1 **Title**

2 # Limited Introgression Supports Division of Giraffe into Four Species

3 **Authors:**

Sven Winter[1,2]*, Julian Fennessy[3] & Axel Janke[1,2]

Affiliations:

[1]Senckenberg Biodiversity and Climate Research Centre, Senckenberganlage 25, 60325 Frankfurt am Main, Germany

[2]Goethe University, Institute for Ecology, Evolution & Diversity, Max-von-Laue-Str. 13, 60438 Frankfurt am Main, Germany

4 [3]Giraffe Conservation Foundation, PO Box 86099, Windhoek, Namibia

5

6 *corresponding author

7 Contact:

8 Sven Winter*: sven.winter@senckenberg.de

9 Julian Fennessy: julian@giraffeconservation.org

10 Axel Janke: axel.janke@senckenberg.de

11

12

13

14

15

16

Abstract

All giraffe (*Giraffa*) were previously assigned to a single species (*G. camelopardalis*) and nine subspecies. However, multi-locus analyses of all subspecies have shown that there are four genetically distinct clades and suggest four giraffe species. This conclusion might not be fully accepted due to limited data and lack of explicit gene flow analyses. Here we present an extended study based on 21 independent nuclear loci from 137 individuals. Explicit gene flow analyses identify less than one migrant per generation, including between the closely related northern and reticulated giraffe. Thus, gene flow analyses and population genetics of the extended dataset confirm four genetically distinct giraffe clades and support four independent giraffe species. The new findings call for a revision of the IUCN classification of giraffe taxonomy. Three of the four species are threatened with extinction, mostly occurring in politically unstable regions, and as such, require the highest conservation support possible.

Keywords

Conservation, gene flow, hybridization, speciation

Introduction

Traditionally, giraffe were classified as a single species (*Giraffa camelopardalis*) with up to eleven subspecies proposed (Lydekker, 1904). However until recently, the classification into nine subspecies was most widely accepted (Dagg & Foster, 1976). It has been shown that in captivity some giraffe subspecies hybridize (Gray, 1972; Lackey, 2011; Lönnig, 2011), which seemed to supported the traditional single species concept for giraffe. However, multi-locus analyses of wild giraffe nuclear loci identified four monophyletic, distinct and evolutionary old groups of giraffe that should be recognized as four distinct species (Fennessy et al., 2016). This finding conflicts with former classifications and has been questioned based on the limited interpretation of traditional data e.g. pelage pattern, number of ossicones and geographic distribution (Bercovitch et al., 2017). The initial findings of four giraffe species (Fennessy et al., 2016) could better be criticized, because it did not involve explicit gene flow analyses.

Imperative to understanding speciation in general from a genetic perspective is gene flow analyses, especially as the keystone of the biological species concept (BSC) is reproductive isolation (Coyne & Orr, 2004). The BSC implies that there is no or only very limited gene flow between species. It has been proposed that one or a limited number of effective migrants (up to 10) per generation (N_em) avoids genetic differentiation of populations and escapes a substantial loss of genetic diversity for neutral traits (Lacy, 1987; Mills & Allendorf, 1996; Vucetich & Waite, 2000; Wright, 1969). Thus, it is a conservative estimate that limited gene flow of less than one migrant per generation ($N_em < 1$) can lead to speciation, despite the occurrence of hybridization between mammal species. As shown, the BSC might need to be revised as some species naturally hybridize in the wild and produce fertile offspring e.g. bears (Arnold, 2016; Kelly, Whiteley, & Tallmon, 2010; Kumar et al., 2017) and whales (Bérubé &

Aguilar, 1998; Spilliaert et al., 1991), and divergence can occur under genetic exchange (Arnold, 2016). Whilst the distinction of four giraffe species is consistent with population genetic analyses (Fennessy et al., 2016), gene flow among giraffe species has not yet been sufficiently analyzed. Here, we revisit the hypotheses of four giraffe species using population genetic methods that explicitly involve gene flow analyses with an increased dataset of 21 nuclear loci and 137 giraffe individuals from 21 locations across Africa (Fig. 1, Supplementary Table 1).

Materials and Methods

Sampling and DNA Extraction

Tissue samples from all giraffe species and five subspecies were collected by the Giraffe Conservation Foundation (GCF) and partners using remote biopsy darts with country-specific research permits between 2009 and 2016 in accordance with ethical guidelines and regulations of the respective governments and institutions. All samples were stored either in RNAlater (Invitrogen) or > 95% ethanol. Additional giraffe samples were added to the dataset of Fennessy et al. (2016) resulting in a total number of 217 giraffe individuals, mostly southern giraffe. The geographical origins and individual IDs are shown in Supplementary Table 1. Sample locations and geographical distributions are shown in Fig. 1. Additional southern giraffe individuals were only included if they are from a hitherto unrepresented region. DNA was extracted using either a Macherey-Nagel NucleoSpin Tissue Kit or a standard phenol-chloroform extraction method. All experimental protocols are in compliance with the guidelines for the best ethical and experimental practices of the Senckenberg Society, as well as with national guidelines of the respective countries.

86

87 **Amplification and sequencing**

88 We PCR amplified and sequenced the seven intron markers previously published (Fennessy et
89 al., 2016) for 32 new individuals and developed 14 additional intron markers as described
90 (Fennessy et al., 2016). The 14 new intron markers were amplified and sequenced for a total
91 number of 137 individual giraffe and the okapi (*Okapia johnstoni*). PCRs were performed with
92 10 ng genomic DNA giraffe and okapi specific primers (see Supplementary Table 2 for primer
93 sequences and PCR conditions). We also amplified and sequenced the mitochondrial
94 cytochrome b and control region for all new individuals as described previously (Bock et al.,
95 2014). Each PCR was examined using agarose gel electrophoresis on a 1% agarose gel with
96 ethidium bromide.
97 Sanger sequencing was performed for the forward and reverse strand using the BigDye
98 terminator sequencing kit 3.1 (Applied Biosystems) with 5 ng of PCR product for each reaction
99 and analyzed on an ABI 3730 DNA Analyzer.
100 The sequences were manually edited and aligned in Geneious v.6.1.8 (Kearse et al., 2012).
101 Heterozygous insertions/deletions of nuclear sequences were resolved by hand or using
102 Indelligent v.1.2 (Dmitriev & Rakitov, 2008) and verified by allele-specific primers if necessary.
103 PHASE implemented in DnaSP v.5.10.01 (Librado & Rozas, 2009) was used to derive the allele
104 haplotypes of the nuclear sequences using a threshold of 0.6 and allowing for recombination.
105 All analyses, except of a mtDNA tree analysis, were performed using only nuclear allele
106 haplotype data.

107

108

Tree Analyses

The mitochondrial cytochrome b and control region sequences of 217 giraffe including newly generated as well as already published sequences (Bock et al., 2014; Brown et al., 2007; Fennessy et al., 2016, 2013; Hassanin et al., 2012, 2007; Winter, Fennessy, Fennessy, & Janke, 2018) (Supplementary Table 1) were aligned, as well as concatenated, and a Neighbor-Joining analysis was reconstructed in Geneious v.6.1.8 (Kearse et al., 2012). We used the HKY model of sequence evolution (Hasegawa, Kishino, & Yano, 1985), as suggested by jModelTest v.2.1.1 (Darriba, Taboada, Doallo, & Posada, 2012) with 1,000 Bootstrap replicates and sequences of two okapis were used as an outgroup.

A multi-locus Bayesian phylogenetic tree of the 21 intron markers for 137 individuals and the okapi as outgroup was generated with the StarBEAST2 (Ogilvie, Bouckaert, & Drummond, 2017) package in BEAST v.2.4.5. (Bouckaert et al., 2014, p. 2) under the JC model of nucleotide evolution as suggested as best fitting model by jModelTest v.2.1.1 (Darriba et al., 2012). A lognormal relaxed clock was used with 10^9 generations and sampling every 20,000th iteration. Convergence of the MCMC runs was analyzed with Tracer v.1.6.0 (Rambaut, Suchard, Xie, & Drummond, 2014), and TreeAnnotator v.2.4.5 (Rambaut & Drummond, 2016) was used to construct a maximum clade credibility tree with 30% burn-in for the nuclear markers.

Population Genetic Analyses

Haplotype information for each locus deduced by DnaSP (Librado & Rozas, 2009) was used to code each individual. The haplotype matrix was then used to infer admixture with the Bayesian clustering algorithm implemented in STRUCTURE v.2.3.4 . For the maximum number of populations (K) between 1-10, we sampled 250,000 steps following a 100,000-step burn-in,

with 40 replicates each. The CLUMPAK webserver (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015) was used to average the results, and to infer the most likely K based on the posterior probability of K (Pritchard et al., 2000) and ΔK (Evanno et al., 2005). Additionally, the most likely K was deduced based on the estimated Ln probability of data (Ln Pr(X|K)) (Pritchard et al., 2010) using Structure Harvester (Earl & vonHoldt, 2012). Principal Component Analyses were performed with the R package adegenet (Jombart, 2008) in R v.3.2.3 (R Core Team, 2015) to assess the degree of similarity between defined population scenarios. Pairwise Fixation index (F_{st}) values were calculated in Arlequin v.3.5.2.1 (Excoffier & Lischer, 2010) based on the nuclear haplotypes.

Gene flow analyses

Long-term average gene flow among and within the giraffe species were calculated in the coalescent genealogy sampler MIGRATE-N v.3.6.11 (Beerli, 2006; Beerli & Felsenstein, 2001) by estimating the mutation-scaled population sizes (Θ) for each population and migration rates (M) for each direction between a pair of populations. We used the Brownian motion mutation model and the Bayesian inference analysis strategy, as some parameter combinations are better estimated using the Bayesian approach compared to the Maximum-likelihood approach (Beerli, 2012). The transition/transversion ratio was set to 2.31 as estimated in MEGA v.7.0.16 (Kumar, Stecher, & Tamura, 2016) based on a concatenated alignment of all 21 loci. Variable mutation rates were considered amongst loci. We used the default settings for the Θ uniformpriors, and adjusted the M uniformpriors (0; 5,000; 10,000; 1,000), because the upper prior boundary appeared to be too small in initial first analyses. Several short-runs were performed to check for convergence of the runs. A long-chain run was

performed for 6 million Markov Chain Monte Carlo (MCMC) iterations (60,000 recorded steps) and a burn-in of 600,000 iterations. An adaptive heating scheme was used with four chains and temperatures set by default with a swapping interval of 1. Convergence of the runs was checked by the posterior distributions, Effective Sample Size (ESS) and consistency of results between runs. In addition, we estimated short-term gene flow, as well as the probability of recent hybridization for each individual in BayesAss v.3.0.4 (Wilson & Rannala, 2003) using 100 million MCMC iterations, a burn-in of 10 million and a sampling interval of 1000 iterations. Mixing of the chain was improved by adjusting the acceptance rates for proposed changes to the parameters (allele frequencies and inbreeding coefficient) by adapting the mixing parameters for allele frequencies (ΔA) and inbreeding coefficients (ΔF) to 0.30. Convergence was checked in Tracer v.1.6.0 (Rambaut et al., 2014) and by consistency of results of several runs with different initial seeds. Results for short-term gene flow were visualized in circos plots using the Circos Table Viewer v.0.63-9 (Krzywinski et al., 2009).

Calculation of gene flow rate

We calculated the effective number of migrants per generation (N_em) or rate of gene flow using two different methods. The first method to calculate N_em was based on the pairwise F_{st} values using equation (1) by Wright (1951):

$$(1)\ N_e m = \frac{(1/F_{st} - 1)}{4}.$$

In addition, we calculated the N_em using the coalescent-based estimates for the mutation-scaled population size Θ and the mutation-scaled immigration rate M derived from MIGRATE-

N. For autosomal markers equation (2) expresses the relationship between Θ_j (population size of the population receiving migrants) and M_{ij} (corresponding migration rate) (Marko & Hart, 2011):

$$(2)\ N_e m_{i>j} = \frac{M_{i>j} \times \Theta_j}{4}.$$

Results

The putatively independent 21 nuclear gene loci are on different chromosomes or are clearly separated from each other in the bovine genome, a close relative with available chromosome level genome data (Supplementary Table 2). A Bayesian multi-locus tree analysis of the 21 nuclear loci (total of 16,969 nucleotides) for 137 giraffe, including all traditionally recognized giraffe subspecies (Fig. 2a), implies a clear separation into four giraffe clades: (1) a northern giraffe cluster including West African (*G. c. peralta*), Kordofan (*G. c. antiquorum*), and Nubian giraffe (*G. c. camelopardalis*) which includes the former Rothschild´s giraffe (*G. c. rothschildi*), (2) the reticulated giraffe (*G. reticulata*), (3) the Masai giraffe (*G. tippelskirchi*) including former Thornicroft´s giraffe (*G. c. thornicrofti*), and (4) a southern giraffe cluster (*G. giraffa*) including South African (*G. g. giraffa*) and Angolan giraffe (*G. g. angolensis*). The monophyly of each of these four clades is supported by a posterior probability of $p \geq 0.95$. However, in the analyses the exact relationships of southern and Masai giraffe relative to northern and reticulated giraffe could not be determined with significant probability ($p \approx 0.81$) (not shown). A mtDNA Neighbor-Joining tree (Supplementary Fig. 1) confirms the reciprocal monophyly of seven distinct subspecies clusters with Bootstrap support of ≥ 80% (Bock et al., 2014; Brown et al., 2007; Fennessy et al., 2016). MtDNA do not support two subspecies of Masai giraffe because individuals which are designated Masai giraffe individuals, disrupt a possible

reciprocal monophyly. For reticulated giraffe, three individuals do not group as expected but rather fall within the northern giraffe, indicating possible hybridization. However, two of these individuals are from zoos, where hybridization is common, and have an unknown history. The third individual is a wild giraffe from a geographic range adjacent to the northern giraffe and is a possible natural hybrid, but during sampling was identified as a phenotypically reticulated giraffe.

Multi-locus population STRUCTURE analyses (Pritchard, Stephens, & Donnelly, 2000) of 21 nuclear loci (Fig. 2b) proposes the best clustering into four distinct populations (optimal K = 4) based on the graphical display. At K = 3 the analyses merge the reticulated and the northern giraffe and at K ≥ 5 the analyses do not produce further clustering. Three different statistical methods to interpret the STRUCTURE results (Evanno, Regnaut, & Goudet, 2005; Pritchard et al., 2000; Pritchard, Wen, & Falush, 2010) confirm K = 4 being significantly the best fitting number of populations (Supplementary Fig. 2). These four clusters conform with the four giraffe clades identified by tree analyses. Intriguingly, STRUCTURE also identifies three potential hybrids between the northern and reticulated giraffe within the reticulated giraffe clade (Fig 2a, Fig 2b). The distinctness of four unique giraffe clades is in addition supported by Principal Component Analyses (PCAs) (Fig. 2c) with significant non-overlapping 95% confidential intervals. PCAs using groups of the seven mtDNA clades do not find more than four distinct clusters (Supplementary Fig. 3). Finally, pairwise fixation indices (F_{st}) of ≥ 0.237 (statistically significant at $p < 0.001$) are consistent with the four distinct clusters of giraffe in the tree analyses (Supplementary Table 3).

Separate PCAs and STRUCTURE analyses for each species (Supplementary Fig. 4-7) indicate population substructure within northern giraffe, and potentially in the Masai giraffe, but no

224 further population substructure in southern and reticulated giraffe. Within northern giraffe
225 STRUCTURE and PCAs up to four clusters can be identified. However, the sample sizes of some
226 populations (three for Ethiopia) are arguably insufficient to draw definitive conclusions.
227 Within the Masai giraffe STRUCTURE and PCAs identify potentially two separate clusters,
228 indicating a possible separation of the two geographically most distant populations that have
229 been analyzed for nuclear SNPs (single nucleotide polymorphisms) to date. Consistent with
230 the STRUCTURE and PCA analyses, pairwise F_{st} analyses within each giraffe species finds a high
231 level of population differentiation within northern and possibly Masai giraffe, and little
232 differentiation within southern and reticulated giraffe (Supplementary Table 4).
233
234 We estimated long-term gene flow within all four giraffe clades, as well as among subspecies
235 within each of the giraffe species which show population substructure in STRUCTURE and
236 PCAs using MIGRATE-N (Beerli, 2006; Beerli & Felsenstein, 2001). Assuming similar mutation
237 rates among all giraffe species, the mutation-scaled population size theta (Θ) estimates for
238 the four species suggest that the effective population size (N_e) is smaller in southern giraffe
239 and Masai giraffe than in northern and reticulated giraffe (Supplementary Table 5a). Thus, the
240 population size in the northern and reticulated giraffe had been larger in the past. The
241 calculated effective numbers of migrants per generation or gene flow rate (N_em) based on Θ
242 and the mutation-scaled migration rate (M) (Supplementary Table 5b) indicate generally very
243 low level of gene flow among most of the four giraffe clades with a maximum of one migrant
244 per five generations ($N_em \leq 0.179$), with one exception. A higher N_em occurs between the
245 northern and reticulated giraffe, with nearly one migrant per generation in the direction of
246 the reticulated giraffe ($N_em = 0.945$), but much less migration is observed in the opposite

direction to northern giraffe (N_em = 0.179). There is little (ca. one in ten) directional gene flow from Masai to reticulated giraffe (N_em = 0.107) and from southern to reticulated giraffe (N_em = 0.104) with nearly zero gene flow in the opposite direction. The gene flow rates for all other species pairs are extremely low (N_em < 0.065). Within species long-term gene flow rates are on average higher (N_em >1) (Supplementary Table 6b). However, between some subspecies gene flow is also limited, in particular the geographically extremely isolated West African giraffe (WA).

Gene flow rates that were calculated on pairwise F_{st} values between species corroborate the MIGRATE-N analyses, but provide no information about the direction of gene flow and the rates are somewhat higher (Supplementary Table 3). In agreement with the MIGRATE-N analyses, the F_{st} based analyses find the highest rate of gene flow between northern and reticulated giraffe (N_em = 0.804), and a much lower gene flow rate (N_em ranges from 0.113 to 0.186) is observed among all other population pairs.

Finally, short-term migration rates (m) estimated with BayesAss (Wilson & Rannala, 2003) (Figure 3, Supplementary Table 5b) confirm low levels of gene flow among the four giraffe species for the past three generations. The highest migration rates occur from northern, Masai and southern giraffe in the direction of reticulated giraffe. The data suggest that approximately 2% (m = 0.021) of the reticulated giraffe population are derived from each of these neighboring species, which is expected. In comparison, to the other gene flow analyses, BayesAss identifies somewhat higher recent migration rates (m) among subspecies within species (Supplementary Table 6b). This is consistent with the lack of genetic differentiation

identified by PCA and F_{st} analyses. The recent migration rates estimated by BayesAss analyses suggest directional gene flow between West African and Kordofan giraffe (m = 0.064), and find gene flow between South African and Angolan giraffe (m = 0.052). Most importantly, however is that BayesAss does not find any first or second generation hybrids.

Discussion

Morphology and genetic analyses suggest that there is more than one giraffe species (Brown et al., 2007; Fennessy et al., 2016; Groves & Grubb, 2011). Here we expand our previous dataset three-fold and improve the sampling of northern and reticulated giraffe, to further study if there are indeed more than one species. The new data set allows for the first-time detailed gene flow and migration analyses. Among the four giraffe species[6], gene flow and migration is very limited. As such, the new analyses of the extended nuclear data corroborate the identification of four genetically distinct giraffe species (Fennessy et al., 2016).

Several attempts have been made to define a species, but a unequivocal consensus has not yet been reached (Coyne & Orr, 2004; De Queiroz, 2007). The most commonly applied model is the BSC, which suggests that reproductive isolation is essential to delineate species (Dobzhansky, 1970; Mayr, 1942). By contrast, subspecies or evolutionary significant units (ESU) are sometimes arbitrary distinctions within a species. Reproductive isolation is also a cornerstone of other species concepts that define species as distinct ESUs with limited gene flow to other such units (Avise & Ball, 1990). Therefore, analyzing gene flow among species is a central analysis to delineate species, especially if, like in giraffe, they possibly hybridize in nature. It has been suggested that gene flow among species must be limited to allow genetic differentiation, and a value below one migrant per generation (< 1 N_em) is a conservative

293 estimate (Wright, 1969), even if other studies are more liberal and suggest that gene flow
294 rates of < 5 N_em (Lacy, 1987) or even < 10 N_em (Mills & Allendorf, 1996; Vucetich & Waite,
295 2000) can allow genetic differentiation and consequently speciation.
296 The initial finding of four giraffe species was unexpected (Fennessy et al., 2016), because
297 giraffe seem to be a morphologically homogenous group, can interbreed in captivity (Gray,
298 1972; Lackey, 2011; Lönnig, 2011), and are highly mobile (Flanagan, Brown, Fennessy, &
299 Bolger, 2016). To avoid differentiation between populations, and if giraffe was in fact one
300 species, gene flow rates in excess of 1-10 migrants per generation are expected based on
301 mathematical models (Lacy, 1987; Mills & Allendorf, 1996; Vucetich & Waite, 2000; Wright,
302 1969).
303 Our study show that the long-term average estimates of gene flow rate among giraffe are
304 below one migrant per generation (N_em < 1). The introgression and population genetic
305 analyses of the expanded data set are thus consistent with the previously proposed
306 classification of four giraffe species by Fennessy et al. (2016). The highest gene flow rates are
307 observed between the northern and reticulated giraffe, but the rate is below Wright's (1969)
308 conservative estimate of < 1 N_em. Compared to other giraffe species, higher gene flow rates
309 among northern and reticulated giraffe is not unexpected, because they are the closest
310 related (Fennessy et al., 2016) and their current neighboring geographic ranges might have
311 overlapped historically (Fig. 1).
312 Yet, even among the closely related and neighboring northern and reticulated giraffe lower
313 gene flow rates than < 1 N_em are observed, which is consistent with being genetically
314 differentiated species. In addition, among all 137 individuals from a wide geographic

315 distribution, only one natural hybrid has been genetically identified yet. The rare occurrence
316 of hybrids further supports the existence of four giraffe species.
317 Population genetic analyses, such as STRUCTURE and PCA of the data set support the results
318 from the gene flow analyses. The new results are inconsistent with past suggestions of
319 possibly six or seven distinct giraffe species (Brown et al., 2007). These results were based on
320 non-stringent conclusions from STRUCTURE analyses with 11 separate genetic clusters at K =
321 13 based on 14 microsatellites, and results of six to seven giraffe clusters based on mtDNA
322 phylogeny (Brown et al., 2007): “11 of the 18 sampling localities resolved as distinct genetic
323 clusters at K=13”, however the authors concluded that only “the seven lineages that are
324 reciprocally monophyletic in the mtDNA tree need to be considered evolutionary significant
325 units if not species”. Other findings of up to eight giraffe species were proposed based on a
326 combination of limited genetic analyses (Brown et al., 2007; Hassanin, Ropiquet, Gourmand,
327 Chardonnet, & Rigoulet, 2007) and morphological characteristics (Groves & Grubb, 2011),
328 however the location of some samples were inaccurate.
329 Both Fennessy et al. (2016) and Bock et al. (2014) suggested to subsume Rothschild’s giraffe
330 (MF) into the Nubian giraffe, as well as Thornicroft’s giraffe (LVNP) into the Masai giraffe,
331 because they lack differentiation at mtDNA sequences. Evolutionary differentiation of
332 populations is often first evident in mtDNA, because theory suggests that this locus, due to its
333 maternal inheritance and non-recombining nature, reaches fixation 4-times more rapidly than
334 nuclear loci (Zink & Barrowclough, 2008). Thus, differentiation into subspecies is often first
335 evident on mtDNA, rather than nuclear sequences. Such population differentiation processes
336 have been reported in natural population of bears (Hailer et al., 2012), humpback whales
337 (Palumbi & Baker, 1994) and macaques (Melnick & Hoelzer, 1992).

While the current mtDNA analyses support previous findings (Fennessy et al., 2016) of Thornicroft's giraffe being subsumed into the Masai giraffe, new and extended nuclear gene datasets identify some substructure among them. We emphasize however, that the nuclear loci have only been sampled from across a limited distribution of the Masai giraffe[6]. Additional sampling of intermediate Masai giraffe populations and additional nuclear gene loci will be necessary to yield more definite results. The first detailed mtDNA analyses on Thornicroft's giraffe (Fennessy, Bock, Tutchings, Brenneman, & Janke, 2013) proposed that while they are not reciprocal monophyletic, the geographic location in Zambia's Luangwa Valley is unique and should, for conservation efforts, tentatively maintain its subspecies status as Thornicroft's giraffe within Masai giraffe (*Giraffa tippelskirchi thornicrofti*).

Within the northern giraffe some substructure is evident in PCAs and STRUCUTURE analyses for nuclear sequences (Supplementary Fig. 6). However, the West African giraffe is a geographically very isolated and small population of ~600 individuals. As described (Fennessy et al., 2016), the geographic distinction between the former Nubian and Kordofan giraffe is unclear and current data suggest that they are not genetically isolated.

Multi locus phylogenies, population genetic and gene flow analyses support the hypothesis of four genetically distinct giraffe species (Fennessy et al., 2016). The molecular data show that there is only very limited gene flow between the four species, which is in agreement with the BSC.

With little more than 5,000 northern giraffe, <8,700 reticulated giraffe and ~34,000 Masai giraffe remaining in the wild (Giraffe Conservation Foundation, 2017), recognizing these – and the southern giraffe – as separate species has an impact on giraffe conservation. Their decline in numbers over the last thirty years (three generations) – northern giraffe (~95 %), reticulated

giraffe (~80%) and Masai giraffe (~52 %), highlight that these species are threatened with extinction(IUCN, 2017). Giraffe, as a single species, and not four, were recently listed as "Vulnerable" on the IUCN Red List (Muller et al., 2016). The mounting evidence of four giraffe species proposes a re-evaluation of the current IUCN giraffe taxonomy to raise the conservation classification to a higher level of threat, and in turn increased conservation management actions.

Acknowledgements

We thank Susanne Gallus, for support in the laboratory, and Fritjof Lammers as well as Markus Pfenninger for support with analyses. Stephanie Fennessy and Maria Nilsson gave valuable comments on the manuscript. We thank Jon B. Hlidberg for the artwork of a Nubian giraffe. This study was supported by the Leibniz Association, the Giraffe Conservation Foundation-USA, the Leiden Conservation Foundation, and various African government partners and international zoo and private supporters.

References

Arnold, M. L. (2016). *Divergence with Genetic Exchange*. Oxford University Press.

Avise, J. C., & Ball, R. M., Jr. (1990). Principles of genealogical concordance in species concepts and biological taxonomy. In *Oxford Surveys in Evolutionary Biology* (Vol. 7, pp. 45–67).

Beerli, P. (2012). Migrate documentation version 3.2. 1. *Florida State University, Tallahasee, FL*, 119.

Beerli, Peter. (2006). Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, *22*(3), 341–345. doi:10.1093/bioinformatics/bti803

Beerli, Peter, & Felsenstein, J. (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences*, *98*(8), 4563–4568. doi:10.1073/pnas.081068098

389 Bercovitch, F. B., Berry, P. S. M., Dagg, A., Deacon, F., Doherty, J. B., Lee, D. E., … Tutchings, A.
390 (2017). How many species of giraffe are there? *Current Biology*, *27*(4), R136–R137.
391 doi:10.1016/j.cub.2016.12.039
392 Bérubé, M., & Aguilar, A. (1998). A New Hybrid Between a Blue Whale, Balaenoptera
393 Musculus, and a Fin Whale, B. Physalus: Frequency and Implications of Hybridization.
394 *Marine Mammal Science*, *14*(1), 82–98. doi:10.1111/j.1748-7692.1998.tb00692.x
395 Bock, F., Fennessy, J., Bidon, T., Tutchings, A., Marais, A., Deacon, F., & Janke, A. (2014).
396 Mitochondrial sequences reveal a clear separation between Angolan and South African
397 giraffe along a cryptic rift valley. *BMC Evolutionary Biology*, *14*(1), 1–12.
398 doi:10.1186/s12862-014-0219-7
399 Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., … Drummond, A. J. (2014).
400 BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS Computational*
401 *Biology*, *10*(4), e1003537. doi:10.1371/journal.pcbi.1003537
402 Brown, D. M., Brenneman, R. A., Koepfli, K.-P., Pollinger, J. P., Milá, B., Georgiadis, N. J., …
403 Wayne, R. K. (2007). Extensive population genetic structure in the giraffe. *BMC Biology*,
404 *5*(1), 1–13. doi:10.1186/1741-7007-5-57
405 Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer.
406 Dagg, A. I., & Foster, J. B. (1976). *The giraffe: its biology, behavior, and ecology*. Van Nostrand
407 Reinhold Company New York.
408 Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new
409 heuristics and parallel computing. *Nature Methods*, *9*(8), 772–772.
410 doi:10.1038/nmeth.2109
411 De Queiroz, K. (2007). Species Concepts and Species Delimitation. *Systematic Biology*, *56*(6),
412 879–886. doi:10.1080/10635150701701083
413 Dmitriev, D. A., & Rakitov, R. A. (2008). Decoding of Superimposed Traces Produced by Direct
414 Sequencing of Heterozygous Indels. *PLOS Computational Biology*, *4*(7), e1000113.
415 doi:10.1371/journal.pcbi.1000113
416 Dobzhansky, T. (1970). *Genetics of the Evolutionary Process*. Columbia University Press.
417 Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for
418 visualizing STRUCTURE output and implementing the Evanno method. *Conservation*
419 *Genetics Resources*, *4*(2), 359–361. doi:10.1007/s12686-011-9548-7
420 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals
421 using the software structure: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620.
422 doi:10.1111/j.1365-294X.2005.02553.x
423 Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to
424 perform population genetics analyses under Linux and Windows. *Molecular Ecology*
425 *Resources*, *10*(3), 564–567. doi:10.1111/j.1755-0998.2010.02847.x
426 Fennessy, J., Bidon, T., Reuss, F., Kumar, V., Elkan, P., Nilsson, M. A., … Janke, A. (2016). Multi-
427 locus Analyses Reveal Four Giraffe Species Instead of One. *Current Biology*, *26*(18),
428 2543–2549. doi:10.1016/j.cub.2016.07.036
429 Fennessy, J., Bock, F., Tutchings, A., Brenneman, R., & Janke, A. (2013). Mitochondrial DNA
430 analyses show that Zambia's South Luangwa Valley giraffe (Giraffa camelopardalis
431 thornicrofti) is genetically isolated. *Afr J Ecol*, *51*. doi:10.1111/aje.12085

Flanagan, S. E., Brown, M. B., Fennessy, J., & Bolger, D. T. (2016). Use of home range behaviour to assess establishment in translocated giraffes. *African Journal of Ecology*, *54*(3), 365–374. doi:10.1111/aje.12299

Giraffe Conservation Foundation. (2017). *Africa's Giraffe – A Conservation Guide*. Windhoek, Namibia: Giraffe Conservation Foundation. Retrieved from https://giraffeconservation.org/

Gray, A. P. (1972). Mammalian hybrids: a check-list with bibliography. *Mammalian Hybrids: A Check-List with Bibliography.*, (2nd edition).

Groves, C., & Grubb, P. (2011). *Ungulate taxonomy*. JHU Press.

Hailer, F., Kutschera, V. E., Hallström, B. M., Klassert, D., Fain, S. R., Leonard, J. A., … Janke, A. (2012). Nuclear Genomic Sequences Reveal that Polar Bears Are an Old and Distinct Bear Lineage. *Science*, *336*. doi:10.1126/science.1216424

Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, *22*(2), 160–174. doi:10.1007/BF02101694

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., Jansen Van Vuuren, V., Matthee, C., … Couloux, A. (2012). Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasitheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C R Biol*, *335*. doi:10.1016/j.crvi.2011.11.002

Hassanin, A., Ropiquet, A., Gourmand, A. L., Chardonnet, B., & Rigoulet, J. (2007). Mitochondrial DNA variability in Giraffa camelopardalis: consequences for taxonomy, phylogeography and conservation of giraffes in West and central Africa. *C R Biol*, *330*. doi:10.1016/j.crvi.2007.02.008

IUCN. (2017). IUCN. The IUCN Red List of Threatened Species. Version 2017-3. <http://www.iucnredlist.org>. Downloaded on 05 December 2017. Retrieved from http://www.iucnredlist.org

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. doi:10.1093/bioinformatics/btn129

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. doi:10.1093/bioinformatics/bts199

Kelly, B. P., Whiteley, A., & Tallmon, D. (2010). The Arctic melting pot. *Nature*, *468*(7326), 891–891. doi:10.1038/468891a

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. doi:10.1111/1755-0998.12387

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., … Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, *19*(9), 1639–1645. doi:10.1101/gr.092759.109

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. doi:10.1093/molbev/msw054

476 Kumar, V., Lammers, F., Bidon, T., Pfenninger, M., Kolter, L., Nilsson, M. A., & Janke, A. (2017).
477 The evolutionary history of bears is characterized by gene flow across species.
478 *Scientific Reports*, *7*. doi:10.1038/srep46487
479 Lackey, L. B. (2011). *Giraffe Studbook Giraffa camelopardalis North American Regional /*
480 *Global*.
481 Lacy, R. C. (1987). Loss of Genetic Diversity from Managed Populations: Interacting Effects of
482 Drift, Mutation, Immigration, Selection, and Population Subdivision. *Conservation*
483 *Biology*, *1*(2), 143–158. doi:10.1111/j.1523-1739.1987.tb00023.x
484 Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA
485 polymorphism data. *Bioinformatics*, *25*(11), 1451–1452.
486 doi:10.1093/bioinformatics/btp187
487 Lönnig, W.-E. (2011). *The Evolution of the Long-necked Giraffe (Giraffa Camelopardalis L.):*
488 *What Do We Really Know? ; Testing the Theories of Gradualism, Macromutation, and*
489 *Intelligent Design ; a Scientific Treatise*. MV-Verlag.
490 Lydekker, R. (1904). On the subspecies of Giraffa camelopardalis. *Proc Zool Soc Lond*, *74*.
491 doi:10.1111/j.1469-7998.1904.tb08288.x
492 Marko, P. B., & Hart, M. W. (2011). The complex analytical landscape of gene flow inference.
493 *Trends in Ecology & Evolution*, *26*(9), 448–456. doi:10.1016/j.tree.2011.05.007
494 Mayr, E. (1942). *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*.
495 Harvard University Press.
496 Melnick, D., & Hoelzer, G. (1992). *Difference in male and female macaque dispersal lead to*
497 *contrasting distributions of nuclear and mitochondrial DNA variation* (Vol. 13).
498 doi:10.1007/BF02547824
499 Mills, L. S., & Allendorf, F. W. (1996). The One-Migrant-per-Generation Rule in Conservation
500 and Management. *Conservation Biology*, *10*(6), 1509–1518.
501 Muller, Z., Bercovitch, F., Brand, R., Brown, D., Brown, M., Bolger, D., … Wube, T. (2016).
502 Giraffa camelopardalis. The IUCN Red List of Threatened Species 2016:
503 e.T9194A51140239. doi:http://dx.doi.org/10.2305/IUCN.UK.2016-
504 3.RLTS.T9194A51140239.en.
505 Ogilvie, H. A., Bouckaert, R. R., & Drummond, A. J. (2017). StarBEAST2 brings faster species
506 tree inference and accurate estimates of substitution rates. *Molecular Biology and*
507 *Evolution*. doi:10.1093/molbev/msx126
508 Palumbi, S. R., & Baker, C. S. (1994). Contrasting population structure from nuclear intron
509 sequences and mtDNA of humpback whales. *Molecular Biology and Evolution*, *11*(3),
510 426–435. doi:10.1093/oxfordjournals.molbev.a040115
511 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using
512 Multilocus Genotype Data. *Genetics*, *155*(2), 945–959.
513 Pritchard, J., Wen, X., & Falush, D. (2010). Documentation for STRUCTURE software, version
514 2.3. University of Chicago, Chicago, IL.
515 R Core Team. (2015). R: A language and environment for statistical computing: R Foundation
516 for Statistical Computing. Available from http://www.R-project.org. *Vienna, Austria.*
517 Retrieved from http://www. R-project. org
518 Rambaut, A., & Drummond, A. (2016). TreeAnnotator v2.4.5 as part of the BEAST software
519 package. *University of Edinburgh, Institute of Evolutionary Biology*.

Rambaut, A., Suchard, M., Xie, D., & Drummond, A. (2014). Tracer v1. 6 http://beast. bio. ed. ac. uk. *Tracer>(Online 2015, May 29)*.

Spilliaert, R., Vikingsson, G., Arnason, U., Palsdottir, A., Sigurjonsson, J., & Arnason, A. (1991). Species Hybridization between a Female Blue Whale (Balaenoptera musculus) and a Male Fin Whale (B. physalus): Molecular and Morphological Documentation. *Journal of Heredity*, *82*(4), 269–274. doi:10.1093/oxfordjournals.jhered.a111085

Vucetich, J. A., & Waite, T. A. (2000). Is one migrant per generation sufficient for the genetic management of fluctuating populations? *Animal Conservation*, *3*(3), 261–266. doi:10.1111/j.1469-1795.2000.tb00111.x

Wilson, G. A., & Rannala, B. (2003). Bayesian Inference of Recent Migration Rates Using Multilocus Genotypes. *Genetics*, *163*(3), 1177–1191.

Winter, S., Fennessy, J., Fennessy, S., & Janke, A. (2018). Matrilineal population structure and distribution of the Angolan giraffe in the Namib desert and beyond. *Ecological Genetics and Genomics*, *7–8*, 1–5. doi:10.1016/j.egg.2018.03.003

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, *15*(4), 323–354.

Wright, Sewall. (1969). *Evolution and the Genetics of Populations, Volume 2: Theory of Gene Frequencies*. University of Chicago Press.

Zink, R. M., & Barrowclough, G. F. (2008). Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, *17*(9), 2107–2121. doi:10.1111/j.1365-294X.2008.03737.x

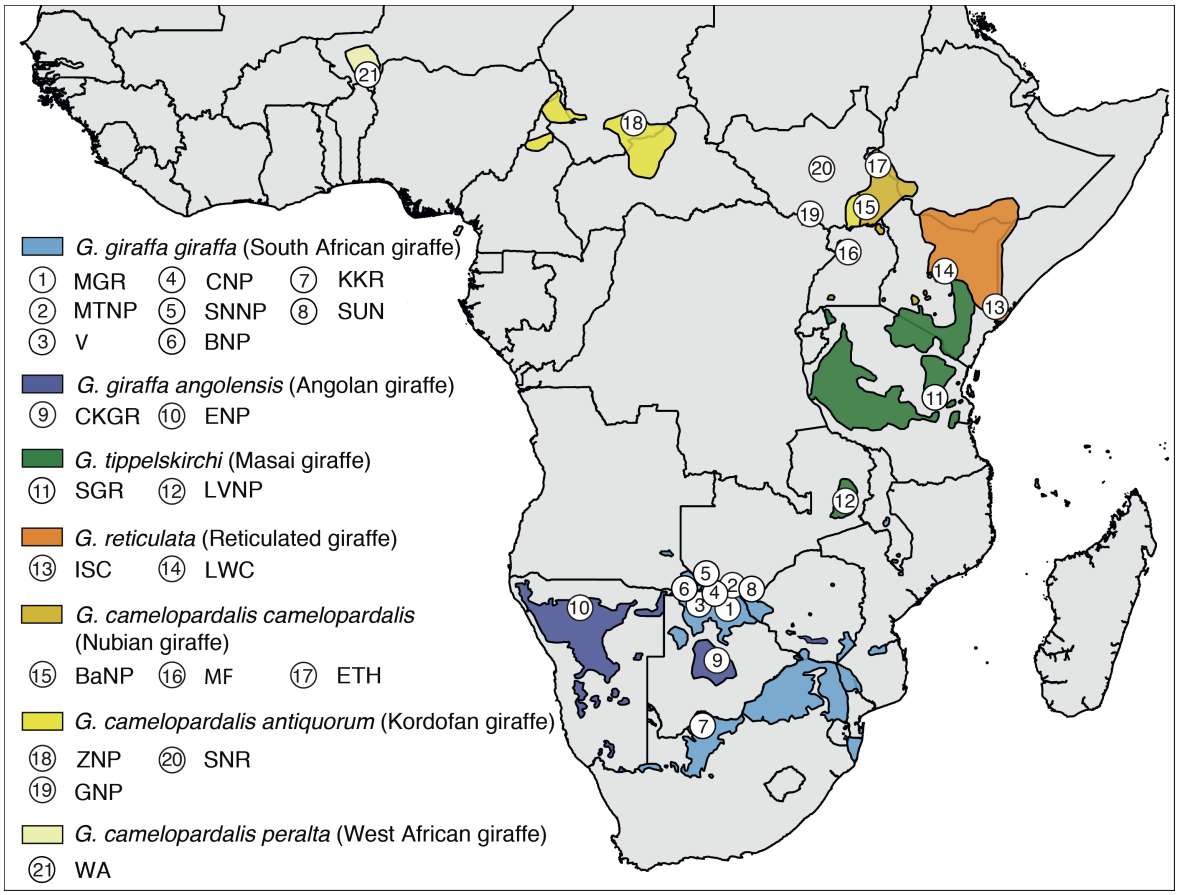
Data Accessibility Statement

The authors declare that all the data supporting the findings of this study are available within the article and its Supplementary Information files. Sequences generated during the study are available at GenBank (https://www.ncbi.nlm.nih.gov/nucleotide/) under the Accession Numbers MG257948 - MG262301.

Author Contributions

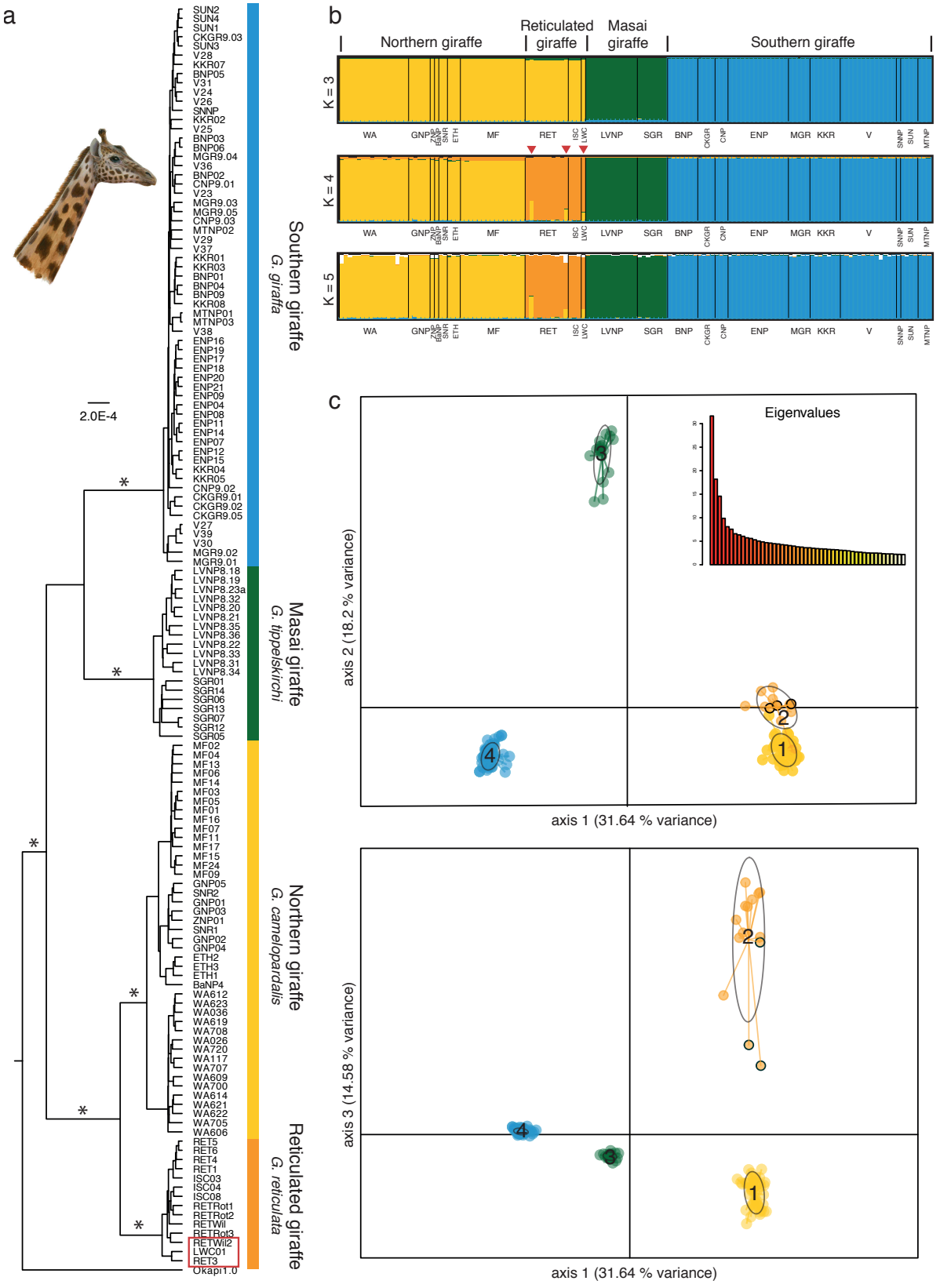
AJ, JF and SW designed and conceived the study. AJ and JF funded the project, JF collected the samples and provided biological data. SW developed markers, generated and analyzed the data. SW and AJ wrote the manuscript with input from JF.

554 **Figures**



555

556 **Fig. 1 Map of Sub-Saharan Africa with giraffe (sub)species distributions and sampling**
557 **locations.**
558 Geographic ranges (colored shadings) of giraffe as identified by the Giraffe Conservation
559 Foundation (2017) were plotted on a map of Sub-Saharan Africa. Numbered circles represent
560 sampling locations (for details see Supplementary Table 1). Species and common names as
561 per Fennessy et al. (2016).



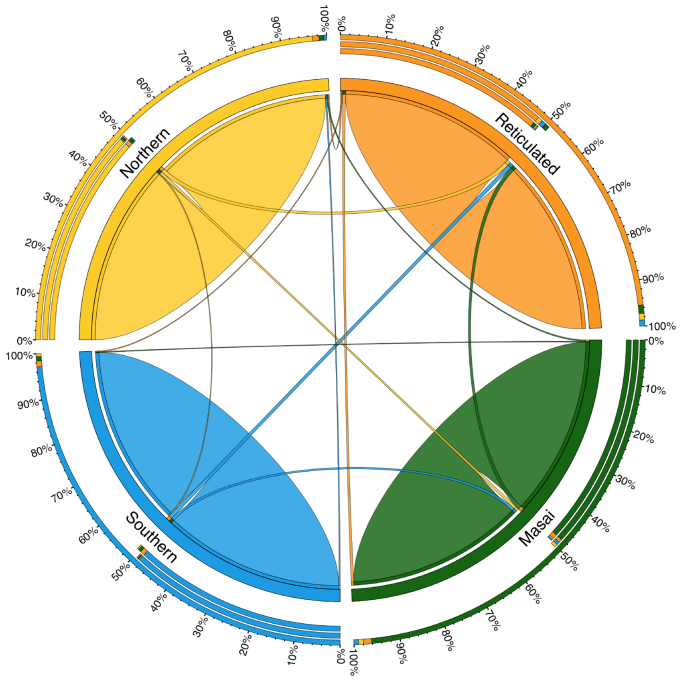
564 **Fig. 2 Nuclear phylogeny and population structuring of giraffe**
565 (a) Bayesian multi-locus tree from 21 nuclear loci and 137 giraffe individuals reconstruct four
566 significant supported ($p \geq 0.95$) giraffe clades, corresponding to the four giraffe species
567 (Fennessy et al., 2016). The okapi is used as the outgroup. The asterisks indicate branches
568 with statistical significant support ($p \geq 0.95$). The red frame indicates the potential hybrids.
569 (b) STRUCTURE analysis of the dataset, excluding the okapi. The colors indicate the
570 membership in a cluster for each sampling location and individual. K = 4 shows four well-
571 resolved groups and is supported as best fitting number of clusters by several statistical
572 methods (see Supplementary Fig. 2). The grouping into four clusters is consistent with the
573 Bayesian multi-locus analysis: yellow: northern giraffe, orange: reticulated giraffe, green:
574 Masai giraffe, and blue: southern giraffe. Three individuals within the reticulated giraffe
575 cluster (red arrowheads) indicate potential hybridization with admixture from the northern
576 giraffe. K = 3 merges northern and reticulated giraffe, and at $K \geq 5$ no further clustering is
577 evident.
578 (c) PCA axes 1-2 and axes 1-3 for four distinct giraffe clusters (1: northern; 2: reticulated; 3:
579 Masai; 4: southern). Colors as in Fig 2b. The 95% confidence intervals are shown as oval
580 outlines. Note that the non-overlapping confidential intervals in the PCA axes 1-2, as well as,
581 axes 1-3 indicate significantly different clusters. Potential hybrids are indicated by black
582 circles.
583 Note – The drawing by Jon B. Hlidberg shows a Nubian giraffe.
584

585

586

587

588



589

590 **Fig. 3 Circular migration plot of recent migration rates among four giraffe clades.**
591 Recent directional migration rates (m) as estimated by BayesAss and indicated by ribbons
592 connecting one species to another. The color coding of the four species is according to the
593 STRUCTURE clusters (Fig 2b). Peripheral concentric stack bars show relative migration rates in
594 percent. Whereas the inner stack bar shows the outgoing ribbon sizes, the middle stack bar
595 the incoming ribbon sizes and the outer stack bar the combination of both.