Recent genetic connectivity and clinal variation in chimpanzees

Jack D. Lester1, Linda Vigilant1, Paolo Gratton1, Maureen S. McCarthy1, Christopher D. Barratt2, Paula Dieguez1, Anthony Agbor1, Paula Álvarez-Varona3, Samuel Angedakin1, Emmanuel Ayuk Ayimisin1, Emma Bailey1, Mattia Bessone1, Gregory Brazzola1, Rebecca Chancellor4, Heather Cohen1, Emmanuel Danquah5, Tobias Deschner1, Villard Ebot Egbe1, Manasseh Eno-Nku6, Annemarie Goedmakers7, Anne-Céline Granjon1, Josephine Head1, Daniela Hedwig8, R. Adriana Hernandez-Aguilar3,9, Kathryn J. Jeffery10, Sorrel Jones1, Jessica Junker1, Parag Kadam11, Michael Kaiser1, Ammie K. Kalan1, Laura Kehoe12, Ivonne Kienast1, Kevin E. Langergraber13, Juan Lapuente1,14, Anne Laudisoit15,16, Kevin Lee1, Sergio Marrocoli1, Vianet Mihindou17,18, David Morgan19, Geoffrey Muhanguzi20, Emily Neil1, Sonia Nicholl1, Christopher Orbell21, Lucy Jayne Ormsby1, Liliana Pacheco3, Alex Piel22, Martha M. Robbins1, Aaron Rundus23, Crickette Sanz24,25, Lilah Sciaky1, Alhaji M. Siaka26, Veronika Städele1, Fiona Stewart27, Nikki Tagg28, Els Ton7, Joost van Schijndel7, Magloire Kambale Vyalengerera1, Erin G. Wessling29, Jacob Willie28, Roman M. Wittig1,30, Yisa Ginath Yuh1, Kyle Yurkiw1,31, Klaus Zuberbuehler20,32,33, Christophe Boesch1, Hjalmar S. Kühl1,2 & Mimi Arandjelovic1

Much like humans, chimpanzees occupy diverse habitats and exhibit extensive behavioural variability. However, chimpanzees are recognized as a discontinuous species, with four subspecies separated by historical geographic barriers. Nevertheless, their range-wide degree of genetic connectivity remains poorly resolved, mainly due to sampling limitations. By analyzing a geographically comprehensive sample set amplified at microsatellite markers that inform recent population history, we found that isolation by distance explains most of the range-wide genetic structure of chimpanzees. Furthermore, we did not identify spatial discontinuities corresponding with the recognized subspecies, suggesting that some of the subspecies-delineating geographic barriers were recently permeable to gene flow. Substantial range-wide genetic connectivity is consistent with the hypothesis that behavioural flexibility is a salient driver of chimpanzee responses to changing environmental conditions. Finally, our observation of strong local differentiation associated with recent anthropogenic pressures portends future loss of critical genetic diversity if habitat fragmentation and population isolation continue unabated.
Humans have been characterized as a genetically continuous species, which is expected to hamper local adaptation. Our species has largely relied on behavioural flexibility to become among the most widely distributed species, inhabiting a diverse range of climates and habitats. Among our closest living relatives, the chimpanzee (Pan troglodytes), also exhibits extensive behavioural variation, both at a local and regional scale, and occupies a broad range of habitats and climates, while displaying little associated morphological variation. However, to date, their pattern of genetic diversity is equivocal: some studies provide evidence of connectivity among all populations, while others have concluded that chimpanzees are taxonomically divided into four geographical subspecies. Given the high degree of behavioural variability and flexibility across chimpanzee populations, characterizing range-wide patterns of genetic diversity in chimpanzees is important for understanding how they adapt to changing environmental conditions. In particular, a signal of genetic connectivity across their range would suggest that, like in humans, local trait fixation allows them to quickly and dynamically respond to ecological challenges.

Prior to the recent growth and expansion of human agriculturalist settlements across Africa (ca. 5–2 ka BP), chimpanzees were nearly continuously distributed across Equatorial Africa. Rivers appear to be the main barriers to migration, separating chimpanzees and bonobos (Pan paniscus), as well as three of the four currently recognized chimpanzee subspecies: Nigeria–Cameroon (Pan troglodytes ellioti), central (Pan troglodytes troglodytes) and eastern (Pan troglodytes schweinfurthii) chimpanzees (Fig. 1). River systems, however, are dynamic and may become permeable to dispersal during arid periods, or when natural bridges form. Moreover, the presence of this type of barrier does not preclude the possibility of gene flow occurring around it, which has not been tested on the subspecies-delineating barriers in any previous studies of chimpanzees. Similarly, the Dahomey Gap, an arid, 200-km-wide forest-savannah mosaic that separates western chimpanzees (Pan troglodytes verus) and P. t. ellioti, is also dynamic, as this region hosted rainforest as recently as 4 ka BP (ref. 21; ~160 chimpanzee generations).

In a continuously distributed population, limited dispersal drives a predictable pattern of genetic diversity, known as isolation by distance (IBD), which manifests as a continuous gradient (cline) of declining similarity as geographical distance increases. Detectable departures from this pattern are evidence of geographic or behavioural barriers that reduce or impede dispersal, thereby leading to increased geographical rates of genetic differentiation. Although chimpanzees are thought to have had a historically continuous geographic distribution across their range, studies of chimpanzee genetic diversity have tended to show discontinuities consistent with the subspecies classifications or some degree of population structuring. To detect geographic clustering of genetic data, these studies relied on several different approaches not accounting for the spatial distribution of samples. These include Bayesian clustering algorithms (e.g., STRUCTURE), the results of which can be biased when IBD is present in the data and principal component analyses, a method specifically intended to maximize between-group differences. However, to date no studies of species-wide...
The chimpanzee population structure have featured spatially explicit methods, which incorporate the geographic location of the genetic samples into the analysis model and assume IBD in the null model37,38. Importantly, when using spatially agnostic methods in the presence of IBD, unbalanced sampling will detect discrete stratification of the data, regardless of the actual pattern29,30,33–36, and this cannot be compensated for by analyzing more loci (i.e., a large number of SNPs)37. Mostly due to the logistical challenge of obtaining non-invasive samples from chimpanzees’ wide range, which largely encompasses politically unstable or remote regions, previous studies of chimpanzee population structure have relied on small and clustered datasets, and often made use of zoo and sanctuary samples of unknown provenance, which introduced further uncertainty11–15. Therefore, previous reports of chimpanzee population structure were likely biased by applying spatially agnostic analyses to datasets comprised of dispersed and uneven sampling in the presence of IBD37,38, and were possibly further confounded by incorrect knowledge about populations of origin in some samples. In fact, even the most comprehensive studies of chimpanzee population history, which utilized genome-wide sequence data4,15, were based on a relatively small number of samples from captive individuals, and focused more on fitting models of divergence among subspecies than investigating recent genetic connectivity in space.

In this study, we aimed at assessing recent genetic connectivity across the chimpanzees range by analyzing an extensive sample providing an unprecedented level of geographic coverage. As part of the Pan African Programme: The Cultured Chimpanzee (PanAf)38, over an 8-year period, we non-invasively collected and genotyped >5000 wild chimpanzee faecal samples from 55 localities in 18 countries across the entire species range. We scored individual genotypes (allele length polymorphism) at up to 14 microsatellite loci, as these are cost effective and allow for accurate genotyping of non-invasively collected samples39. Previous studies in other species (fruit flies40, fish41,42, birds43, amphibians44, wild boars34, felids45 and beetles46) have demonstrated that 6–14 microsatellite loci are informative enough to even detect subtle population structure. Furthermore, a previous study of how sampling affects detection of chimpanzee population structure found that eleven loci were sufficient to overcome false signals of population structure caused by sampling bias and uncover a signal of IBD9. As microsatellite markers evolve rapidly and are expected to be selectively neutral, they are highly sensitive to patterns of gene flow, especially within shallow evolutionary timescales. For our analyses, we modeled our georeferenced genetic data in space by employing spatially explicit approaches to inform the distribution, thereby avoiding a priori definitions of the population structure or assumptions of homogeneous sampling, thus minimizing biases associated with previous studies.

Results

The PanAf team and collaborators collected 5397 chimpanzee faecal samples from 55 temporary or long-term research sites spanning the geographic range of chimpanzees from 2010 to 2018 (Fig. 1). After DNA extraction and microsatellite amplification, 2497 extracts were successfully typed at up to 14 loci (range: 7–14, mean = 11.3, Supplementary Tables 1.1–1.4) and represent 939 unique individuals from 48 sampling locations (see Supplementary Note 1: ‘Genotype reconstruction’). We found that including individuals typed at seven loci (~8% of all samples) in all analyses did not affect the observed patterns of genetic differentiation compared to limiting the minimum to eight or nine loci, and this inclusion allowed for the addition of several key sampling locations in comparisons of population diversity levels. Reducing the minimum to six loci (n = 69) noticeably reduced differentiation across analyses.

Influence of geographic distance on genetic distance. In an idealized model with uniform population density and dispersal, genetic differentiation among local demes is expected to increase with geographic distance (D). Roux57 showed that, in a linear one-dimensional (1d) space, $F_{ST}/(1 – F_{ST})$ ~ D and $F_{ST}/(1 – F_{ST})$ ~ log(D) in an infinite two-dimensional (2d) space, when mutation is negligible. In principle, one may statistically fit these equations to actual data and interpret deviations from fits as signals of non-uniform spatial patterns of density and gene flow.

Actual geographical spaces are neither 1d nor infinite 2d spaces, and populations inhabiting an elongated 2d space are expected to show relationships of $F_{ST}$ and D somewhat in between linear and logarithmic37. In addition, mutation is often not negligible. Microsatellite DNA markers, in particular, have a typically fast mutation rate, and the number of repeats at a given microsatellite marker tends to change according to a stepwise mutation model (SMM)48, which leads to an overestimation of diversity. While a strict SMM lends itself to unbiased estimation of demographic parameters from microsatellite variation (e.g., Goldstein49), it rarely accurately describes the actual mutation process at microsatellite loci. Constraints to the maximum and minimum size number of repeats at a given marker50 are especially important, as they determine an upper limit to the genetic differentiation among pairs of populations. With these limitations in mind, modeling genetic differentiation as a function of geographical distance is still a powerful tool to identify discontinuities in the spatial structure of biological populations.

In a global, unstratified test of our data (G ~ D) in which subspecies were not defined, we found that D explained >30% (adjusted r² = 0.314) of G, and as much as 58% (adjusted r² = 0.579), when three obvious outlier sites were removed (Mt. Sangbé (Côte d’Ivoire), Gashaka (Nigeria) and Issa (Tanzania)). These sites were consistent outliers in all comparisons and analyses (Supplementary Fig. 3b and Supplementary Fig. 4a, b). A log-transformed model (G ~ log(D)) gave very similar results (adjusted r² = 0.320 and 0.579 for the global model and for the model excluding outliers, respectively), as it could be expected given that chimpanzee range’s shape is fairly elongated and somewhat intermediate between a 1d and a 2d space.

Full-range data were then stratified according to whether the two sites in each pair belonged to the same or different subspecies (Fig. 2b). However, further analyses in which each subspecies pairs were separately considered (e.g., P. t. verus–P. t. verus, P. t. verus–P. t. troglodytes, etc., Fig. 2c, d) revealed that this pattern was largely driven by the overrepresentation of site pairs within P. t. verus, which is known to have the lowest diversity of all subspecies15,33, resulting in a very low intercept. Fitting G ~ D regressions within each subspecies pair, instead, did not reveal clear evidence of abrupt discontinuities across the chimpanzees range (Fig. 2c, d), but rather highlighted ‘locally’ different patterns. In particular, P. t. verus populations displayed slow slope and intercept, which, combined with their low genetic diversity, is consistent with a relatively recent and fast demographic and spatial expansion, while P. t. troglodytes had low intercept but a higher slope (consistent with a more stable pattern of IBD) and there was no clear IBD pattern within P. t. schweinfurthii (where G tends to be high, irrespective of geographic distance, which can be expected given the rugged landscape in the great rift area, where three sampled sites are located). Notably, between-subspecies regressions involving P. t. verus (the most geographically isolated and genetically divergent subspecies35) were essentially flat, suggestive of saturated genetic differentiation at
outliers, while the *P. t. troglodytes*–*P. t. schweinfurthii* regression was highly similar to within-subspecies regressions, with coefficients in-between *P. t. troglodytes* and *P. t. schweinfurthii* in the linear regression and almost identical to within *P. t. troglodytes* in the log-transformed regression, Fig. 2c, d.

**Historical estimated effective migration rates.** To visualize chimpanzee migration patterns and test for spatial genetic structure, we employed the spatially explicit analysis implemented in the Estimated Effective Migration Surfaces (EEMS) software\(^1\). This analysis combines spatial and genetic data and relies on MCMC to estimate relative effective migration rates ($m$; between-deme genetic dissimilarity measured against geographic distance) modeled on a 2D surface, using a mean-centred relative scale. In this model, a ‘flat’ uniform migration surface would indicate perfect IBD across the considered range, and non-uniformity of $m$ in space is interpreted as increased or decreased genetic connectivity. Besides plotting estimated migration surfaces, EEMS results differentiate areas where the probability of $m$ differing from the mean rate is statistically significant (i.e., local $m >$ mean $m$ in >95% of the MCMC samples). Importantly, the EEMS model assumes that microsatellite markers mutate according to a strict SMM and violations of this assumption may lead to an overestimation of effective migration rates, while homoplasy caused by allelic size saturation may lead to an underestimation. Despite having some deviations from strict SMM and the presence of homoplasy that are apparent in our data, we found the EEMS model to be a good fit to our chimpanzee dataset (adjusted $r^2 = 0.381$, Supplementary Fig. 7a).

However, the observed genetic dissimilarity tended to be less than predicted by the model for very large distances, i.e., mostly involving *P. t. verus* vs. *P. t. schweinfurthii* pairs, and in particular, *P. t. verus*–Issa comparisons (Supplementary Fig. 7b), similar to our observations in the regression analyses.

Although we observed spatial variation in $m$, we did not detect any significant signal of genetic discontinuity, or ‘barriers’ between-subspecies populations (Fig. 3a). Consistent with the divergent results from the $G ~ D$ regression models (Fig. 2a–d), $m$ was generally higher in the *P. t. verus* range than in the rest of the species range (*P. t. elliottii*, *P. t. troglodytes* and *P. t. schweinfurthii*, hereafter, collectively, *ETS*) and two patches of significantly high $m$ were identified in the most densely sampled areas of the *P. t. verus* range. In contrast, significantly low migration was inferred.
for the region surrounding the Gashaka sampling location within the range of *P. t. elliotti* and in the mountainous regions of southern Uganda and Rwanda centered around Gishwati in the *P. t. schweinfurthii* range.

Since EEMS estimates relative *m* and the considerable contrast between *P. t. verus* and ETS may subdue more local patterns, we also performed separated analyses for *P. t. verus* and for the remaining *ETS* populations (Fig. 3b, c). The *P. t. verus* analysis (Fig. 3b) highlighted significantly low *m* associated with the outlier location of Mt. Sangbé, while the ETS analysis confirmed significantly lower gene flow in the mountainous region in the southeastern portion of the *P. t. schweinfurthii* range relative to the average *ETS* migration landscape. To determine whether local effects drove these effective barriers, we removed Mbe, Mt. Sangbé and Gishwati, and re-analyzed the data; all of the barriers separating the subspecies’ ranges were not detected. Effective migration rates within much of *Pan troglodytes verus* were significantly higher than average. b Historical EEMS map of *P. t. verus*. There was a significant barrier associated with Mt. Sangbé in Côte d’Ivoire. When we removed Mt. Sangbé the barrier was no longer significant, but it was still present, suggesting the possibility of reduced historical gene flow across Côte d’Ivoire. c Historical EEMS map of *P. t. elliotti*, *P. t. troglodytes* and *P. t. schweinfurthii* (collectively ETS). As in a, removing Gishwati resulted in the barrier no longer being significant (Extended data Fig. 6c). Rates of *m* between the panels are relative and not directly comparable.

**Diversity rates.** We visualized spatial patterns of chimpanzee genetic diversity by plotting estimated relative diversity rates (*q*; pairwise within-deme genetic dissimilarity between individuals) from EEMS. At full scale, *q* in *P. t. verus* populations were significantly lower than species average (Fig. 4a). We calculated mean *q* from the EEMS output for both *P. t. verus* and ETS, and found that ETS was 1.7-fold higher, in agreement with previously published reports. When analyzed alone (Fig. 4b), the overall distribution of diversity in *P. t. verus* was generally homogeneous, with the exception of very local significant differences in parts of Liberia (Sapo and Grebo) and Côte d’Ivoire (Djouroutou). A focused analysis of *ETS* (Fig. 4c) revealed that diversity rates tended to be higher in populations closest to the equator, and
areas where rates diverged significantly were generally localized
between two neighbouring populations. Interestingly, although m
was low in the southeastern ETS range, diversity appeared to be
relatively high. All of the populations sampled within the P. t.
elioti range had low relative q, with two sites being significantly
lower than the average ETS. Two sampling locations in western
Uganda (Budongo and Ngogo) in the P. t. schweinfurthii range
also displayed significantly lower q.

Discussion
Our results show an essentially clinal pattern of genetic variation,
largely predicted by geographic distance (IBD), with strong local
effects driving differentiation in a few highly isolated populations.
Importantly, we did not find evidence of major discontinuities
conforming to the current taxonomic subdivision. However,
although effective barriers were not evident in our data, the P. t.
verus population appears to have had a divergent history, as they
display much lower genetic diversity in tests of same-subspecies
among-site differentiation, while also sharing widespread higher
effective migration rates across the subspecies in the whole-
species EEMS analysis. Importantly, our observations of high
effective migration rates in P. t. verus may simply be a con-
sequence of the overall paucity of diversity in this subspecies. The
EEMS model utilizes variation in genetic dissimilarity against
geographic distance to estimate effective migration, therefore,
high levels of genetic similarity (low diversity) in a geographically
clustered subpopulation will appear to have a relatively high
migration rate relative to the rest of the sampled populations.
Therefore, our observation of relatively high effective migration
throughout P. t. verus may indicate higher levels of connectivity
in western chimpanzees, but may alternatively or additionally be
a signal of a relatively recent (likely northward) population
expansion. Indeed, much of the P. t. verus range is savannah
mosaic habitat with discontinuous gallery forests, where
chimpanzees tend to have both (1) larger home ranges (over 63
km²) than forest-dwelling populations, to seemingly compensate
for lower fruit tree density and suitable habitat availability,
and (2) a concomitant increase in dispersal distances, thereby
possibly driving a signal of increased migration. However, effec-
tive migration was uniform across most of the P. t. verus range, as
similar rates were also detected among the forest-dwelling
populations in southern Liberia and western Côte d’Ivoire,
where home ranges have been reported to be between 6 and 37
km² in size, suggesting that dynamics other than ecology are
involved. This and, especially, the low diversity in P. t. verus,
suggest that the high effective migration we observed in this
subspecies should be interpreted as evidence of a population
bottleneck followed by a recent and fast range expansion. Indeed,
this interpretation is corroborated by among-site comparisons of
mean standard deviation of allele sizes (SD; large allele-size var-
iation) and Garza Williamson’s M (Supplementary Figs. 5a–d
and 6a–d, and see Supplementary Note 2: ‘Detection of a population
bottleneck in P. t. verus’ for detailed explanation), whereby the
low SD:M ratio observed in P. t. verus suggests a loss of alleles and
allele-size ranges, which can be expected for populations that have maintained a relatively small population size for a substantial number of generations (Supplementary Fig. 6a–d). A similar pattern was also observed in most P. t. ellioti samples and in the population in the Budongo/Ngogo region of Uganda, while Gashaka has both low SD and moderate M, indicating small long-term population size, possibly, a recent bottleneck. Conversely, populations that maintain a comparatively high SD, but display a low value of M in the central and eastern subspecies (e.g., Gishwati and Issa), are suggested to have experienced a more recent bottleneck (Supplementary Fig. 6a–d).

Our estimates of effective migration rates did not detect significant discontinuities corresponding with the Dahomey Gap, the Sanaga or Ubangi Rivers, the three proposed subspecies-delimiting barriers. Although our sampling scheme endeavoured to collect data from as many localities as was practicable, unavoidable sampling gaps occurred in the P. t. ellioti range, between P. t. ellioti and P. t. troglodytes, and in northwestern DRC between P. t. troglodytes and P. t. schweinfurthii. However, these gaps were of moderate extension and spatial homogeneity in sampling is not an assumption of either the EEMS analysis or our IBD regression, so that it is extremely unlikely that sampling gaps represent a main influence on our results. Our analyses, therefore, indicate that previous observations of recent genetic discontinuities in chimpanzees likely resulted, at the minimum, from the use of spatially agnostic methods that were biased by geographically clustered genetic samples. Notably, these results do not at all suggest an overall lack of genetic differentiation among chimpanzee populations set apart by geography. Though we found that genetic variation follows a generally clinal pattern across the species, we want to stress that we were specifically focused on patterns occurring on a recent timescale, i.e., within the current interglacial period. Populations are spatially and temporally dynamic, and it is likely that the range of suitable habitat for chimpanzees has fluctuated with the glacial–interglacial cycles, causing repeated population contractions and subsequent expansions. It is important to remark that our analyses utilized microsatellite markers, which, mainly due to both their high mutation rates and polymorphism, can be very effective at revealing subtle genetic structure. However, fast mutation rate, coupled with constraints on maximum and minimum allele size, restrict the information that microsatellites may provide about deeper demographic events, leading to a shallow effective timescale. In particular, a deep divergence between populations, e.g., chimpanzee subspecies, may appear relatively shallow if there are heavy constraints on allele size (which causes differentiation to become asymptotic). However, if previously isolated populations came into contact in the more recent past, demes that are closer to the interface will exchange more migrants, and thus more alleles. The admixed alleles would diffuse along a gradient, since they would become more dispersed as distance from the interface increases. In such cases, recent gene flow and the relative underestimation of deep divergences by microsatellites may, in principle, create a smooth pattern of IBD, whereas other markers might otherwise reveal steeper gradients. This makes microsatellites highly effective for assessing recent patterns of population structure, as signals of ancient contraction–expansion events are flattened by allele size constraints and relatively high mutation rates. Finally, we wish to stress that gradients are not expected to occur by chance alone. Therefore, observations of IBD in our data represent, at least, clear evidence of recent connectivity among populations.

The chimpanzee populations that appear considerably more differentiated than expected by IBD alone, e.g., Mt. Sangbé, Gashaka and Issa, are known to have experienced isolation and decline in recent generations as a consequence of local anthropogenic pressures. Therefore, their observed differentiation most likely stems from loss of alleles due to random drift in extremely small populations.

Our data suggest that, much like humans, chimpanzees likely experienced periods of substantial genetic connectivity across a geographic range encompassing a broad range of habitats, a finding that may point to important evolutionary implications. Environmental heterogeneity and population size act synergistically to influence rates of adaptive evolution. Advantageous mutations are more likely to become fixed in small populations occupying uniform habitats, and are less likely to become fixed in a large population occupying a diverse habitat. Therefore, as humans and chimpanzees are both widely distributed across a variety of habitat types, evolutionary adaptation is expected to be slow in these species, especially given their long life histories. Hominins have long employed behavioural flexibility to mitigate biological limitations, exploit new resources, inhabit challenging environments (i.e., use of fire) and migrate out of Africa on several occasions before modern times. Since our results have demonstrated that recent genetic variation in chimpanzees contains large geographic structure on a broad scale, behavioural flexibility may also play an important evolutionary role in their adaptability to environmental heterogeneity. Chimpanzees have diverse and variable behavioural repertoires that often vary among nearby communities, they utilize learned techniques to harvest otherwise unobtainable food, and their behavioural diversity increases with environmental variability. Indeed, chimpanzees living in arid habitats employ specialized behaviours to regulate the thermal exposure and spontaneously innovate novel behaviours in response to increased environmental complexity arising from anthropogenic pressure.

Although our data point to diffuse genetic connectivity in chimpanzees, this signal originates from recent historical patterns in previous generations and does not represent the current potential for gene flow. Indeed, we found that several local chimpanzee populations are very strongly isolated from a genetic perspective, and many more are known to have become fragmented and/or have undergone precipitous declines in recent decades due to anthropogenic factors, such as habitat fragmentation, hunting and disease transmission. P. t. schweinfurthii, P. t. troglodytes and P. t. ellioti are recognized as endangered, while P. t. verus is recognized as critically endangered. Though our results indicate a high level of recent genetic connectivity in chimpanzees, we do not suggest that the distinction among the subspecies populations is irrelevant. Each subspecies range hosts populations that contribute unique behaviours to the species as a whole, and our findings suggest that genetic diversity in chimpanzees is distributed throughout the taxon, underscoring the need to maintain regionally driven conservation approaches. However, these results also highlight the need to preserve and restore corridors to facilitate connectivity among remaining populations to avoid inbreeding depression and the accumulation of deleterious alleles that negatively affect fitness. All populations possess unique traits that cumulatively reflect the full genetic and behavioural diversity observed in chimpanzees. Continued isolation and further population decline will irrevocably erode the overall viability of this critical keystone species, ultimately imperiling their long-term fitness. If connectivity between populations is not restored or deteriorates, the geographic distribution we observe in chimpanzees today will shape the genetic structure of future generations, which may have negative evolutionary consequences, and ultimately jeopardizes the long-term survival of wild populations.

Methods
Sample collection and preparation. As part of the PanAf, 5397 chimpanzee faecal samples were non-invasively collected at 55 temporary or long-term research sites
across the species range (Fig. 1, Supplementary Table 1.3 and Supplementary Fig. 3c), preserved according to the two-step ethanol–silica method57 and stored in the field for up to 2 years. Upon arrival to the lab, samples were stored at −20 °C. DNA extracts were isolated either manually, using the QIAamp Stool Kit (Qiagen), or using an automated process employing the QIAamp 96 PowerFecal QIAcube HT robot (Qiagen), per manufacturer instructions, modified by incorporating a pre-treatment step to improve the DNA quality and yield (Supplementary Note 1: ‘Laboratory methods’). To determine loci (amelogenin) were amplified using a two-step multiplex process37 with slight modifications (Supplementary Note 1: ‘Laboratory methods’ and Supplementary Fig. 1). PCR products were analyzed using an ABI Prism 3730 genetic analyzer (Applied Fisher Scientific). Loci were characterized relative to ROX labeled HD400 internal size standard using Genemapper version 5.0 (Thermo Fisher Scientific). Homozygotes were identified by three identical PCR replicates and heterozygotes were confirmed by at least two unambiguous PCR replicates of each allelic79 (Supplementary Note 1: ‘Genotype reconstruction’).

Analyses. Cervus 3.0.7 (ref. 78) was used to calculate allelic frequencies, determine the minimum number of loci needed to discriminate individuals (PDMa)79, and to assess the degree of relatedness of individuals within sites (Supplementary Note 1: ‘Genotype reconstruction’ and Supplementary Table 5). Based on the allelic frequency data, we determined that we needed eight loci to be 99.9 percent certain that two identical genotypes originated from the same individual rather than full siblings79. In cases where genotypes were unique (not matching to any other genotype), individuals were typed at a minimum of seven loci.

We employed two strategies to assess genetic continuity at the species, subspecies and site levels: (1) we examined the relationship between genetic and geographic distances by applying simple linear regression functions to multiple levels of our data, and (2) we used a spatially explicit analysis that identifies significant spatial deviations from clinal variation (genetic discontinuity). We also performed Mantel, partial Mantel and STRUCTURE analyses on the dataset, as they are standard practice in population structure studies; however, reliable inferences could not be drawn as results were severely confounded by biases in the data. Instead, we provide in-depth details about the modeling and results of these analyses in Supplementary Note 2: ‘Cluster (STRUCTURE) analysis and stratified and partial Mantel tests’ (Supplementary Fig. 2a–c and Supplementary Table 2.1).

We utilized linear regression analyses (G − D) to assess the relationship between genetic distance (Fst)56 and geographic distance (least cost paths; LCPs) at the species, between and within-subspecies and subspecies-comparisons levels. For our analyses, we calculated Fst as a suitable estimator of Fst, which adjusts for differing levels of genetic diversity arising from variation in effective population sizes (Ne) among sampling locations and mutation rate among loci, while not being dependent on within-site diversity56. Fst (hereafter G1 − Fst) was modeled as a function of LCP estimators of D, which calculate pairwise distance between sampling localities based on the presence of suitable chimpanzee habitat spanning back 120 ky BP79, and include topological features that could cause resistance to gene flow. We used LCPs specifically to force distance measurements around the presence of large water bodies, such as the Atlantic Ocean, where chimpanzee migration is not possible (Supplementary Fig. 3a, c).

To test for effective barriers to migration and to plot the distribution of chimpanzee genetic diversity across their range, we used the spatially explicit EEMS program11. EEMS estimates historical migration rates that would have given rise to the observed spatial distribution of genetic diversity in the data under idealized conditions, i.e., random mating, neutral selection and constant population size. This method uses georeferenced genetic data to assess diversity in a spatial context; therefore, in contrast to Bayesian clustering algorithms, it does not require sampling homogeneity, ideal for our unbalanced, heterogeneous dataset.

Furthermore, the EEMS analysis is not biased by the presence of IBD in the data, rather, decreasing genetic similarity as a function of distance is the null assumption. Furthermore, the EEMS analysis is not biased by the presence of IBD in the data, therefore, in contrast to Bayesian clustering algorithms, it does not require such random sampling. The historical range of chimpanzees has likely expanded and contracted over time11, so we accordingly extended the current northern boundaries of the species range by 300 km in our EEMS analyses. We also included the Dahomey Gap as a functional part of the chimpanzee range, as it has fluctuated in size and featured rainforest as recently as 4 ka BP, a period within the timescale of this study21. For grid density, we performed iterative analyses to ascertain the optimal number of demes that provided the best model fit for analyses at the species (800 demes), P. t. versus (300 demes) and ETS (490 demes) population levels21. For all analyses, to ensure the models reached stability, three independent MCMC chains were run preceded by 500,000 steps of burn-in, at lengths of 6,000,000 repetitions, each beginning with different random seeds.

Statistics and reproducibility. We performed all analyses from a pool of 939 individual genotypes originating from 48 sampling locations with a minimum inter-site distance of 41 km (mean pairwise distance 2843 km; estimated from LCPs). When necessary, we balanced our dataset by imposing a lower limit of 6 and an upper limit of 20 (using a randomized drawing process) on the number of genotypes originating from the same sampling location (see Supplementary Note 2: ‘isolation by distance (IBD) and distance estimators’). For the EEMS analyses, since balancing was unnecessary for samples for which geospatial data were recorded (932) were used in the species-wide scale and were divided for focused analyses of P. t. versus (522) and ETS (410). Detailed descriptions of all analyses and p values are provided in the text and Supplementary Note 2: ‘Population genetics analyses’.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The data from the present study are available from the corresponding authors upon reasonable request.

Received: 25 June 2020; Accepted: 4 February 2021;
Published online: 05 March 2021

References

1. Serre, D. & Paabo, S. Evidence for gradients of human genetic diversity within and among continents. Genome Res. 14, 1679–1685 (2004).
2. Ohta, T. Population size and rate of evolution. J. Mol. Evol. 1, 305–314 (1972).
3. Slotkin, M. Gene flow and selection in a clinic. Genetics 75, 733–756 (1973).
4. Potts, R. Variability selection in hominid evolution. Evol. Anthropol. 7, 81–96 (1998).
5. Potts, R. Hominin evolution in settings of strong environmental variability. Quat. Sci. Rev. 73, 1–13 (2013).
6. Prüfer, K. et al. The bonobo genome compared with the chimpanzee and human genomes. Nature 486, 527–531 (2012).
7. Hill, W. C. O. in The Nomenclature, Taxonomy and Distribution of Chimpanzees, Vol. 1 (ed. Bourne, G. H.) 22–49 (Karger, 1969).
8. Pruetz, J. D. & Bertolani, P. Chimpanzee (Pan troglodytes versus) behavioral responses to stresses associated with living in a savannah-mosaic environment: implications for hominin adaptations to open habitats. Paleoanthropology 2009, 252–262 (2009).
9. Funästuck, T. et al. The sampling scheme matters: Pan troglodytes troglodytes and P. t. schweinfurthii are characterized by clinal genetic variation rather than a strong subspecies break. Am. J. Phys. Anthropol. 156, 181–191 (2014).
10. Fischer, A., Pollack, J., Thalmann, Ö., Nickel, B. & Paabo, S. Demographic history and genetic differentiation in apes. Curr. Biol. 16, 1133–1138 (2006).
11. Gonder, M. K. et al. A new west African chimpanzee subspecies? Nature 388, 337 (1997).
12. Beaugé, C., Patterson, N., Stone, A. C., Przeworski, M. & Reich, D. E. Genetic structure of chimpanzee populations. PLoS Genet. 3, e66 (2007).
13. Bowden, R. et al. Genomic tools for evolution and conservation in the chimpanzee: Pan troglodytes ellioti is a genetically distinct population. PLoS Genet. 8, e1002504 (2012).
14. Prado-Martinez, J. et al. Great ape genetic diversity and population history. Nature 499, 471–475 (2013).
Acknowledgements

We wish to thank the Max Planck Society, Max Planck Society Innovation Fund and the Heinz L. Kerekder Foundation for providing funding to the PanAf project.

We also would like to thank the following authorities who granted permission to conduct research: Agence Nationale des Parcs Nationaux, Gabon, Centre National de la Recherche Scientifique (CNEREST), Gabon, Conservation Society of Mbe Mountains (CAMM), Nigeria, Department of Wildlife and Range Management, Ghana, Direction des Eaux, Forêts et Chasses, Senegal, Eaux et Forêts, Mali, Forestry Commission, Ghana, Forestry Development Authority, Liberia, Institut Congolais pour la Conservation de la Nature, DR-Congo, Instituto della Biodiversità e delle Aree Proteggiate (IBAP), Makerere University Biological Field Station (MUBFS), Uganda, Ministere de l’Economie Forestiere, R-Congo, Ministere de la Recherche Scientifique et de l’Innovation, Cameroon, Ministere de la Recherche Scientifique, DR-Congo, Ministère de l’Economie Forestière du gouvernement de la République du Congo, Institut National de Recherche Forestière, Agence Congolaise de la Faune et des Aires ProtégéesMinistère de l’Agriculture de l’Élevage et des Eaux et Forêts, Guinea, Ministère de la Recherche Scientifique et Technologique, R-Congo, Ministère des Eaux et Forêts et Ministère de l’Enseignement Supérieur et de la Recherche Scientifique, Cote d’Ivoire, Ministère des Forêts et de la Faune, Cameroon, Ministre de l’Environnement et de l’Assainissement et du Développement Durable du Mali, Ministre de l’Agriculture et du Développement Rural, Guinea-Bissau, Ministry of Agriculture, Forestry and Food Security, Sierra Leone, Ministry of Education, Rwanda, National Forestry Authority, Uganda, National Park Service, Nigeria, National Protected area Authority, Sierra Leone, Rwanda Development Board, Rwanda, Société Equatoriale d’Exploitation Forestière (SEEF), Gabon, Tanzania Commission for Science and Technology, Tanzania, Tanzania Wildlife Research Institute, Tanzania, Uganda National Council for Science and Technology (UNCST), Uganda, Uganda Wildlife Authority, Uganda and Ministere de l’Enseignement Supérieur et de la Recherche Scientifique in Côte d’Ivoire.

We thank the following people who were instrumental in providing help with the PanAf project: for analytical guidance, we would like to thank Tomas Marques-Bonet, Aida Andrès Moran, Benjamin Peter, Joshua Schmidt and Lauren C. White. For Tanzania, Uganda National Council for Science and Technology (UNCST), Uganda, Rwanda, National Forestry Authority, Uganda, National Park Service, Nigeria, National Protected area Authority, Sierra Leone, Rwanda Development Board, Rwanda, Société Equatoriale d’Exploitation Forestière (SEEF), Gabon, Tanzania Commission for Science and Technology, Tanzania, Tanzania Wildlife Research Institute, Tanzania, Uganda National Council for Science and Technology (UNCST), Uganda, Uganda Wildlife Authority, Uganda and Ministere de l’Enseignement Supérieur et de la Recherche Scientifique in Côte d’Ivoire.

We thank the following people who were instrumental in providing help with the PanAf project: for analytical guidance, we would like to thank Tomas Marques-Bonet, Aida Andrès Moran, Benjamin Peter, Joshua Schmidt and Lauren C. White. For Tanzania, Uganda National Council for Science and Technology (UNCST), Uganda, Rwanda, National Forestry Authority, Uganda, National Park Service, Nigeria, National Protected area Authority, Sierra Leone, Rwanda Development Board, Rwanda, Société Equatoriale d’Exploitation Forestière (SEEF), Gabon, Tanzania Commission for Science and Technology, Tanzania, Tanzania Wildlife Research Institute, Tanzania, Uganda National Council for Science and Technology (UNCST), Uganda, Uganda Wildlife Authority, Uganda and Ministere de l’Enseignement Supérieur et de la Recherche Scientifique in Côte d’Ivoire.

We thank the following people who were instrumental in providing help with the PanAf project: for analytical guidance, we would like to thank Tomas Marques-Bonet, Aida Andrès Moran, Benjamin Peter, Joshua Schmidt and Lauren C. White. For Tanzania, Uganda National Council for Science and Technology (UNCST), Uganda, Rwanda, National Forestry Authority, Uganda, National Park Service, Nigeria, National Protected area Authority, Sierra Leone, Rwanda Development Board, Rwanda, Société Equatoriale d’Exploitation Forestière (SEEF), Gabon, Tanzania Commission for Science and Technology, Tanzania, Tanzania Wildlife Research Institute, Tanzania, Uganda National Council for Science and Technology (UNCST), Uganda, Uganda Wildlife Authority, Uganda and Ministere de l’Enseignement Supérieur et de la Recherche Scientifique in Côte d’Ivoire.

Funding

Open Access funding enabled and organized by Project DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-01806-x.

Correspondence and requests for materials should be addressed to J.D.L. or M.A.

Reprints and permission information is available at http://www.nature.com/reprints

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