Barriers to chimpanzee gene flow at the south-east edge of their distribution

Noémie Bonnin1 | Alex K. Piel2 | Richard P. Brown1 | Yingying Li3 | Andrew Jesse Connell3 | Alexa N. Avitto3 | Jean P. Boubli4 | Adrienne Chitayat5 | Jasmin Giles3 | Madhurima S. Gundlapally3 | Iddi Lipende6 | Elizabeth V. Lonsdorf7,8 | Deus Mjungu9 | Dismas Mwacha9 | Lilian Pintea10 | Anne E. Pusey11 | Jane Raphael12 | Serge A. Wich1,5 | Michael L. Wilson13,14,15 | Emily E. Wroblewski16 | Beatrice H. Hahn3 | Fiona A. Stewart1,2

1School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, UK
2Department of Anthropology, University College London, London, UK
3Departments of Medicine and Microbiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA
4School of Science, Engineering & Environment, University of Salford, Salford, UK
5Institute of Biodiversity and Ecological Dynamics, University of Amsterdam, Amsterdam, The Netherlands
6Tanzania Wildlife Research Institute (TAWIRI), Arusha, Tanzania
7Department of Psychology, Franklin and Marshall College, Lancaster, Pennsylvania, USA
8Department of Anthropology, Emory University, Atlanta, Georgia, USA
9Gombe Stream Research Centre, The Jane Goodall Institute–Tanzania, Kigoma, Tanzania
10Conservation Science Department, The Jane Goodall Institute, Washington, District of Columbia, USA
11Department of Evolutionary Anthropology, Duke University, Durham, North Carolina, USA
12Gombe National Park, Kigoma, Tanzania
13Department of Anthropology, University of Minnesota, Minneapolis, Minnesota, USA
14Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, Minnesota, USA
15Institute on the Environment, University of Minnesota, St. Paul, Minnesota, USA
16Anthropology Department, Washington University, St. Louis, Missouri, USA

Abstract

Populations on the edge of a species’ distribution may represent an important source of adaptive diversity, yet these populations tend to be more fragmented and are more likely to be geographically isolated. Lack of genetic exchanges between such populations, due to barriers to animal movement, can not only compromise adaptive potential but also lead to the fixation of deleterious alleles. The south-eastern edge of chimpanzee distribution is particularly fragmented, and conflicting hypotheses have been proposed about population connectivity and viability. To address this uncertainty, we generated both mitochondrial and MiSeq-based microsatellite genotypes for 290 individuals ranging across western Tanzania. While shared mitochondrial
haplotypes confirmed historical gene flow, our microsatellite analyses revealed two distinct clusters, suggesting two populations currently isolated from one another. However, we found evidence of high levels of gene flow maintained within each of these clusters, one of which covers an 18,000 km² ecosystem. Landscape genetic analyses confirmed the presence of barriers to gene flow with rivers and bare habitats highly restricting chimpanzee movement. Our study demonstrates how advances in sequencing technologies, combined with the development of landscape genetics approaches, can resolve ambiguities in the genetic history of critical populations and better inform conservation efforts of endangered species.

**KEYWORDS**
biogeography, genetic diversity, great apes, microsatellites, mitochondrial DNA, Tanzania

# INTRODUCTION

Barriers to gene flow can profoundly influence the future of a species. Without genetic exchange, neighbouring populations accumulate genetic differences through genetic drift and natural selection that can favour local adaptation and ultimately result in divergence and speciation (Coyne, 1992). However, genetic isolation can also lead to the fixation of deleterious alleles, potentially accelerating the extinction of small populations, a process described as the ‘extinction vortex’ (Gilpin & Soulé, 1986). This is particularly relevant as species are facing unprecedented climatic and environmental pressures in an increasingly fragmented landscape (Chase et al., 2020; Haddad et al., 2015; Pimm et al., 2014). Maintaining intraspecific genetic diversity and ensuring connectivity have been recognized as key conservation goals in the post-2020 Global Biodiversity Framework (CBD, 2021). By investigating gene flow and barriers to animal movement, scientists can provide conservationists with the information required to mitigate the deleterious effects of isolation (Frankham, 2005, 2015).

Landscape genetics methods, which combine techniques from landscape ecology and population genetics, have been developed to help scientists assess the influence of landscape features on gene flow and genetic diversity (Manel et al., 2003; Storfer et al., 2007). These approaches are powerful and accurate tools used to detect barriers to animal movement and have been used across a wide range of species, from insects (Marchi et al., 2013; Trense et al., 2021), to large terrestrial mammals (Baden et al., 2019; Connor et al., 2021; Epps et al., 2013; Fedorca et al., 2019).

Geographic barriers are thought to have played an important role in determining genetic diversity of chimpanzees (*Pan troglodytes*), with the distribution of the four recognized subspecies correlating with known natural barriers (Fontsere et al., 2022; Gonder et al., 2006; Lester et al., 2021; Mitchell et al., 2015). Despite being widely distributed across Africa, all subspecies are classified as either ‘Endangered’ (i.e. Nigeria-Cameroon chimpanzees *P. t. ellioti*, central chimpanzees *P. t. troglodytes*, and eastern chimpanzees *P. t. schweinfurthii* or ‘Critically Endangered’ (i.e. Western Chimpanzees *P. t. verus*) by the IUCN Red List of Threatened Species ([www.iucnredlist.org](http://www.iucnredlist.org)). Poaching, infectious disease and the loss and fragmentation of their habitats are the leading causes of their decline (Keele et al., 2009; Kühl et al., 2017; Strindberg et al., 2018). Climate change is an additional threat of increasing concern, shifting suitable habitats and resource availability and subsequently leaving individuals with no other choice than to adapt or migrate (Carvalho et al., 2019; Davis et al., 2005; Korstjens & Hillyer, 2016; Lehmann et al., 2010).

Chimpanzees at the edge of their distribution often occur close to the limits of their ecological range (van Leeuwen et al., 2020) and are therefore likely to experience different regimes of natural selection (Lesica & Allendorf, 1995). Therefore, they may represent an important source of adaptive diversity, yet these populations are often fragmented and more likely to be geographically isolated (Lester et al., 2021; Plumptre et al., 2010). Tanzania marks the south-eastern limit of *Pan* distribution with a census population of ~3000 wild chimpanzees (*P. t. schweinfurthii*; Carvalho et al., 2022; Humle et al., 2016; Moyer et al., 2006; Piel & Stewart, 2014). Despite two national parks – Gombe National Park (GNP) and Mahale Mountains National Park (MMNP) – and a network of village and district forest reserves protecting wild chimpanzee habitats in the country, regional habitat loss and fragmentation raise serious concerns about population isolation and long-term viability (Lasch et al., 2011; TAWIRI, 2018). Along with habitat loss, poaching is also threatening Africa’s most southerly chimpanzee population, Lwazi, where ~100 chimpanzees survive under heavy anthropogenic pressures (Davenport et al., 2010; Ogawa, 1997).

The majority (75%) of Tanzanian chimpanzees are found at low density within an 18,000 km² ecosystem known as the Greater Mahale Ecosystem (GME) and regarded as one of the driest and most seasonal habitats in which the species is found (Lindshield et al., 2021; van Leeuwen et al., 2020). Dominated by miombo woodland, the landscape is thought to have fostered greater behavioural diversity compared to chimpanzee populations inhabiting more ‘typical’ forested landscapes (Kalan et al., 2020). Along
with their genetic diversity, this behavioural diversity is now under pressure from habitat fragmentation (Kühl et al., 2019).

Initial studies on gene flow and genetic diversity of Tanzanian chimpanzees have provided contrasting findings regarding population connectivity. Based on analysis of mitochondrial DNA (mtDNA), GME chimpanzees were first described as one panmictic population (Inoue et al., 2011), while later studies suggested potential barriers restricting chimpanzee movements between the northern and southern parts of the ecosystem (Moyer et al., 2006; Piel et al., 2013; Rudicell et al., 2011). The Malagarasi River, a pre-riift time tributary of the Congo River (Kullander & Roberts, 2011), was originally thought to represent a barrier to chimpanzee movement, separating GNP and GME chimpanzees. Its permeability was later argued by Piel et al. (2013), who reported shared mtDNA haplotypes between chimpanzees in GNP and the GME and circumstantial evidence that chimpanzees could cross the river using natural, shallow fords. Additionally, chimpanzees in GNP and northern GME are infected with closely related strains of the simian immunodeficiency virus (SIVcpz; Rudicell et al., 2011), which suggests, at the very least, historical movement across the river.

More recently, two studies of chimpanzee connectivity across Africa reported that chimpanzees from Issa, a sampled community from western Tanzania, used as a representative of the GME, clustered separately from other eastern chimpanzee sites (figures S36 and S45 in Fontsere et al., 2022 and figure S2 in Lester et al., 2021). Based on microsatellite markers, Lester et al. (2021) identified Issa as a genetic outlier, divergent from all other sampled chimpanzees regardless of distance (Lester et al., 2021). The authors suggested a loss of genetic diversity due to small population size. Genomic analyses by Fontsere et al. (2022), based on chromosome 21 sequences, demonstrated identical-by-descent-like shared fragments between Issa and communities from Rwanda and Uganda suggesting that Issa chimpanzees have been connected since the Last Glacial Maximum with northern populations and have been isolated only recently (~1000 years ago). Given the increasing pressure on chimpanzee habitat, clarifying connectivity and genetic diversity is vital to the species’ long-term survival in this part of their range. Here, we generated an extensive dataset using both mtDNA and 10 MiSeq-based microsatellite loci (sequenced genotypes) to clarify the population genetic structure and diversity of chimpanzees living at the south-east edge of their distribution. We also applied landscape genetics approaches to detect barriers to gene flow and estimate their resistance to chimpanzee movement. We predicted that genetic structure should align with previously proposed barriers to movement and lead to lower genetic diversity in isolated populations. Our findings confirm part of our predictions and demonstrate the importance of using multiple molecular markers and integrating population and landscape genetics approaches to determine conservation priorities. This is especially important when traditional population genetic analyses that rely on a single marker have been unable to resolve previous uncertainties regarding the connectivity of critical populations.

2 | MATERIALS AND METHODS

2.1 | Study species

Chimpanzees live in multi-male, multi-female communities of about 20 to 200 individuals (Watts, 2012; Wilson, 2012). Each community exhibits fission-fusion grouping, whereby individuals form temporary parties of various age/sex compositions (Goodall, 1968; Nishida, 1968). Once reaching sexual maturity, female chimpanzees generally disperse from their natal communities and reproduce in another community (Emery Thompson, 2013). Wild chimpanzees can live up to 60 years (Emery Thompson et al., 2007). They feed principally on ripe fruits, but also eat plant parts, insects and vertebrate prey (Wrangham, 1975). To meet their nutritional needs, communities range across large territo ries spanning from 5 km² in moist, forest environments (Williams et al., 2002) to ~90 km² in drier savanna landscapes (Pruetz & Herzog, 2017).

2.2 | Sampling

We analysed 234 faecal samples from 16 different sampling locations within the GME (Figure 1). For each sample, we collected approximately 15 g of chimpanzee faeces into a tube containing 15-20 mL of RNAlater (Ambion), which was frozen on site at approximately −20°C on the day of collection. Samples were subsequently frozen at −80°C after transportation to the laboratory.

We also included genotypes from 136 individuals from GNP. Samples from habituated GNP chimpanzees have been collected since 2002 for SIVcpz diagnostics (Keele et al., 2009; Rudicell et al., 2011) and 19 were previously genotyped using the MiSeq-based approach for earlier analyses (Barbian et al., 2018). An additional 177 individuals were genotyped using a MiSeq-based approach for this study with samples collected through 2022. Because GNP samples originated from a long-term study of GNP chimpanzees during which females arrived from outside of the park or changed communities, we assigned individuals to the community in which they were born or first sampled. For the purposes of this study, the three GNP chimpanzee communities (i.e. Mitumba (MT), Kasekela (KK) and Kalande (KL)) are referred to as ‘sample sites’; GNP and GME are hereafter referred to as habitat patches.

2.3 | Molecular techniques

2.3.1 | DNA extraction

We extracted faecal DNA using the QIAamp Fast DNA Stool Mini Kit and the automated QIAcube system (Qigagen). Briefly, 400 µL of faecal RNAlater mixture was resuspended in InhibitEx buffer, clarified
by centrifugation, treated with proteinase K, and passed through a DNA binding column. Bound DNA was finally eluted in 200 μL elution buffer.

2.3.2 | mtDNA sequencing

Polymerase chain reaction (PCR) was carried out to amplify a 498 bp fragment of the first hypervariable control region (HV1) of the mitochondrial genome using the primers L15997 (5′-CACCATAGCACC AAGCT-3′) and H16498 (5′-CCTGAAGT AGGAACCAGATG- 3′; Keele et al., 2006). PCR conditions followed those described by Morin et al. (1994) except that an annealing temperature of 55°C was used and 55 amplification cycles were performed. PCR amplicons were gel-purified and sequenced directly with L15997 used as the sequencing primer by the Sanger method. Quality control was performed in Genious using a QC threshold of 30 to remove ambiguous base pair (Illumina, 2011). We assembled and aligned the resulting sequences with Mega 7.0.26 (Kumar et al., 2016), along with geo-referenced sequences from previous studies (Keele et al., 2006; Liu et al., 2008; Rudicell et al., 2011).

2.3.3 | Microsatellite genotyping

Microsatellites were preferred to single nucleotide polymorphisms (SNPs) as this allowed the inclusion of available GNP data, which increased the geographic scope of this study. Furthermore, microsatellites have been shown to perform comparably to SNPs for quantifying population divergence (Lemopoulos et al., 2019).

Here, we used MiSeq sequencing of microsatellites using the approach developed by Barbian et al. (2018). Sequencing of microsatellites allowed us to overcome common drawbacks of more commonly used capillary electrophoresis-based microsatellites: amplification artefacts, imprecise sizing, length homoplasy and limited multiplex capability. The MiSeq-based approach reduces PCR artefacts, allows fragment sizes to be precisely determined, discerns cryptic alleles that would have been hidden by length homoplasy and efficiently allows multiplexing of several individuals and loci into a single high-throughput run. By uncovering a greater number of unique alleles, high-throughput microsatellite genotyping can also better resolve population genetic structure by discriminating populations that appear to cluster together when using length-based microsatellites (Darby et al., 2016). While length-based microsatellites are generally not comparable across
laboratories due to differing allelic designations between different sequencing machines (Ellis et al., 2011). MiSeq-based microsatellite genotyping can be effectively shared and compared across multiple studies and laboratories (Barbian et al., 2018).

We amplified 10 polymorphic autosomal microsatellite loci (Table S1) following the MiSeq-based approach developed by Barbian et al. (2018). Briefly, loci were amplified in a one-step multiplex reaction in three independent PCRs. PCR products were then combined in equal volume and diluted in nuclease-free sterile water (1:10) prior to MiSeq sequencing. Read files were subsequently processed using the CHIIMP analysis pipeline (Barbian et al., 2018). Homozygous genotypes and rare alleles (i.e. only represented by a single individual) were confirmed by sequencing the amplicons from at least two independent multiplex reactions.

2.4 | Dataset preparation

The CHIIMP pipeline generates allele codes based on sequence length and content. To fit the required integer input format of most population genetic software, we used sequential index numbers as identifiers.

Unhabituated chimpanzees were sampled non-invasively across the GME, so some individuals may have been sampled more than once. We conducted an identity analysis in CERVUS 3.0.7 (Kalinowski et al., 2007) to distinguish individuals. We determined that eight loci were necessary to obtain a probability of detecting identity among sibs to be <0.001 (Waits et al., 2001). Genotypes that mismatched at one or two loci were re-examined for possible genotyping errors or allelic dropout before merging the data into consensus genotypes. From the 234 samples collected across the GME, we identified 154 individuals.

The presence of related individuals can increase the signal of genetic differentiation. We thus calculated pairwise estimates of relatedness for all individuals (GNP and GME) using the R package RELATED (Pew et al., 2015). We used the package function ‘compar_estimators’ to test the performance of different estimators on simulated data. The function uses allele frequencies to generate virtual pairs of individuals with specified genetic relationships (i.e. parent-offspring, full-sibs, half-sibs and unrelated) and estimates the Pearson’s correlation coefficient between observed and expected relatedness values for each estimator. Wang’s estimator (Wang, 2002) showed the highest correlation coefficient and was used to infer first degree relatives ($r \geq 0.5$). We used 1000 bootstrap permutations to estimate 95% confidence intervals for the relatedness index and found 17 pairs of individuals in which the lower 95% confidence limit was $\geq 0.5$. We then examined the effects of relatedness by removing 14 individuals (5 GME and 9 GNP) identified within these pairs. We performed subsequent population genetic structure analyses using both sets of data (i.e. 276 unrelated individuals and the full dataset of 290) and found no effect when adding related individuals. Hence, we present the results from the full dataset in the main text (results from the ‘unrelated dataset’ can be found in Figure S1).

We filtered potential stutter sequences and other PCR artefacts through the CHIIMP pipeline (Barbian et al., 2018) and additionally tested the presence of possible null alleles, large allelic dropout or scoring error due to stuttering using MICRO-CHECKER 2.2.3 (Van Oosterhout et al., 2004). None of the loci showed evidence of genotyping errors across the GNP or the GME. We also tested for linkage disequilibrium (LD) and deviation from Hardy–Weinberg equilibrium (HWE) between all pairs of loci using GENEPOP 4.7.5 (Rousset, 2008). We observed no deviation from HWE and no evidence of LD after sequential Bonferroni correction when GNP and the GME were analysed separately, thus justifying the use of all loci.

2.5 | Population structure and genetic diversity

To identify genetic structure, we performed clustering analyses for our mtDNA sequences using BAPS 6.0 (Cheng et al., 2013; Corander et al., 2003) and for our microsatellites genotypes using STRUCTURE 2.3.4 (Pritchard et al., 2000).

Using BAPS, we performed genetic mixture analyses without prior information on geographic location. We used an upper bound of $K = 20$ (the number of haplotypes detected) and performed 5 independent runs for each value of $K$. Results from mixture clustering were used to determine the optimal number of clusters and were followed by an admixture analysis.

Similarly, we estimated microsatellites clustering via STRUCTURE with an upper bound of $K = 19$ (the number of sampling sites) and without prior information on geographic location. Because STRUCTURE tends to underestimate the number of contributing populations when using unbalanced sample sizes (full representation of GNP communities while ~10% of GME chimpanzees sampled), we used the recommendations by Wang (2017) and specified the population-specific ancestry prior, decreased the initial $\alpha$ to $1/K$ and used the uncorrelated allele frequency model. We conducted 10 independent runs with 100,000 Markov Chain Monte Carlo (MCMC) iterations, after an initial burn-in of 10,000. Independent runs within each cluster were merged and visualized using the R package POPHELPER (Francis, 2017). Optimal $K$ values were determined using POPHELPER based on the log probability of the data $\ln \Pr (X|K)$ (Pritchard et al., 2000) and the ad hoc statistic $\Delta K$ which is the rate of change in the log probability (Evanno et al., 2005). To investigate substructure, main clusters were subsequently analysed using the same method. As all individuals were available for GNP communities, Wang (2017) recommendations were not applied for clusters comprising only GNP individuals.

We also constructed a Median-joining haplotype network (Bandelt et al., 1999) using POPART 1.7 (Leigh & Bryant, 2015) to visualize mtDNA haplotype partitioning and performed a principal component analysis (PCA) using the R package ADEGENET (Jombart, 2008) to summarize the microsatellite genetic variability across Tanzania.

We used ARLEQUIN 3.5 (Excoffier & Lischer, 2010) to perform an analysis of molecular variance (AMOVA; Excoffier et al., 1992)
and estimate measures of mtDNA molecular diversity (haplotype diversity (h) and nucleotide diversity (SD)). To evaluate differences in molecular diversity between clusters, we performed a permutation test with 10,000 replicates using a custom R script (Alexander, 2015).

We obtained general statistics of microsatellite diversity including number of alleles (Na), rarefied allelic richness (Ar), expected and observed heterozygosity (Hₑ and Hₒ, respectively) and inbreeding coefficients (F₁ₛ) using the R package DIVERSITY 1.9.90 (Keenan et al., 2013). The divBasic function calculates Ar by normalizing all populations to the smallest sample size and subsampling 1000 times, with replacement, and thus provides 95% confidence intervals. We tested for significant differences in allelic richness and observed heterozygosity between clusters using t-tests or Wilcoxon signed rank tests depending on equality of variances and normality of differences (Alexander et al., 2016).

To examine how Tanzanian populations cluster within the subspecies range and compare genetic diversity across eastern chimpanzees, we combined our data with published eastern chimpanzee HV1 mtDNA sequences (Figure S3; Table S2). We obtained sample numbers and locations for each published haplotype from previous studies (Inoue et al., 2011; Li et al., 2012; Liu et al., 2008) and performed population structure and genetic diversity analyses across the eastern chimpanzee range. As the samples in these studies were collected non-invasively, there is a possibility that some samples represent the same individual. Additionally, some of the sequences were up to 167bp shorter, especially those retrieved from Lwazi (Inoue et al., 2011), and may have been missing some of the variable sites detected in our haplotypes. While acknowledging these limitations, we present the results of these analyses in the supplemental Information (Figures S4 and S5 and Tables S3–S5) to provide important context for newly sequenced samples.

2.6 | Landscape genetic analysis

To test for the presence of isolation by distance (IBD), we performed a Mantel test using the R package ADEGENET (Jombart, 2008). Genetic distances were expressed as the proportion of shared alleles between individuals and Euclidean geographic distances between each location. Because correlation between genetic and geographic distances can occur under different biological scenarios (e.g., continuous clines or distant patches), we visualized local densities by scatterplots of two-dimensional kernel density estimations of genetic and geographic distances using the R package MASS (Venables & Ripley, 2002).

To further investigate spatial patterns of genetic variability, we performed a multivariate spatial analysis using MULTISPATI-PCA (Dray et al., 2008) implemented in the R package ADESPATIAL (Dray et al., 2018). This approach constructs vectors that maximize the product of the variance and the pattern of spatial autocorrelation (measured by Moran’s I) among individuals. Spatial information is introduced through a spatial weighting matrix built from a connection network. Here we used a distance-based connection network with a maximum distance between any two neighbours of 8km (representing the maximum diameter calculated from known chimpanzee home ranges in Tanzania, Table S6). Monte Carlo tests (1000 permutations) were used to assess the existence of ‘global’ and ‘local’ spatial structures (these terms refer to genetic divergence between non-neighbours and neighbours, respectively, relative to those between randomly-selected individuals). To detect potential barriers to gene flow, we estimated effective migration surfaces using the program EEMS (Petkova et al., 2016). EEMS infers relative effective migration rates between evenly spaced demes and produces visualizations that highlight deviations from migration rates under isolation by distance alone. We used pilot runs to determine the appropriate Markov Chain Monte Carlo (MCMC) iterations and burn-in as well as the minimum number of demes. We subsequently performed two runs for each chosen number of demes (500, 800 and 1000) with a burn-in of $5 \times 10^5$ and $2 \times 10^6$ MCMC iterations. We averaged the results of the final six iterations to plot the effective migration and diversity surfaces.

To evaluate the influence of specific landscape features on chimpanzee movements, we used the R package RESISTANCEGA (Peterman, 2018; Peterman et al., 2014). The package uses a genetic algorithm to parameterize resistance surfaces based on pairwise genetic data. The optimization process generates several resistance surfaces and starts by assigning random values to landscape features. Then, a linear mixed effects model with a maximum likelihood population effects parameterization (MLPE) is used to determine the relationship between pairwise genetic distance and pairwise cost distances. The best model is retained, and new resistance surfaces are iteratively generated from the best previous resistance surfaces until there is no further improvement to the relationship between genetic and geographic distances. Here, we estimated genetic distances based on the proportion of shared alleles between individuals found within the same raster cell (Dps; Bowcock et al., 1994) using the R package GRAPH4LG (Savary et al., 2021). This allowed grouping of individuals from known GNP communities and avoided the need to group GME individuals a priori, based on subjective sample sites. The index was initially developed as an inter-individual genetic distance but has been adapted for inter-population comparison and used repeatedly in landscape genetics (Balkenhol et al., 2016). We optimized resistance surfaces based on the commuteDistance function as it is substantially faster than the optimization with the equivalent Circuitscape program (Peterman, 2018). We considered eight landscape surfaces (Table S7; Figure S6). Distance from riparian forests, elevation, distance from steep slopes and proportion of riparian forests were used to allow direct comparison with a recent study modelling habitat connectivity for chimpanzees in the GME (Bonnin et al., 2020). Original variables were resampled at 1km resolution as a trade-off between retaining detail across the landscape and minimizing processing time for analyses. Vegetation layers were derived using a landcover product created from 1973 Landsat MSS images (Bonnin et al., 2020). To
avoid losing important narrow riparian forests from our original
categorical vegetation layers, we created new layers reflecting the
proportion of each vegetation class within 1 km² using focal sta-
(
statistics in ArcGIS Desktop. Major roads and rivers were also consid-
erned based on their known influence on chimpanzee movement.
We chose not to include current vegetation surfaces and human
population surfaces as the region was historically sparsely popu-
ulated (Kano, 1971). That is, population growth since the 1980s
and recent habitat loss is unlikely to have shaped the current ge-
netic structure of chimpanzees given their long generation time
(Langergraber et al., 2012). The package also includes a surface
based on Euclidean distance only to identify the proportion of our
data influenced by geographic distance alone.

We first optimized each surface separately using the SS_
optim function considering all possible resistance transformation
equations implemented in the package. We ranked the perform-
ance of these single surfaces using AICc (Akaike’s Information
Criterion corrected for small/finite population size) and consid-
ered models with ΔAICc < 2 to be competing models (Burnham
& Anderson, 2002). We then used the MS_optim function to opti-
imize every possible combination of the competitive variables and
ranked all final single-surface and multi-surface variable combi-
nations according to AICc. We performed analyses both across
the full dataset and within the GME. For the latter, we used the
inter-individual version of Dps, since this provided a better fit to
the sampling design across the ecosystem. We also included the
connectivity model outputs for 1973 from Bonnin et al. (2020)
as landscape surfaces to test whether habitat connectivity could
better explain the genetic structure of chimpanzee within the
ecosystem. Finally, we performed a pseudo-bootstrap procedure
using the resist.boot function to assess the relative support for
each optimized resistance surface. During this procedure, sample
locations and resistance distance matrices are sub-sampled and
fitted to the MLPE model using the previously optimized resist-
ance surfaces. We ran 1000 bootstrap iterations with 75% of the
observations re-sampled each time. The bootstrap function was
only used within the GME as resampling a proportion of the
observations was likely to remove the three sample locations rep-
resenting GNP.

3 | RESULTS

3.1 | Population genetic structure and genetic
diversity

Mitochondrial and microsatellite data revealed different levels of
genetic structure. The BAPS mixture analyses indicated that the
mtDNA sequences were optimally described by K=11 clusters
(highest Log likelihood). The BAPS plot illustrates some geographic
structuring with most clusters mainly found in GNP or in the GME
(Figure 2a). However, individuals did not cluster by habitat patch (i.e.
GNP and GME) at K=2.

The median-joining haplotype network revealed a similar pat-
tern, although the majority of mtDNA haplotypes (16 out of 20)
were specific to either GNP or the GME. Haplotypes did not form
geo-graphically partitioned haplogroups, with GNP and the GME
clustering together either by haplotype sharing or sequence simi-
larly (Table S8; Figure 2b). Haplotype divergence was generally
low (1–4 mutational steps) except for UG59 which was more diver-
gent (7 mutations). We also observed two predominant haplotypes
(GM7 and MH32) shared by two-thirds of the individuals sampled
within the GME, while a balanced distribution was found within GNP
(Figure 2b).

The STRUCTURE analysis favoured K=2 microsatellite clus-
ters (Figure S7). Individuals were correctly assigned to either GNP
or GME with high probability (mean estimate membership coef-
ficient q=0.99; Figure 2c). The highest proportion of admixture was
25% for one individual sampled in KL. The use of higher values of
K did not reveal additional substructure (Figure S8). The PCA cor-
rorated STRUCTURE results and clearly separated GNP and the
GME along the first axis. The first component explained 11.95% of
the total variation, while the second axis represented 4.35% of the
variation. Separate PCAs on GNP and the GME did not reveal clear
evidence of substructure (Figure S9), nor did running STRUCTURE
within each cluster separately (Figure S10).

Analysis of molecular variation (AMOVA) revealed that most of
the mtDNA and microsatellite variation was found among indi-
viduals within sample sites (i.e. GNP communities and GME sample
sites), both across and within GNP and the GME (Table S9). Genetic
variation between GNP and the GME accounted for 15.99% of
the total mtDNA variation and 9.63% of the total microsatellite
variation. The fixation indices between habitat patches were
moderate but significant for both markers (mtDNA: ΦCT = 0.160;
p-value = 0.003; microsatellites: FCT = 0.096; p-value < 0.01). Genetic
differentiation between sample sites was higher within the GME
than within GNP (Table S9). Only 2.47% (for mtDNA) and 0.32%
(for microsatellites) of the total genetic variation found within the
GME was attributed to a north–south separation and the corre-
sponding fixation index was not significant (p-value ΦST = 0.171; p-
value FST = 0.074).

Measures of mtDNA molecular diversity were within the range of
the other eastern chimpanzee populations (Tables S5 and S10) and
significantly higher in GNP than in the GME (Haplotypic diversity:
HGNP = 0.908 and HGME = 0.752, p-value < 0.001; nucleotide diversity:
SDGNP = 0.016 and SDGME = 0.014, p-value = 0.002). Among sample
sites, all molecular diversity indices were highest in MT (H = 0.897,
SD = 0.016, MPD = 8.138) and lowest in ISB (H = 0.200, SD = 0.001,
MPD = 0.040).

All 10 microsatellite loci were polymorphic with 8–22 alleles
each (Figure S11, Table S11). Allelic richness and observed het-
erozygosity did not differ significantly between GNP and the
GME (ArGNP = 9.24 and ArGME = 9.25, t(19) = −0.014, p-value = 0.989;
H0GNP = 0.803 and H0GME = 0.811, t(9) = −0.291, p-value = 0.777).
We observed no significant population-level inbreeding within
habitat patches (i.e. GNP and GME) or sample sites (i.e. GNP
communities and GME sample sites) based on $F_{ST}$ confidence intervals overlapping with negative values in all cases (Table S12).

### 3.2 | Barriers to gene flow and estimated resistances

Genetic distances were significantly correlated to geographic distances ($r = 0.50$, $p$-value < 0.001); however, the kernel density plot showed a clear pattern of discontinuity with two clouds of points (Figure S12). This rejects the presence of a clinal population structure by IBD and supports spatial genetic structure.

Both Multispati and EEMS analyses corroborated this result and identified a barrier to gene flow between GNP and the GME. Multispati analysis revealed a significant global structure ($p < 0.001$) with the first positive axis explaining most (48.5%) of the variation (Figure S13). A clear genetic break between the two habitat patches was evident following interpolation of vector scores from the first positive axis (Figure 3a). In contrast, no local structure was detected across the full dataset ($p = 0.110$) and repeating the analyses within the GME did not reveal significant substructure ($P_{Global} = 0.120, P_{Local} = 0.129$). EEMS analyses detected deviations from isolation-by-distance, with a lower-than-expected migration rate between GNP and the GME and a higher-than-expected migration rate along Lake Tanganyika within the GME (Figure 3b).

To investigate how landscape features impact chimpanzee movements, we used ResistanceGA. The package optimized the resistance of eight landscape surfaces and identified rivers and proportion of bare habitat (i.e. areas with tree cover less than 25%) as the best predictive features to explain the genetic structure observed, with competing predictive power ($\Delta AICc < 2$, Table 1). Optimized multi-surface composite of rivers and bare habitat explained less genetic variation than the single-surface models (Table S13). All other surfaces had $\Delta AICc > 33$ relative to the top ranked model, but proportion of woodland, distance from riparian forest, roads and distance from steep slope had higher explanatory power than the distance model alone ($\Delta AICc > 2$ relative to the distance model). Rivers resisted gene flow with an optimized cost of 1659 for the Malagarasi River and 4 for Lugufu (cost is dimensionless but optimized costs provide a relative scale to contextualize and compare landscape features tested, Table 1). Cost increased with proportion of bare habitat, distance from forest and distance from steep slope, while...
it decreased with proportion of woodland, elevation values and proportion of forest (Table 1, Figure S15).

We repeated the analyses within the GME. Here, the distance model had greatest explanatory power, ranking as the top model in 89% of bootstrap samples. The habitat connectivity models generated from Bonnin et al. (2020) ranked below the distance model ($\Delta$AICc > 3) and were not supported by the bootstrap iterations (Table S14).

### Table 1. Model output from ResistanceGA.

<table>
<thead>
<tr>
<th>Landscape surfaces</th>
<th>Avg. AICc</th>
<th>$\Delta$AICc</th>
<th>Avg.weight</th>
<th>k</th>
<th>Resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rivers</td>
<td>$-6215.41$</td>
<td>0.00</td>
<td>0.427548</td>
<td>4</td>
<td>Malagarasi: 1659; Lugufu: 4</td>
</tr>
<tr>
<td>Proportion of bare habitat</td>
<td>$-6215.01$</td>
<td>0.40</td>
<td>0.350354</td>
<td>4</td>
<td>Inv-Rev. Mono. (max: 111)</td>
</tr>
<tr>
<td>Proportion of woodland</td>
<td>$-6181.58$</td>
<td>33.83</td>
<td>0</td>
<td>4</td>
<td>Inv. Mono. (max: 27)</td>
</tr>
<tr>
<td>Distance from riparian forest</td>
<td>$-6176.53$</td>
<td>38.88</td>
<td>0</td>
<td>4</td>
<td>Inv. Mono. (max: 2489)</td>
</tr>
<tr>
<td>Roads</td>
<td>$-6154.44$</td>
<td>60.97</td>
<td>0</td>
<td>3</td>
<td>80</td>
</tr>
<tr>
<td>Distance from steep slope</td>
<td>$-6148.04$</td>
<td>67.37</td>
<td>0</td>
<td>4</td>
<td>Inv. Rev. Mono. (max: 2495)</td>
</tr>
<tr>
<td>Proportion of forest 1973</td>
<td>$-6132.13$</td>
<td>83.28</td>
<td>0</td>
<td>4</td>
<td>Inv. Mono. (max: 1155)</td>
</tr>
<tr>
<td>Euclidean distance</td>
<td>$-6131.45$</td>
<td>83.96</td>
<td>0</td>
<td>2</td>
<td>Inv. Mono. (max: 2354)</td>
</tr>
<tr>
<td>Elevation</td>
<td>$-6130.76$</td>
<td>84.66</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Note: The ‘Euclidean distance’ surface corresponds to a model built with Euclidean distance only (testing for correlation with geographic distance alone).

Resistance: optimized cost value for the landscape feature tested with categorical surfaces and the shape of the transformation for continuous surfaces as well as their maximum resistance value (Inv. Mono.: Inverse Monomolecular, Inv-rev. mono: Inverse-Reverse Monomolecular). See Figure S15 for transformation plots created for the continuous surfaces.

Abbreviations: AICc, AIC values adjusted for the number of populations sampled generated from the MLPE mixed effects mode; $\Delta$AICc, difference in the AICc values between the best supported model and each subsequent model; k, number of parameters; weight, weight of support for the corresponding surface averaged.
4 | DISCUSSION

Despite nearly seven decades of research into two chimpanzee communities – Kasekela in GNP and M-group in MMNP (Nakamura et al., 2015; Wilson et al., 2020) – we know relatively little about chimpanzee movement in Tanzania. Previous studies had used mtDNA or a limited number of size-based microsatellite loci to build hypotheses about barriers to regional chimpanzee movement. By generating both mtDNA- and MiSeq-based microsatellite genotypes for 136 GNP and 154 GME individuals collected across 19 sampling sites, we resolved previous contrasting findings and drew reliable conclusions on the gene flow and genetic diversity of these critically endangered populations.

We identified significant mtDNA and microsatellite structure between GNP and the GME. However, each marker illustrates different levels of structure. MtDNA haplotypes did not form geographically partitioned haplogroups, while microsatellite genotypes revealed two distinct clusters representing two populations likely currently isolated from one another. Our landscape genetic analyses confirmed a genetic discontinuity with a deviation from exact isolation by distance between the two genetic clusters. From the landscape features tested, rivers and proportion of bare habitat had the greatest explanatory power with the Malagarasi River and bare habitats highly restricting gene flow. Yet, we found evidence of high levels of gene flow and genetic diversity maintained within each cluster with similar microsatellite diversity and higher mtDNA diversity within the 36km² GNP than across the 18,000km² GME.

4.1 | Population genetic structure and gene flow

The finding of mtDNA haplotype-sharing suggests either recent or historical gene flow across Tanzania. In contrast, analyses of microsatellite genotypes revealed two distinct clusters corresponding to GNP and the GME, which indicates the two populations are likely currently isolated from one another. Nuclear microsatellite markers are more likely to reveal contemporary genetic patterns due to their higher evolutionary rates, while mtDNA is more useful for revealing historical events (Frankham et al., 2004). Given that at least five generations are needed for the genetic signatures of a landscape feature to be detectable (Westphal et al., 2021), it can be inferred that the two populations experienced restricted gene flow for at least 125 years (the generation time for chimpanzee being 25years (Langergraber et al., 2012)).

Evidence of widespread historical gene flow among eastern chimpanzee populations is well established with shared mtDNA haplotypes across the entire range of the subspecies (Goldberg & Ruvolo, 1997; Morin et al., 1994). However, regional structuring of mtDNA haplotypes was greater between GNP and the GME than between GNP and other eastern chimpanzee populations despite the latter being separated by far greater geographic distances (Langergraber et al., 2011). Such differentiation could be explained by the geographic barriers highlighted in our landscape genetic analyses. Across the 12 landscape features tested, rivers (more specifically, the Malagarasi) and proportion of bare habitats appeared as the primary drivers of chimpanzee’s genetic structure in Tanzania.

The impact of the Malagarasi River on chimpanzee gene flow was previously debated following evidence that chimpanzees could cross the river using natural, shallow fords (Piel et al., 2013). Although not an absolute barrier, the 50–200m wide river still limits gene flow and appears to be a primary driver of genetic structure. The biogeographical importance of the Malagarasi River has been noted by previous studies on other species. For example, olive (Papio anubis) and yellow baboons (P. cynocephalus) are parapatric with their distributions being delimited by this river (Kano, 1971; Zinner et al., 2015). Nonetheless, the proportion of bare habitat had almost equivalent explanatory power over the genetic structure observed. Chimpanzees feed principally on fruits (Wrangham, 1975), so it is not surprising that areas without trees are associated with limited chimpanzee movement. Together with the Malagarasi River, the 50km of open habitat separating GNP and the GME restricts gene flow and impacts the genetic structure of eastern chimpanzees to a greater extent than the 500km of mosaic habitats separating GNP and Uganda. The isolated effect of the Malagarasi River over the proportion of bare habitat cannot be precisely estimated because of the geographic structure of our samples. Additional samples directly above the Malagarasi River are required to determine with confidence which of these landscape features – if either – represents a more important barrier relative to the other.

Our results also suggest a high level of gene flow within the GME, which corroborates a previous proposal that the 18,000km² ecosystem hosts a single panmictic population of chimpanzee (Inoue et al., 2011). Although other studies had hypothesized potential barriers limiting chimpanzee movement between northern and southern parts of the GME (Moyer et al., 2006; Piel et al., 2013; Rudicell et al., 2011), our mtDNA and microsatellite results showed negligible north–south structuring within the ecosystem and no evidence of barriers to gene flow. Given the time lags for a new barrier to become detectable in the genetic structure, recent anthropogenic disturbance is unlikely to be observable in the current genetic structure but may appear in the future.

Chimpanzee behavioural divergence corroborates our results with the grooming hand clasp (i.e. two chimpanzees sitting face to face, clasping their hands above their heads and grooming one another with their other hand (McGrew, 2017)) observed in Issa and MMNP but not in GNP (Piel et al., 2017). The distribution of this behaviour within Tanzania supports a barrier to chimpanzee movement between GNP and the GME and suggests connectivity across the GME.

A high level of gene flow within the GME is also consistent with habitat connectivity modelling, suggesting the entire ecosystem is linked by a series of corridors with high probabilities of chimpanzee movement (Bonnin et al., 2020). However, this habitat connectivity model had less explanatory power for our genetic data than geographic distance alone. The connectivity model was built using chimpanzee occurrences skewed toward sleeping sites and may
have undervalued landscape features used for feeding and travelling (Bonnin et al., 2020). It may also be that dispersing individuals cross unsuitable habitats that are not used frequently for other daily activities (McCarthy et al., 2015).

Although SNPs have been shown to perform comparably to microsatellites in quantifying population divergence, they have greater power for resolving family structure within a population (Lemopoulos et al., 2019). Given the low genetic variation across the GME, future use of genomic markers and the inclusion of additional landscape surfaces may help us better understand how landscape features impact chimpanzee movement across the GME.

4.2 Genetic diversity and population viability

Our data provide no evidence of inbreeding at any sites and even suggest greater mtDNA genetic diversity within GNP than across the GME. This is very surprising given that GNP is 0.2% the size of GME and was thought to be isolated from other populations (Pintea et al., 2011; Pusey et al., 2007). Morin et al. (1994) also reported high within-community mtDNA genetic diversity in GNP and attributed it to females mediating gene flow through migration between communities. Inoue et al. (2011) reported genetic diversity in GNP to be greater than in five other sampled Tanzanian habitats. Evidence from Gombe suggests that female chimpanzees favour genetically dissimilar mates, which reduces the chances of inbreeding and optimizes genetic diversity (Walker et al., 2016). However, greater genetic diversity may also suggest connectivity between GNP chimpanzees and northern populations (e.g. Burundi). With chimpanzees now found ~15 km north of the park (Wilson et al., 2020) and females of unknown origin periodically arriving in GNP (Walker et al., 2016), it is very likely that GNP chimpanzees are receiving new alleles from extra-park populations. Furthermore, despite the park being primarily surrounded by human settlements and agriculture, chimpanzees exhibit high behavioural flexibility (Hockings et al., 2015) and have been observed in other parts of their range dispersing through and ranging across human-dominated landscapes (McCarthy et al., 2015; McLennan et al., 2021) as well as feeding on and making nests close to cultivated food resources (McCarthy et al., 2017; McLennan et al., 2020).

The lower mtDNA diversity observed within the GME may be due to its unusual geographic location, marking the southeastern limit of Pan distribution (IUCN, 2018). With Lake Tanganyika separating the GME from central populations and the Malagarasi River reducing gene flow from the north, new alleles coming from neighbouring populations are less likely to reach GME chimpanzees. It is also possible that due to the opportunistic nature of sampling within the GME, some genetic variability was missed. Contrary to GNP, for which we had access to genotype data for nearly all individuals, we estimate that we sampled less than 10% of chimpanzees present in the GME. This may have led to an underestimation of the genetic variability and the genetic structure across the ecosystem. Although our sample size is much higher than the minimum recommended (i.e. 30 individuals: Hale et al., 2012), more extensive sampling could provide additional support for the described pattern.

MtDNA diversity was lower in the GME, but nuclear diversity estimates were equivalent and mtDNA diversity indexes were in the range of other eastern chimpanzee populations (Table S5). Furthermore, we observed no sign of inbreeding across the 16 GME sample sites. One of our sample sites, Issa, was used to represent the GME in a recent study of chimpanzee connectivity across the species’ range (Lester et al., 2021). The authors found that Issa represented a consistent outlier, genetically divergent from all sites regardless of distance. They attributed this result to a loss of alleles due to random drift in extremely small populations. In contrast, our results provide evidence that Issa chimpanzees are not at all isolated and are genetically connected to a large population (~2600 chimpanzees within the GME). As Issa was the only representative of the GME in Lester et al. (2021), it is also feasible that the GME chimpanzee population, as a whole, represents an isolated population. Our findings corroborate those of Fontseré et al. (2022) who suggested historical connectivity across eastern chimpanzees and only recent isolation of Issa (thus the GME).

4.3 Biogeography and conservation of an isolated population at the edge of their distribution

GNP chimpanzees appear to be genetically closer to almost all other eastern chimpanzee populations than they are to GME chimpanzees to their immediate south, while the latter are closer to populations in the Democratic Republic of Congo (DRC) and northern Uganda (Table S4). This genetic structure is partially explained by the geographic barriers separating GNP and GME but may also reveal two origins of chimpanzees in Tanzania. GNP chimpanzee ancestors likely colonized the area by coming from central Africa to the north around Lake Tanganyika via Rwanda-Uganda-Burundi while GME chimpanzees could have reached Tanzania from the south of Lake Tanganyika (via Zambia). Although colonization may have involved some gene flow across the Malagarasi River (Piel et al., 2013), this long-standing geographic barrier would have minimized any direct gene flow and would explain the lower genetic distances between GNP and Lwazi compared to GNP and the GME (Inoue et al., 2011).

Reconstruction of historical chimpanzee distribution suggests that the species was found on both sides of Lake Tanganyika, up to Zambia (Barratt et al., 2021), although the current recognized range ends in DRC, north of the DRC-Zambia border. Supporting this hypothesis, Fontseré et al. (2022) found that Issa individuals carry more fragments of central chimpanzee ancestry than other eastern chimpanzees communities. Based on shared genetic drift, Kabogo chimpanzees – from south DRC – were also reported to form a clade with Issa and two other populations from Rwanda (figure S54 in Fontseré et al., 2022). Additional data, especially from chimpanzees inhabiting Lwazi and east DRC, are required to investigate more deeply the origin and the history of Tanzania’s chimpanzees.
Given the degree of isolation of the GME and GNP populations, new questions arise about how best to conserve chimpanzee connectivity and genetic diversity in Tanzania. Indeed, not only do they represent two distinct genetic clusters, but GNP and GME chimpanzees are also found across different habitats (van Leeuwen et al., 2020), have different body sizes (Uehara & Nishida, 1987) and exhibit divergent behaviours, e.g., ranging pattern and diet (Giuliano, 2022), vocalization (Mitani et al., 1992) and grooming culture (Piel et al., 2017). Chimpanzees in the GME may similarly present genetic adaptations to cope with seasonal fluctuations in resource availability. Selection on genes related to food and water scarcity has been found in other mammals (Finch et al., 2014; Orkin et al., 2021; Rocha et al., 2021) and may operate in this isolated population of chimpanzees. Evidence of outbreeding depression in mammals is scarce (Edmands, 2007; Frankham et al., 2004) but should not be neglected (Edmands, 2007). Reconnecting GNP and GME populations may increase genetic variability but decrease local adaptation. Current guidelines for population genetic rescue call for selecting populations that occur in similar habitats and have low population divergence to avoid reducing local adaptation and genetic incompatibilities (Hedrick & Fredrickson, 2010). The Tanzania National Chimpanzee Conservation Action Plan focuses on three key conservation targets, using corridors to preserve chimpanzee population connectivity, conserving chimpanzee populations and protecting core chimpanzee habitats (TAWIRI, 2018). With limited resources for conservation, conservation decision makers therefore have to balance resources across such key conservation targets. Considering the first target, efforts may be best allocated by restoring or strengthening connectivity between GNP and northern populations and preserving important habitats across the GME.

These chimpanzee populations demonstrate the inherent challenges to connectivity conservation. In an era of climate change and massive habitat loss, the need for managing large-scale connectivity has never been more urgent, but with limited resources, researchers generally face a trade-off among areas to focus conservation effort. A thorough understanding of both long-term and contemporary patterns of gene flow represents key information for successful connectivity management. With advances in sequencing technologies as well as population and landscape genetics methods, we are now able to generate reliable genotypes that can be compared across multiple studies and field sites allowing faster and more accurate understanding of gene flow and population genetic diversity.

ACKNOWLEDGEMENTS
We are grateful to TAWIRI, COSTECH and the Uvinza, Tangaanyka and Kigoma Districts for permission to conduct research in western Tanzania. The Nature Conservancy, Frankfurt Zoological Society and the Jane Goodall Institute with support from the United States Agency for International Development (USAID) and National Aeronautics and Space Administration (NASA) provided substantial financial, logistical, and administrative support. Thank you to the numerous field researchers who participated in the surveys conducted across the GME: Kiumbe Ahmedi, Matsumbuku Ahmedi, Mashaka Alimas, Moses Anyelwiye, Justina Boniface, Patrick Hassan, Busoti Juma, Hussen Juma, Shabani Kabangula, Mtunda Kapanda, Mustafa Kazimba, Gabriel Laizer, Kidosi Laurente, Joffrey and Shedrack Lucas, Makasa Mlekwa, Msigwa Rashid, Godfrey Stephano and numerous others for their invaluable assistance and expertise in the field, without whom this research would not have been accomplished. We are also extremely grateful to JGI field researchers who contributed to the long term data collection across GNP. We are also grateful to the UCSD/Salk Center for Academic Research and Training in Anthropogeny (CARTA) for long-term support of GMERC. Many thanks to Dr Adeline Morez, Dr Begoña Martínez-Cruz and Dr Nicolas Bech for helpful discussions on population and landscape genetic analyses. We further acknowledge Dr Jack Lester for his help with the EEMS analysis. The laboratory work was supported by funds from the National institutes of Health to B.H.H. (R01 AI050529, R37 AI150590 and P30 AI045008).

CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
Microsatellite genotypes, environmental layers and the r code used to perform the landscape genetic analyses are available on Dryad (https://datadryad.org/stash/share/L7dI0nN66Kmgt-9N7P3q3Mm_82EkI7w2DrKJN0qu0L). All newly generated Mitochondrial sequences (HVI) are accessible in GenBank under accession numbers OQ813753–0Q813755.

BENEFIT-SHARING STATEMENT
A research collaboration was developed with local scientists and NGOs to conduct research on a priority concern related to the connectivity of endangered great apes’ populations, with all collaborators being acknowledged as co-authors. The contributions of all individuals to the research, including local field assistants, are described in the acknowledgments, and a research report has been provided to the district government and the Tanzania Wildlife Research Institute to support conservation planning. All genetic data have been shared with the broader public via appropriate biological databases. Our group places great importance on engaging in international scientific collaborations and contributing to institutional capacity building efforts.


**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

---