



Development of 79 SNP markers to individually genotype and sex-type endangered mountain gorillas (*Gorilla beringei beringei*)

Benjamin N. Sacks^{1,2} · Tianyi Hu^{1,3} · Elizabeth M. Kierepka¹ · Stevi L. Vanderzwan¹ · Jena R. Hickey^{1,4}

Received: 21 October 2020 / Accepted: 13 May 2021 / Published online: 20 May 2021
© The Author(s) 2021

Abstract

The mountain gorilla (*Gorilla beringei beringei*) is one of two endangered subspecies of eastern gorilla. The principle approach to monitoring the two extant mountain gorilla populations has been to use fecal surveys to obtain DNA profiles for individuals that are then used for capture-recapture-based estimates of abundance. To date, 11 to 14 microsatellites have been used for this purpose. To adapt to ongoing changes in genotyping technologies and to facilitate the analysis of fecal DNA samples by multiple laboratories, we developed a panel of single nucleotide polymorphism (SNP) markers that can be used for future gorilla monitoring. We used published short read data sets for 3 individuals to develop a suite of 79 SNPs, including two sex markers, for a Fluidigm platform. This marker set provided high resolution to differentiate individuals and will facilitate future monitoring, leaving room for additional SNPs to be included in a 96-assay format.

Keywords Fluidigm · *Gorilla beringei beringei* · Illumina · Single nucleotide polymorphism

The mountain gorilla (*Gorilla beringei beringei*) is one of two subspecies of eastern gorilla, and is currently recognized as endangered by the IUCN (Hickey et al. 2018). Today, two isolated populations of mountain gorilla remain, one in Bwindi Impenetrable National Park, Uganda and Sarambwe Nature Reserve, Democratic Republic of Congo (DRC) and the other in the Virunga Massif of Uganda, Rwanda, and DRC. Since the early 2000s, monitoring the abundance of these populations has involved regular noninvasive fecal DNA surveys (e.g., Guschanski et al. 2009). Genetic analyses are used to identify individuals from fecal DNA samples, which enable mark-recapture analyses to estimate total numbers of individuals within the population (Roy et al. 2014;

Hickey and Sollman 2018). In past surveys, microsatellite markers have been used for this purpose (Guschanski et al. 2009; Roy et al. 2014; Hickey and Sollman 2018; Granjon et al. 2020). Although these markers performed adequately in the past, the technology required to support their use is becoming obsolete and will eventually be unsupported (von Thaden et al. 2017). Additionally, microsatellites are challenging to calibrate between laboratories. Therefore, we developed a set of 79 SNPs with high resolution to differentiate even close relatives that can be used by any participating laboratory for future surveys of these and other eastern gorilla populations.

To discover SNPs, we obtained raw sequencing reads from entire genomes (80–98 Gbp each) of 3 female mountain gorillas from the Virunga Massif (none were available from Bwindi): Turimaso (ENA No. ERS525618), Maisha (ENA No. ERS525616), and Uririmo (ENA No. ERS525617) (Xue et al. 2015). We aligned trimmed reads to the western lowland gorilla (*Gorilla gorilla gorilla*) genome (gorGor5; Gordon et al. 2016) using bwa-mem (Li 2013), and called variants using freeBayes (v0.9.21-19-gc003c1e; Garrison and Marth 2012), which resulted in identification of 12,894,536 variable sites (i.e., SNPs). We then filtered SNPs using vcftools (Danecek et al. 2011), retaining biallelic SNPs with 3 of each allele (–mac 3) and with read depths ranging 20–40x (–minDP 20, –maxDP 40); the

✉ Benjamin N. Sacks
bnsacks@ucdavis.edu

¹ Mammalian Ecology and Conservation Unit, Veterinary Genetics Laboratory, School of Veterinary Medicine, University of California, Davis 95616, USA

² Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis 95616, USA

³ Forensic Sciences Graduate Group, University of California, Davis 95616, USA

⁴ International Gorilla Conservation Programme, Kigali, Rwanda

average read depths of genomes were ~25–30x, so our narrow depth range ensured removing sequence repeats and low-coverage sites. We further removed sites that were heterozygous in all 3 individuals, ensuring that all remaining sites had a representative of all 3 possible genotypes, resulting in 132,958 sites. We then mapped contigs to the human genome to obtain orthologous positions, and removed all sites that were on gorilla contiguous sequences < 3 Mbp long (Gordon et al. 2016), leaving 115,106 sites with locations on chromosomes that were known with high confidence. Finally, we selected 129 candidate SNPs that were distributed approximately evenly across the chromosomes (i.e., 37 Mbp [range: 16.5–58.5 Mbp] apart on average). To facilitate primer design, we also limited selection to SNPs for which the regions flanking them for 200 bp in 3' and 5' directions included no other known SNPs or indels (i.e., based on the unfiltered vcf file).

To design 2 sex markers (i.e., bringing the total to 131 candidate SNPs), we used sequences from human X chromosome and Y chromosome introns of the amelogenin and zinc finger genes (Kim et al. 2010). We used basic local alignment search tool (BLAST) in Genbank to obtain gorilla Y chromosome sequences that were orthologous to the Y chromosome introns of the amelogenin gene (Genbank accession Nos. FJ532255.1) and the zinc finger genes (AH014841.2 ZFY). However, this procedure produced no X chromosome orthologs for gorillas. Therefore, we used the human sequences for these X chromosome loci as references to extract reads from a female gorilla whole genome sequence (Turimaso) using bwa-mem. We manually aligned reads and BLAST product sequences, respectively, for gorilla X and Y chromosome paralogs of these genes to verify that the same 3 SNPs differing between human X and Y chromosome introns also differed between the gorilla X and Y chromosomes. However, we also discovered 3 additional variants in gorillas corresponding to sites in the published human SNP primers, which required us to redesign primers specifically for gorillas.

We used Fluidigm's D3 Assay Design Tool (<https://d3.fluidigm.com>) in conjunction with the flanking sequences for the 131 SNPs (Supporting Information, Supporting text 1) to design primers, which we ordered from Fluidigm (Fluidigm, San Francisco USA; Supporting information, Table S1). We used a set of 561 DNA samples extracted from mountain gorilla feces collected from nests in Bwindi during 2018 that had been individually identified with microsatellites and which included ≥ 1 sample from each of 450 putative individuals (i.e., all but one of 451 identified in the survey; Hickey et al. 2019; Hu 2020). This sample set included arbitrarily selected duplicates of the same DNA extracts and extracts from multiple fecal samples assigned with microsatellites to the same individual. After screening all candidate SNPs against 94 of these samples,

we retained 96 of the SNPs for further evaluation against the remaining 467 samples. From the 96 loci screened against the entire samples set, we selected 79 SNP loci based on consistent cluster separation and high call rates.

We used the 96.96 Fluidigm Dynamic Arrays with integrated fluidic circuits (IFCs) run on the Biomark HD system, initially following the manufacturer's guidelines, except that we used 18 specific target amplification (STA) PCR cycles (instead of 14) during a pre-amplification step to increase the concentration of marker-specific DNA to be used as template for the allele specific (ASP) reactions (von Thaden et al. 2017), and we diluted the STA PCR product 1:10 rather than 1:100 before adding to the ASP reactions; finally, we used 45 cycles instead of 38 cycles in the ASP PCR reaction. We called genotypes using Fluidigm SNP Genotyping Analysis Software v4.3.2. We estimated heterozygosity, unbiased probabilities of identity (P_{ID}) and identity of siblings (P_{IDsibs}) in Gimlet (v. 1.3.3; Valière 2002).

Based on the 79 final loci, we obtained > 95% call rates on 488 of the 561 (87%) fecal DNA samples. For 475 genotypes where both sex markers (GorAmelo2-CG, GorZF1-CT) yielded genotypes, 471 pairs (99.16%) agreed, of which 468 of these samples (99.14%) also agreed with sexes typed from the microsatellite study (Hickey et al. 2019), suggesting < 1% sex-typing error rate. Based on 21 pairs of replicate genotypes (same fecal extract), the overall agreement was 98.6% (SD = 0.007%), indicating a genotyping error rate < 2%. Similarly, 41 pairs of fecal samples previously determined based on microsatellites to be from the same individual matched at 99.3% (SD = 0.027) of their SNP alleles on average. Conversely, pairwise comparisons among genotypes from 424 putative distinct individuals (as identified through microsatellites) matched at 73.9% (SD = 0.038) of alleles on average. These metrics, along with the average expected and observed heterozygosity (0.34 and 0.33, respectively) and the combined P_{ID} and P_{IDsibs} (1.6×10^{-23} and 2.2×10^{-12} , respectively; Supplementary information, Table S2), indicate high resolution to differentiate individuals from the Bwindi population. Because the markers were designed using genomes of mountain gorillas from the Virunga Massif and because we did not screen markers on the basis of polymorphism in the Bwindi population, they should perform as well or better on gorillas from the Virunga population, for example exhibiting polymorphism in the 15 SNPs that were monomorphic in the Bwindi population. Despite ascertainment bias favoring the Virunga population in particular and mountain gorillas in general, these markers also would likely perform well for eastern lowland (*G. b. graueri*), and possibly western (*G. gorilla*) gorillas as well, due to their higher effective population sizes (Xue et al. 2015), although empirical confirmation would be necessary.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12686-021-01217-4>.

Acknowledgements Funding was from the International Gorilla Conservation Programme (Grant No. RW-CENS-18-001), and the University of California, Davis, Forensic Sciences Graduate Group. Dr. Andrea Schreier generously shared equipment, and Alisha Goodbla assisted with scheduling of equipment. The 2018 Bwindi-Sarambwe population surveys of mountain gorillas, known as census, were conducted by the Protected Area Authorities in Uganda and the Democratic Republic of Congo (Uganda Wildlife Authority and l’Institut Congolais pour la Conservation de la Nature) under the transboundary framework of the Greater Virunga Transboundary Collaboration. The census was supported by the Rwanda Development Board, International Gorilla Conservation Programme (a coalition of Conservation International, Fauna & Flora International and WWF), Mammalian Ecology and Conservation Unit (MECU) of the UC Davis Veterinary Genetics Laboratory, Max Planck Institute for Evolutionary Anthropology, The Dian Fossey Gorilla Fund, Institute of Tropical Forest Conservation, Gorilla Doctors, Conservation Through Public Health, Wildlife Conservation Society Uganda Country Office, WWF Uganda Country Office, and Bwindi Mgahinga Conservation Trust. The census was funded by Fauna & Flora International, WWF, and Partners in Conservation at the Columbus Zoo & Aquarium.

Funding See acknowledgments above.

Data availability Provided as supporting information.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare no conflicts.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

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/>.

References

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158

- Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv preprint [q-bio.GN]
- Gordon D, Huddleston J, Chaisson MJ, Hill CM, Kronenberg ZN, Munson KM, Malig M, Raja A, Fiddes I, Hillier LW, Dunn C (2016) Long-read sequence assembly of the gorilla genome. *Science* 352:aae0344
- Granjon AC, Robbins MM, Arinaitwe J, Cranfield MR, Eckardt W, Mburanumwe I, Musana A, Robbins AM, Roy J, Sollmann R, Vigilant L, Hickey JR (2020) Estimating abundance and growth rates in a wild mountain gorilla population. *Anim Conserv* 23:455–465
- Guschanski K, Vigilant L, McNeilage A, Gray M, Kagoda E, Robbins MM (2009) Counting elusive animals: comparing field and genetic census of the entire mountain gorilla population of Bwindi Impenetrable National Park, Uganda. *Biol Cons* 142:290–300
- Hickey JR, Basabose A, Gilardi KV, Greer D, Nampindo S, Robbins MM, Stoinski TS (2018) *Gorilla beringei* ssp. *beringei*. The IUCN Red List of Threatened Species 2018: e.T39999A17989719. <https://doi.org/10.2305/IUCN.UK.2018-2.RLTS.T39999A17989719.en>
- Hickey JR, Uzabaho E, Akantorana M, Arinaitwe J, Bakebwa I, Bitariho R, Eckardt W, Gilardi KV, Katutu J, Kayijamahe C, Kierepka EM, Mugabukomeye B, Musema A, Mutabaazi H, Robbins MM, Sacks BN, Zikusoka GK (2019) Bwindi-Sarambwe 2018 Surveys: monitoring mountain gorillas, other select mammals, and human activities. GVTC, IGCP & partners, Kampala, Uganda, 40p. http://igcp.org/wp-content/uploads/Bwindi-Sarambwe-2018-Final-Report-2019_12_16.pdf
- Hickey JR, Sollman R (2018) A new mark-recapture approach for abundance estimation of social species. *Plos One*. <https://doi.org/10.1371/journal.pone.0208726>
- Hu T (2020) Developing and applying single nucleotide polymorphisms (SNP) markers for a noninvasive genetic survey of mountain gorillas (*Gorilla beringei beringei*), Thesis, University of California, Davis, USA
- Kim JJ, Han BG, Lee HI, Yoo HW, Lee JK (2010) Development of SNP-based human identification system. *Int J Legal Med* 124:125–131
- Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. [q-bio.GN]
- Roy J, Vigilant L, Gray M, Wright E, Kato R, Kabano P et al (2014) Challenges in the use of genetic markrecapture to estimate the population size of Bwindi mountain gorillas (*Gorilla beringei beringei*). *Biol Cons* 180:249–261
- Valière N (2002) GIMLET: a computer program for analysing genetic individual identification data. *Mol Ecol Notes* 2:377–379
- von Thaden A, Cocchiararo B, Jarausch A et al (2017) Assessing SNP genotyping of noninvasively collected wildlife samples using microfluidic arrays. *Sci Rep* 7:10768
- Xue Y, Prado-Martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, Frandsen P, Chen Y, Yngvadottir B, Cooper DN, de Manuel M, Hernandez-Rodriguez J, Lobon I, Siegmund HR, Pagani L, Quail MA, Hvilson C, Mudakikwa A, Eichler EE, Cranfield MR, Marques-Bonet T, Tyler-Smith C, Scally A (2015) Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* 348:242–245

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