside the target space ($N = 226,094$), were used since the coverage of the kinship and neutral markers was insufficient.

To assess global levels of heterozygosity, the unfolded SFS (Site Frequency Spectrum) was calculated for each sample separately, including thirteen gorilla whole-genomes representative of all gorilla subspecies [30, 31], using ANGSD [32] and realSFS [70] only in the target space with the following quality filter parameters: `-uniqueOnly 1 -remove_bads 1 -only_proper_pairs 1 -trim 0 -C 50 -baq 1 -minMapQ 20 -minQ 20 -setMaxDepth 200 -doCounts 1 -GL 1 -doSaf 1`. We used the human genome (Hg19) to determine the ancestral state.

Association analysis

The genotype calls obtained with GATK were further filtered with Plink [65] to exclude variants considering their missing rate (`--geno 0.05`), minor allele frequency (`--maf 0.01`) and Hardy-Weinberg equilibrium (`--hwe 0.00001`). The final dataset consists of 27 samples (13 survivors and 14 non-survivors) and 6852 high-quality variants. Nominal significance was set at an alpha of 0.05 and a Bonferroni corrected alpha threshold of 7.3×10^{-6} ($0.05/6852$) was defined to account for family-wise error. Associations were tested for by a chi-square allelic test with one degree of freedom and p -values were estimated by permutation in plink (`plink --assoc -mperm 10,000`) [65]. The p -values were plotted in a Manhattan plot in R (v3.4.1).

The allele frequency for each group (survivors and non-survivors) was obtained using the `-freq2` option in VCFtools [62]. Then, we calculated the allele frequency difference per SNP by subtracting the allele frequency in non-survivors from the allele frequency in survivors

(Δ Frequency). We chose a threshold of ± 0.2 , and plotted the allele frequencies using R. We also reported the allele frequency differences per SNP considering all samples (pre-Ebola outbreak: survivors and non-survivors) and the survivors (post-Ebola outbreak), which results in more modest allele frequency differences. The fixation index (F_{ST}) between both groups was calculated using VCFtools `-weir-fst-pop` option (Weir and Cockerham) [62] with a threshold at 0.15, and results were plotted in R. We retrieved markers with $\alpha \leq 0.05$ in the association test and a p -value < 0.05 in the permutation test. The ANGSD software [32] was used to perform a replication of the association analysis using the following parameters `-minQ 20 -minMapQ 30 -doAsso 1 -GL 1 -out assocGQ_filter -doMajorMinor 1 -doMaf 1 -SNP_pval 1e-6 -minInd 22 -minMaf 0.02`. The output of the association analysis are LRT values (Likelihood Ratio Test), which are chi square distributed with one degree of freedom. Since we set a threshold of significance at 95% confidence, the minimum score to be significant is $LRT = 3.84$. In both association analyses, we linked the nominally significant SNPs with their genes (Supplementary Table S4 and S5). Genes with multiple nominally significant SNPs were considered to be potentially more relevant. Subsequently, we compared the overlap of discovered genes between the datasets (Unified Genotyper and ANGSD) in a Venn diagram (Supplementary Fig. S8) using the R package VennDiagram [71].

Prediction of functional consequences of the significant markers

We used VEP (v91) [72] for the functional annotation of the associated SNPs. We retrieved the predicted consequence of each significant marker found in the potentially related genes to Ebola immune response, as well as PolyPhen-2 [33], Sift [73] and C-scores [35]. Associated loci were intersected with hits in the GWAS catalogue [37] within 5000 bp. We also performed an overrepresentation test using Panther [74] to test whether any of the potentially related genes are overrepresented in biological or functional categories compared to the rest of targeted genes with no apparent association with Ebola. In addition, we mapped previously published RNA sequencing data from six tissues (brain, cerebellum, heart, kidney, liver, testis) in two gorilla individuals [36] to the annotated genes in the human reference genome (using the Ensembl Release 75 gene models) using Tophat2 [75], and estimated the gene expression with htseq-count [76]. Gene expression is reported \log_2 -normalized, and we counted the number of reads overlapping the candidate missense mutations to confirm their transcriptional activity in gorillas. Values presented are the cumulative sums of RNA sequencing reads across individuals and tissues.

Microbiome sequencing

16S RNA sequencing reads were processed using QIIME (v1) (Quantitative Insights Into Microbial Ecology) [77] to analyze the 16S rRNA. First, paired-end raw reads were merged using `fastq-join` from `ea-utils` package [78]. Then, with `usearch` software [79], merged FASTQ reads were filtered (`-fastq_truncLen 253` and `-fastq_maxE 0.5`). Using QIIME environment, the metadata mapping file was constructed and validated (`validate_mapping_file.py`) and QIIME labels were added (`add_qiime_labels.py`). We applied open-reference OTUs (Operational Taxonomic Units) picking (`pick_open_reference_otus.py`). Summary statistics were computed using `biom summarize-table`. The resulting dataset was rarefied to an even depth of 10,000 sequences per extract (6 extracts were excluded in diversity analysis: G191_5782, G344_5827, G314_5824, G489_5834, G489_5835, G344_5828). Finally, we ran diversity analysis with a sequence depth of 10,000 (`core_diversity_analysis.py`). Taxa abundance quantification and significance or relative taxa abundance (T-test and p -values adjusted for multiple testing with Bonferroni-correction) were computed in R.

Deep sequencing of the DNA library (pre-capture) was performed as stated above. To remove sequencing adapters and merge the read pairs, we used Adapter-Removal v2.2.4 with default settings [80]. We then aligned the merged sequences to the gorilla reference genome `Kamilah_GGO_v0` using `bwa mem` [59] to remove host DNA. Subsequently we filtered out potential human contaminant DNA by aligning the unmapped sequences to the human reference genome `hg19`, resulting in 724,738,878 filtered sequences. MALT v0.4.1 (MEGAN Alignment Tool) [40] was used to characterize the microbial profile, using all archaeal, viral, and bacterial reference sequences downloaded from NCBI on 06.05.2019. These were indexed using `malt-build` to build a custom database. `Malt-run` was then used with minimum percent identity (`--minPercentIdentity`) set to 95, the minimum support (`--minSupport`) parameter set to 10, and the top percent value (`--topPercent`) set as 1, other parameters were set to default. The resulting `rma6` files were visualized with MEGAN6 [81] and clustered in a Principal Coordinate analysis (PCoA) and Neighbor Joining Tree analysis according to microbial composition on the species level (Supplementary Figs. S14 and S15).

Abbreviations

EBOV: Ebola virus; EVD: Ebola virus Disease; GWAS: Genome-Wide Association Study; IUCN: International Union for Conservation of Nature; OTUs: Operational Taxonomic Units; PCA: Principal Component Analysis; SFS: Site Frequency Spectrum; SNP: Single Nucleotide Polymorphism

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-08025-y>.

Additional file 1.**Additional file 2.****Acknowledgments**

We are grateful to S. Gatti, F. Levréro, C. Genton, R. Cristescu for their assistance with sample collection on site. We thank the teams of the ECOFAC program (EU) and African Parks Networks for logistic assistance and permission to work in Odzala-Kokoua National Park. The storage and extraction of fecal samples were performed in the molecular ecology platform (UMR 6553 Ecobio, CNRS/UR1) dedicated to non-invasive samples.

Research involving animals

Non-invasive fecal samples were collected from the field following the standard practices. Fecal samples are exempt by the Convention on the Trade in Endangered Species of Wild Fauna and Flora (CITES).

Authors' contributions

D.H., T.M.-B., P.L.G., N.M., D.V., A.N. and M.T.P.G. conceived and designed the study; D.H., J.H.R. and P.F. selected the genes for the target capture design. D.H., D.V. and C.S.O. performed experiments; C.F., M.K., D.H., J.H.R., C.H.S.O. and J.N. analyzed data; P.L.G., N.M. and D.V. collected genetic and demographic data; P.F., H.R.S., C.H., provided analytical support. C.F., M.K., D.H., and T.M.-B. wrote the manuscript with input and approval from the other authors.

Funding

C.F. is supported by "la Caixa" PhD fellowship, fellowship code LCF/BQ/DE15/10360006. M.K. is supported by "la Caixa" Foundation (ID 100010434), fellowship code LCF/BQ/PR19/11700002. J. N is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 676154 (ArchSci2020) and an EMBO short-term fellowship STF-8036. P.F. is supported by the Innovation Fund Denmark. H.R.S. is supported by The Danish Council for Independent Research | Natural Sciences. A.N. is supported by BFU2015-68649-P (MINECO/FEDER, UE). M.T.P.G. is supported by the Danish Basic Research Foundation award DNRF143. T.M.-B. is supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 864203), BFU2017-86471-P (MINECO/FEDER, UE), "Unidad de Excelencia María de Maeztu", funded by the AEI (CEX2018-000792-M), Howard Hughes International Early Career and Secretaria d'Universitats i Recerca and CERCA Programme del Departament d'Economia i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 880). P.L.G., N.M. and D.V. are supported by the French National agency for research via the ANR-11-JV57-015 IDiPop project. D.H. is supported by Wellcome Investigator Award (202802/Z/16/Z) and works in the Medical Research Council Integrative Epidemiology Unit at the University of Bristol, which is supported by the Medical Research Council (MC_UU_000111/1-7). This long-term research on gorillas was funded by the ECOsystèmes FORestiers program (Ministère de l'Ecologie et du Développement Durable France), the Espèces-Phares program (DG Environnement, UE) and Lundbeck Foundation Visiting Professorship R317-2019-5 grant to T.M.-B. and M.T.P.G. This work was supported by: AEI-PGC2018-101927-BI00(FEDER/UE).

Availability of data and materials

The sequencing dataset generated in this study is publicly available at ENA (project ID: PRJEB43265).

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institut de Biologia Evolutiva, (CSIC-Universitat Pompeu Fabra), PRBB, Doctor Aiguader 88, 08003 Barcelona, Catalonia, Spain. ²Research and Conservation, Copenhagen Zoo, 2000 Frederiksberg, Denmark. ³Department of Biology, Section for Computational and RNA Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark. ⁴Genomics of Health Research Group, Hospital Universitari Son Espases (HUSE) and Institut d'Investigacions Sanitàries de Balears (IDISBA), Palma, Spain. ⁵Center for Evolutionary Hologenomics, The GLOBE Institute, University of Copenhagen, 1353 Copenhagen, Denmark. ⁶UMR 6553, ECOBIO: Ecosystems, Biodiversity, Evolution, CNRS/University of Rennes 1, Station Biologique de Paimpont, 35380 Paimpont, France. ⁷Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain. ⁸CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), 08036 Barcelona, Spain. ⁹BarcelonaBeta Brain Research Center, Pasqual Maragall Foundation, 08005 Barcelona, Spain. ¹⁰University Museum, NTNU, Trondheim, Norway. ¹¹Department of Evolutionary Anthropology, University of Vienna, Djerassiplatz 1, 1030 Vienna, Austria. ¹²Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. ¹³MRC Integrative Epidemiology Unit at University of Bristol, Bristol, UK. ¹⁴CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain. ¹⁵Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Edifici ICTA-ICP, c/ Columnes s/n, 08193 Cerdanyola del Vallès, Barcelona, Spain.

Received: 31 May 2021 Accepted: 13 September 2021

Published online: 11 October 2021

References

- Centers for Disease Control and Prevention. History of Ebola Virus Disease. U.S. Department of Health & Human Services; 2018. <https://www.cdc.gov/vhf/ebola/history/summaries.html>. Accessed 4 Feb 2020
- World Health Organization. Ebola virus disease situation report. 2016. <https://apps.who.int/iris/handle/10665/208883>. Accessed 3 Jul 2018.
- World Health Organization. External Situation Report 98 - Ebola Virus Disease. 2020. <https://www.who.int/publications/i/item/10665-332654>.
- Kuhn JH, Becker S, Ebihara H, Geisbert TW, Johnson KM, Kawaoka Y, et al. Proposal for a revised taxonomy of the family Filoviridae: Classification, names of taxa and viruses, and virus abbreviations. Arch Virol. 2010;155:2083-103. <https://doi.org/10.1007/s00705-010-0814-x>.
- Centers for Disease Control and Prevention. What is Ebola Virus Disease? | Ebola (Ebola Virus Disease) | CDC. U.S. Department of Health & Human Services; 2019. <https://www.cdc.gov/vhf/ebola/about.html>. Accessed 4 Feb 2020
- World Health Organization. Ebola virus disease. 2018. <http://www.who.int/en/news-room/fact-sheets/detail/ebola-virus-disease>. Accessed 3 Jul 2018.
- Feldmann H, Geisbert TW. Ebola haemorrhagic fever. Lancet. 2011;377:849-62. [https://doi.org/10.1016/S0140-6736\(10\)60667-8](https://doi.org/10.1016/S0140-6736(10)60667-8).
- Smith KF, Acevedo-Whitehouse K, Pedersen AB. The role of infectious diseases in biological conservation. Anim Conserv. 2009;12:1-12. <https://doi.org/10.1111/j.1469-1795.2008.00228.x>.
- Walsh PD, Abernethy KA, Bermejo M, Beyers R, De Wachter P, Akou ME, et al. Catastrophic ape decline in western equatorial Africa. Nature. 2003;422:611-4.
- Rouquet P, Froment JM, Bermejo M, Kilbourn A, Karesh W, Reed P, et al. Wild animal mortality monitoring and human ebola outbreaks, Gabon and Republic of Congo, 2001-2003. Emerg Infect Dis. 2005;11:283-90. <https://doi.org/10.3201/eid1102.040533>.
- Leroy EM, Rouquet P, Formenty P, Souquière S, Kilbourne A, Froment J-M, et al. Multiple Ebola Virus Transmission Events and Rapid Decline of Central African Wildlife. Science (80-). 2004;303:387-90. <https://doi.org/10.1126/science.1092528>.
- Bermejo M, Rodríguez-Teijeiro JD, Illera G, Barroso A, Vilà C, Walsh PD. Ebola outbreak killed 5000 gorillas. Science (80-). 2006;314:1564. <https://doi.org/10.1126/science.1133105>.
- Caillaud D, Levréro F, Cristescu R, Gatti S, Dewas M, Douadi M, et al. Gorilla susceptibility to Ebola virus: the cost of sociality. Curr Biol. 2006;16:489-91.
- Maisels F, Bergl R, Williamson E. Gorilla gorilla (errata version published in 2016). In: The IUCN Red List of Threatened Species; 2016. <http://dx.doi.org/10.2305/IUCN.UK.2016-2.RLTS.T9404A17963949.en>.

15. Leendertz SAJ, Wich SA, Ancenaz M, Bergl RA, Gonder MK, Humle T, et al. Ebola in great apes – current knowledge, possibilities for vaccination, and implications for conservation and human health. *Mammal Rev.* 2017;47:98–111.
16. Genton C, Cristescu R, Gatti S, Levréro F, Bigot E, Motsch P, et al. Using demographic characteristics of populations to detect spatial fragmentation following suspected ebola outbreaks in great apes. *Am J Phys Anthropol.* 2017;164:3–10. <https://doi.org/10.1002/ajpa.23275>.
17. Genton C, Cristescu R, Gatti S, Levréro F, Bigot E, Caillaud D, et al. Recovery potential of a western Lowland Gorilla population following a major Ebola outbreak: Results from a ten year study. *PLoS One.* 2012;7:e37106. <https://doi.org/10.1371/journal.pone.0037106>.
18. le Gouar PJ, Vallet D, David L, Bermejo M, Gatti S, Levréro F, et al. How Ebola impacts genetics of western lowland gorilla populations. *PLoS One.* 2009;4:e8375. <https://doi.org/10.1371/journal.pone.0008375>.
19. Honda K, Littman DR. The microbiome in infectious disease and inflammation. *Annu Rev Immunol.* 2012;30:759–95. <https://doi.org/10.1146/annurev-immunol-020711-074937>.
20. Fiebigler U, Bereswill S, Heimesaat MM. Dissecting the interplay between intestinal microbiota and host immunity in health and disease: Lessons learned from germfree and gnotobiotic animal models. *Eur J Microbiol Immunol.* 2016;6:253–71. <https://doi.org/10.1556/1886.2016.00036>.
21. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. *Cell Res.* 2020;30:492–506. <https://doi.org/10.1038/s41422-020-0332-7>.
22. Domínguez-Díaz C, García-Orozco A, Riera-Leal A, Padilla-Arellano JR, Fafutis-Morris M. Microbiota and its role on viral evasion: Is it with us or against us? *Front Cell Infect Microbiol.* 2019;9(JUL):256. <https://doi.org/10.3389/fcimb.2019.00256>.
23. Gou W, Fu Y, Yue L, Chen GD, Cai X, Shuai M, et al. Gut microbiota may underlie the predisposition of healthy individuals to COVID-19. *medRxiv.* 2020. <https://doi.org/10.1101/2020.04.22.20076091>.
24. Moeller AH, Peeters M, Ayoub A, Ngole EM, Esteban A, Hahn BH, et al. Stability of the gorilla microbiome despite simian immunodeficiency virus infection. *Mol Ecol.* 2015;24:690–7. <https://doi.org/10.1111/mec.13057>.
25. Genton C, Pierre A, Cristescu R, Lévréro F, Gatti S, Pierre JS, et al. How Ebola impacts social dynamics in gorillas: A multistate modelling approach. *J Anim Ecol.* 2015;84:166–76. <https://doi.org/10.1111/1365-2656.12268>.
26. Reed PE, Mulangu S, Cameron KN, Ondzie AU, Joly D, Bermejo M, et al. A New Approach for Monitoring Ebolavirus in Wild Great Apes. *PLoS Negl Trop Dis.* 2014;8(9):e3143. <https://doi.org/10.1371/journal.pntd.0003143>.
27. Gatti S, Levréro F, Ménard N, Gautier-Hion A. Population and group structure of western Lowland Gorillas (*Gorilla gorilla gorilla*) at Lokoué. Republic of Congo *Am J Primatol.* 2004;63:111–23. <https://doi.org/10.1002/ajp.20045>.
28. Fischer A, Prüfer K, Good JM, Halbwax M, Wiebe V, André C, et al. Bonobos fall within the genomic variation of Chimpanzees. *PLoS One.* 2011;6:1–10.
29. Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Di Rienzo A. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci U S A.* 2005;102:18508–13. <https://doi.org/10.1073/pnas.0507325102>.
30. Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, et al. Great ape genetic diversity and population history. *Nature.* 2013;499:471–5. <https://doi.org/10.1038/nature12228>.
31. Xue Y, Prado-Martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, et al. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science (80-).* 2015;348:242–5.
32. Korneliusen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics.* 2014;15:256. <https://doi.org/10.1186/s12859-014-0356-4>.
33. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7:248–9. <https://doi.org/10.1038/nmeth0410-248>.
34. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc.* 2016;11:1–9. <https://doi.org/10.1038/nprot.2015.123>.
35. Kircher M, Witten DM, Jain P, O’roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46:310–5. <https://doi.org/10.1038/ng.2892>.
36. Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, et al. The evolution of gene expression levels in mammalian organs. *Nature.* 2011;478. <https://doi.org/10.1038/nature10532>.
37. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 2017;45:D896–901.
38. Gomez A, Petrzalkova K, Yeoman CJ, Vlckova K, Mrázek J, Koppova I, et al. Gut microbiome composition and metabolomic profiles of wild western lowland gorillas (*Gorilla gorilla gorilla*) reflect host ecology. *Mol Ecol.* 2015;24:2551–65.
39. Hicks AL, Lee KJ, Couto-Rodriguez M, Patel J, Sinha R, Guo C, et al. Gut microbiomes of wild great apes fluctuate seasonally in response to diet. *Nat Commun.* 2018;9:1786. <https://doi.org/10.1038/s41467-018-04204-w>.
40. Vågøne ÅJ, Herbig A, Campana MG, Robles García NM, Warinner C, Sabin S, et al. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat Ecol Evol.* 2018;2:520–8.
41. Gomez A, Rothman JM, Petrzalkova K, Yeoman CJ, Vlckova K, Umaña JD, et al. Temporal variation selects for diet–microbe co-metabolic traits in the gut of *Gorilla* spp. *ISME J.* 2016;10:514–26.
42. Bittar F, Keita MB, Lagier JC, Peeters M, Delaporte E, Raoult D. Gorilla gorilla gut: A potential reservoir of pathogenic bacteria as revealed using culturomics and molecular tools. *Sci Rep.* 2014;4:1–5.
43. Hernandez-Rodriguez J, Arandjelovic M, Lester J, de Filippo C, Weihmann A, Meyer M, et al. The impact of endogenous content, replicates and pooling on genome capture from faecal samples. *Mol Ecol Resour.* 2018;18:319–33. <https://doi.org/10.1111/1755-0998.12728>.
44. Fontseré C, Alvarez-Estape M, Lester J, Arandjelovic M, Kuhlwil M, Dieguez P, et al. Maximizing the acquisition of unique reads in noninvasive capture sequencing experiments. *Mol Ecol Resour.* 2021;21:745–61. <https://doi.org/10.1111/1755-0998.13300>.
45. Kerviel A, Thomas A, Chaloin L, Favard C, Muriaux D. Virus assembly and plasma membrane domains: which came first? *Virus Res.* 2013;171:332–40. <https://doi.org/10.1016/J.VIRUSRES.2012.08.014>.
46. Lefranc MP. Immunoglobulin and T cell receptor genes: IMGT® and the birth and rise of immunoinformatics. *Front Immunol.* 2014;5(FEB):22. <https://doi.org/10.3389/fimmu.2014.00022>.
47. Saeed MF, Kolokoltsov AA, Freiberg AN, Holbrook MR, Davey RA. Phosphoinositide-3 kinase-akt pathway controls cellular entry of ebola virus. *PLoS Pathog.* 2008;4:e1000141. <https://doi.org/10.1371/journal.ppat.1000141>.
48. Ruibal P, Oestereich L, Lüdtko A, Becker-Ziava B, Wozniak DM, Kerber R, et al. Unique human immune signature of Ebola virus disease in Guinea. *Nat.* 2016;533:7601. <https://doi.org/10.1038/nature17949>.
49. Melin AD, Janiak MC, Marrone F, Arora PS, Higham JP. Comparative ACE2 variation and primate COVID-19 risk. *Commun Biol.* 2020;3:641.
50. Vallet D, Petit EJ, Gatti S, Levréro F, Ménard N. A new 2CTAB/PCI method improves DNA amplification success from faeces of Mediterranean (Barbary macaques) and tropical (lowland gorillas) primates. *Conserv Genet.* 2008;9:677–80. <https://doi.org/10.1007/s10592-007-9361-8>.
51. Meyer M, Kircher M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb Protoc.* 2010;2010:db.prot5448. <https://doi.org/10.1101/pdb.prot5448>.
52. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 2012;40:1–8.
53. Rohland N, Reich D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* 2012;22:939–46. <https://doi.org/10.1101/gr.128124.111>.
54. Hugerth LW, Wefer HA, Lundin S, Jakobsson HE, Lindberg M, Rodin S, et al. DegePrime, a program for degenerate primer design for broad-taxonomic-range PCR in microbial ecology studies. *Appl Environ Microbiol.* 2014;80:5116–23. <https://doi.org/10.1128/AEM.01403-14>.
55. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 2013;41:e1. <https://doi.org/10.1093/nar/gks808>.
56. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 2012;6:1621–4. <https://doi.org/10.1038/ismej.2012.8>.
57. Xu H, Luo X, Qian J, Pang X, Song J, Qian G, et al. FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads. *PLoS One.* 2012;7:1–6.
58. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics.* 2014;30:614–20.
59. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754–60.
60. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078–9.

61. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43:491–8.
62. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156–8.
63. Kuhlwil M, Fontseré C, Han S, Alvarez-Estape M, Marques-Bonet T. HuConTest: Testing Human Contamination in Great Ape Samples. *Genome Biol Evol.* 2021;13. <https://doi.org/10.1093/gbe/evab117>.
64. Meisner J, Albrechtsen A. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics.* 2018;210:719–31.
65. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
66. Vieira FG, Lassalle F, Korneliusen TS, Fumagalli M. Improving the estimation of genetic distances from Next-Generation Sequencing data. *Biol J Linn Soc.* 2016;117:139–49.
67. Paradis E, Claude J, Strimmer K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics.* 2004;20:289–90.
68. Korneliusen TS, Moltke I. NgsRelate: A software tool for estimating pairwise r from next-generation sequencing data. *Bioinformatics.* 2015;31:4009–11.
69. Hanghøj K, Moltke I, Andersen PA, Manica A, Korneliusen TS. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *Gigascience.* 2019;8(5):giz034. <https://doi.org/10.1093/gigascience/giz034>.
70. Nielsen R, Korneliusen T, Albrechtsen A, Li Y, Wang J. SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLoS One.* 2012;7(7):e37558. <https://doi.org/10.1371/journal.pone.0037558>.
71. Chen H, Boutros PC. VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R; 2011. <https://doi.org/10.1186/1471-2105-12-35>.
72. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17:17–122. <https://doi.org/10.1186/s13059-016-0974-4>.
73. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* 2012;40:452–7.
74. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, et al. PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.* 2003;13:2129–41.
75. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;14:R36. <https://doi.org/10.1186/gb-2013-14-4-r36>.
76. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31:166–9. <https://doi.org/10.1093/bioinformatics/btu638>.
77. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010;7:335–6. <https://doi.org/10.1038/nmeth0510-335>.
78. Aronesty E. ea-utils: “Command-line tools for processing biological sequencing data”; 2011.
79. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2017;26:2460–1.
80. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res Notes.* 2016;9:88. <https://doi.org/10.1186/s13104-016-1900-2>.
81. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Res.* 2007;17:377–86.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

