

A comparative approach for species delimitation based on multiple methods of multi-locus DNA sequence analysis: A case study of the genus Giraffa (Mammalia, Cetartiodactyla)

Alice Petzold, Alexandre Hassanin

▶ **To cite this version:**

Alice Petzold, Alexandre Hassanin. A comparative approach for species delimitation based on multiple methods of multi-locus DNA sequence analysis: A case study of the genus Giraffa (Mammalia, Cetartiodactyla). PLoS ONE, 2020, 15 (2), pp.e0217956. 10.1371/journal.pone.0217956 . hal-02862423

HAL Id: hal-02862423

https://hal.sorbonne-universite.fr/hal-02862423

Submitted on 9 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PLOS ONE

RESEARCH ARTICLE

A comparative approach for species delimitation based on multiple methods of multi-locus DNA sequence analysis: A case study of the genus *Giraffa* (Mammalia, Cetartiodactyla)

Alice Petzold[1,2], **Alexandre Hassanin**[1,2]*

1 Institut de Systématique, Évolution, Biodiversité (ISYEB), Sorbonne Université, MNHN, CNRS, EPHE, Paris, France, **2** Muséum national d'Histoire naturelle, CP51, Paris, France

* alexandre.hassanin@mnhn.fr

Check for updates

OPEN ACCESS

Citation: Petzold A, Hassanin A (2020) A comparative approach for species delimitation based on multiple methods of multi-locus DNA sequence analysis: A case study of the genus *Giraffa* (Mammalia, Cetartiodactyla). PLoS ONE 15 (2): e0217956. https://doi.org/10.1371/journal.pone.0217956

Editor: Ulrich Joger, State Museum of Natural History, GERMANY

Received: May 21, 2019

Accepted: January 24, 2020

Published: February 13, 2020

Copyright: © 2020 Petzold, Hassanin. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All alignments generated for this study were deposited in Open Science Framework at https://osf.io/9wv86/.

Funding: The PhD thesis of AP was funded by LabEx BCDiv (https://labex-bcdiv.mnhn.fr) (project written by AH in 2015). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. This work was supported by a grant from Agence Nationale de la Recherche under the LabEx ANR-10-

Abstract

Molecular data are now commonly used in taxonomy for delimiting cryptic species. In the case of giraffes, which were treated as a single species (*Giraffa camelopardalis*) during half of a century, several molecular studies have suggested a splitting into four to seven species, but the criteria applied for taxonomic delimitation were not fully described. In this study, we have analysed all multi-locus DNA sequences available for giraffes using multispecies coalescent (MSC: *BEAST, BPP and STACEY), population genetic (STRUCTURE, allelic networks, haplotype network and bootstrapping, haplowebs and conspecificity matrix) and phylogenetic (MrBayes, PhyML, SuperTRI) methods to identify the number of species. Our results show that depending on the method chosen, different taxonomic hypotheses, recognizing from two to six species, can be considered for the genus *Giraffa*. Our results confirm that MSC methods can lead to taxonomic over-splitting, as they delimit geographic structure rather than species. The 3-species hypothesis, which recognizes *G. camelopardalis* sensu strico A, *G. giraffa*, and *G. tippelskirchi*, is highly supported by phylogenetic analyses and also corroborated by most population genetic and MSC analyses. The three species show high levels of nucleotide divergence in both nuclear (0.35–0.51%) and mitochondrial sequences (3–4%), and they are characterised by 7 to 12 exclusive synapomorphies (ES) detected in nine of the 21 nuclear introns analysed for this study. By contrast, other putative species, such as *G. peralta*, *G. reticulata*, *G. thornicrofti* or *G. tippelskirchi* sensu stricto, do not exhibit any ES in the nuclear genes. A robust mito-nuclear conflict was found for the position and monophyly of *G. giraffa* and *G. tippelskirchi*, which is interpreted as the result of a mitochondrial introgression from Masai to southeastern giraffe during the Pleistocene and nuclear gene flow mediated by male dispersal between southern populations (subspecies *G. g. giraffa* and *G. g. angolensis*).

LABX-0003-BCDiv, in the program "Investissements d'avenir" n ANR-11-IDEX-0004-02.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Biologically, speciation implies reproductive isolation through barriers preventing or limiting gene flow between populations [1]. Over the process of genetic differentiation, reproductively isolated populations may accumulate distinct phenotypic features that facilitate their recognition as different species. However, separated populations facing similar selective environments often converge phenotypically and show no visible differences (see Fišer et al. [2] for a review on cryptic species), which complicates their recognition as distinct species.

The development of DNA sequencing techniques over the last three decades allowed for studying species delimitation with molecular data in order to uncover cryptic taxa. Mitochondrial genes, and in particular the *COX1* gene (cytochrome *c* oxidase subunit 1), have been intensively used for species delimitation [3, 4]. However, numerous molecular studies have revealed that the mitochondrial tree may deviate from the species tree. Indeed, the maternal inheritance of the mtDNA genome can be misleading for species delimitation in mammals, because females and males have usually different dispersal behaviours (female philopatry versus male dispersal) [5, 6], and because interspecific hybrid females are generally fertile, whereas hybrid males are often sterile (Haldane's rule), facilitating mitochondrial introgression between closely related species [7–9]. To overcome these limitations, most recent taxonomic studies dealing with the delimitation between cryptic mammal species have focused on multi-locus datasets [10–12], as the use of multiple independent DNA markers has been shown to provide a strong and reliable signal for deciphering relationships among closely related taxa [13, 14]. However, interpreting the results from multi-locus datasets can be difficult, especially when the DNA markers show low genetic variation or conflicting relationships between them. These difficulties have led to the development of a plethora of new methodological approaches for multi-locus species delimitation [15, 16], which may be subdivided into three categories: (1) phylogenetic methods, (2) multispecies coalescent (MSC) approaches, and (3) population genetic methods (Table 1). Phylogenetic methods were not originally developed for studying species delimitation, but the species monophyly criterion has been widely used since the origin of molecular taxonomy [17]. For multi-locus datasets, several phylogenetic approaches can be considered: the concatenation of all markers into a supermatrix (although this approach has been widely criticized [18]), the separate analyses of the markers, or more sophisticated methods, such as *BEAST [19] or SuperTRI [20]. Based on the coalescent theory, some authors have suggested that species can be delimited without monophyletic gene trees [21]. The incorporation of the coalescent model [22] in certain software (e.g. *BEAST [19], BPP [23] and STACEY [24]) enabled the inference of species limits from multi-locus data by accounting for incongruences among gene trees in the presence of incomplete lineage sorting [19]. MSC approaches often require prior assignments of samples to populations or taxa and are hence restricted to the validation of proposed delimitations [25]. Population genetic approaches are generally applied to detect "cryptic substructure" between groups showing very similar phenotypes. The program STRUCTURE [26] is probably the most popular approach for Bayesian clustering using multi-locus data. It has recently gained new interest as the clusters identified with STRUCTURE can be used as preliminary hypothesis for assigning individuals to populations or taxa, which represents the first step of most MSC analyses [27]. In addition, geographic clusters detected with STRUCTURE are often interpreted (perhaps wrongly [28]) as reproductively isolated populations, which may constitute a strong argument in favour of a division into several species (e.g., Brown et al. [29]). Allele-sharing methods, such as haplowebs [30] and conspecificity matrix (CM) [31], can be also used to detect reproductively isolated populations.

The systematics of giraffes is a controversial issue, since at least nine different hypotheses of species delimitation were proposed on the basis of morphological characters and, more

Table 1. Species delimitation methods based on multi-locus nuDNA sequences used in this study.

Method	Input	Category	Reference	Description	SD Criteria
STRUCTURE	alignment of phased alleles	**PG**	Pritchard et al. [26] Falush et al. [53]	Bayesian clustering method based on the estimation of allele frequencies	ΔK and plateau methods
Haplowebs	alignment of phased alleles	**PG**	Flot et al. [30]	Allele-sharing method, in which groups of individuals sharing a unique pool of alleles are connected by curves.	Group of individuals constituting a field for recombination [FFR]
Conspecificity matrix (CM)	Conspecificity scores	**PG**	Debortoli et al. [31]	The conspecificity matrix was generated by calculating for each pair of individuals a conspecificity score, i.e., the number of markers supporting conspecificity in haploweb analyses.	Number of markers supporting the hypothesis of conspecificity
BPP (Bayesian Phylogenetics and Phylogeography)	alignment of consensus sequences	**MSC**	Yang and Rannala [64]	Bayesian method based on the MSC model, in which a reversible-jump Markov chain Monte Carlo algorithm is used to calculate the posterior probabilities of species delimitations.	Probability $\geq$ 0.95
STACEY (Species Tree and Classification Estimation, Yarely)	alignment of phased alleles	**MSC**	Jones [24] Individual assignment of alleles: present study	Bayesian method implemented in BEAST 2 [65] for the inference of a "species or minimal clusters tree" (SMC) under the birth-death-collapse tree prior and without the requirement of a guide tree.	Probability $\geq$ 0.95
***BEAST** (Species Tree Ancestral Reconstruction in BEAST)	alignment of phased alleles	**MSC + P**	Program: Heled and Drummond [19]; Individual assignment of alleles: present study	Bayesian method implemented in BEAST 2 [65] based on the MSC model	Probability $\geq$ 0.95
Bootstrap Analysis of Haplotypes	alignment of phased haplotypes	**PG+ + P**	Present study	Bootstrap consensus tree reconstructed with ML, MP or NJ methods	Bootstrap $\geq$ 90
Supermatrix MrBayes	alignment of consensus sequences	**P**	Ronquist et al. [46]	Bayesian inference of phylogeny	Probability $\geq$ 0.95
PhyML			Guindon et al. [47]	ML method for tree construction	Bootstrap $\geq$ 90
SuperTRI (SuperTree with Reliability Indices)	Weighted binary matrix of node support for each locus	**P**	Ropiquet et al. [20]	Three measures are calculated to estimate the reliability of the nodes (SBP, MPP and NRep) using the branch support values (PP) of all phylogenetic hypotheses produced during the separate Bayesian analyses of the 21 introns	(1) SBP $\geq$ 90 (2) MPP $\geq$ 0.1 (3) NRep $\geq$ 2

SD: Species delimitation; PG: Population Genetics; MSC: Multispecies Coalescent; P: Phylogenetic methods; SBP: Supertree Boostrap Percentage; MPP: Mean Posterior Probability

Rep: Reproducibility Index; SPR: Subtree Pruning and Re-grafting

https://doi.org/10.1371/journal.pone.0217956.t001

recently, molecular data (see Table A in S1 Appendix). The existence of several giraffe species was first proposed by Geoffroy Saint-Hilaire [32], who noted that differences in coat pattern, horn shape and skull can be used to distinguish the Nubian giraffe (from the Sennaar region in Sudan) from the Southern giraffe (from the Cape region). Thomas [33] proposed another arrangement in two species, in which Nubian and Southern giraffes were assigned to *Giraffa camelopardalis*, whereas the reticulated giraffe was treated as a full species, *Giraffa reticulata*. Lydekker [34] shared this view, but recognized 12 subspecies in *G. camelopardalis* and two in *G. reticulata*. However, Dagg and Foster [35] indicated that phenotypic features are highly variable between and within populations, and recognized therefore a single species, *G. camelopardalis*. Subsequently, this point of view was accepted by most other taxonomists, despite persisting controversy regarding the number of subspecies [36, 37]. However, the taxonomy of giraffes has been challenged by recent genetic studies: based on the analyses of mitochondrial sequences and 14 nuclear microsatellite loci, Brown et al. [29] proposed a minimum of six species, corresponding to *Giraffa angolensis*, *G. giraffa*, *G. peralta*, *G. reticulata*, *G. rothschildi*, and

G. tippelskirchi (N.B. the subspecies *camelopardalis*, *antiquorum* and *thornicrofti* were not included in their study); whereas Fennessy et al. [38] and Winter et al. [12] suggested a division into four species, i.e., *G. camelopardalis*, *G. giraffa*, *G. reticulata* and *G. tippelskirchi*, based on multi-locus analyses of 7 and 21 nuclear introns, respectively. However, the four-species hypothesis proposed by Fennessy et al. [38] has previously elicited concerns and controversy (see Bercovitch et al. [39]).

In this study, we reanalysed all multi-locus data available for the nine giraffe subspecies (i.e., *camelopardalis*, *angolensis*, *antiquorum*, *giraffa*, *peralta*, *reticulata*, *rothschildi*, *thornicrofti* and *tippelskirchi*) using various phylogenetic (MrBayes, PhyML, SuperTRI), population genetic (STRUCTURE, allelic networks, haplotype network and bootstrapping, haplowebs and conspecificity matrix) and MSC (*BEAST, BPP, STACEY) methods. Our intention was to provide sound scientific evidence about the number of species of *Giraffa* by comparing different methods currently used for molecular species delimitation that rely on different species concepts (Phylogenetic species concept [40], Genetic species concept [41] and Genealogical species concept [42]). Such a strategy is especially important for taxa of conservation concern like giraffes, since the application of a single species concept has been shown to "overlump" or "oversplit" species, which can entail negative consequences for conservation management [43, 44].

Our five main goals were (1) to test if the different methods converge towards the same conclusion or if they support divergent taxonomic hypotheses, (2) to examine if one hypothesis is more supported by the analyses than the others (comparative approach of species delimitation), (3) to understand why some methods or models can lead to taxonomic over-splitting, (4) to know if available molecular data are sufficient to conclude on the number of species, and (5) to determine which data, methods and operational criteria are relevant for delimiting species with molecular data.

Material and methods

Nuclear and mitochondrial datasets used for the analyses

Seven giraffe datasets were generated for our analyses using the sequences available in the NCBI nucleotide database:

1. the mtDNA-G507 dataset, which contains a mitochondrial fragment covering the whole cytochrome b (*Cytb*) gene and the 5'part of the control region (length = 1742 nucleotides [nt]) for 507 individuals (listed in Table A in S2 Appendix), and its reduced version including only the 82 different mitochondrial haplotypes, named mtDNA-GH82;
2. the mtDNA-GH82O3, in which the mtDNA-GH82 dataset was aligned to three outgroup species: *Bos taurus* (NCBI accession number KT184464); *Ovis canadensis* (NC_015889) and *Okapia johnstoni* (JN632674) (length = 1776 nt);
3. a nuclear dataset, named nuDNA-G274, including 274 phased alleles of 21 introns (*ACP5*, *C1orf74*, *CCT2*, *COL5A2*, *CTAGE5*, *CWF19L1; DDX1*, *DHX36*, *IGF2B1*, *MACF1*, *NOTCH2*, *NUP155*, *OTOF*, *PLCE1*, *RASSF4*, *RFRC5*, *SAP130*, *SOS1*, *UBN2*, *USP33*, *USP54)* for 137 giraffes (accession numbers LT596685-LT598170, MG257969–MG262280) (length = 16966 nt);
4. the nuDNA-G274O6, in which the nuDNA-G274 dataset was aligned to the alleles of three outgroup species: the okapi (*Okapia johnstoni*, published sequences [12, 38]), and two bovid species, (i) *Bos taurus*, for which the sequences were extracted by BLAST from the whole genome version UMD3.1.1 (http://bovinegenome.org/) or, in case of unavailability

of certain genes, from the genome of *Bos mutus* available on NCBI (SAMN08580377); and (ii) *Ovis canadensis*, for which the sequences were extracted by BLAST from the genome available on NCBI (CP011888.1) (length = 17276 nt);

5. the nuDNA-G137 dataset, comprising the alignments of original consensus sequences of 21 introns for the 137 giraffes (length = 16966 nt), which were recovered by detecting heterozygous sites in Geneious R10 (Biomatters, Auckland, New Zealand);
6. the nuDNA-G137O3 dataset, in which the nuDNA-G137 dataset was aligned to the three outgroup species mentioned above (length = 17276 nt);
7. the nuclear haplotype dataset, named nuDNA-GH274, which was inferred from the nuDNA-G280 dataset in DNASP v5.0 [45] using the "Generate Haplotype Data File" option under non-consideration of gaps/ missing data and subsequent exclusion of outgroup taxa (length = 1362 nt; it contains only the sites found to be variable between haplotypes).

The datasets were aligned automatically using the MAFFT algorithm implemented in Geneious R10 (Biomatters, Auckland, New Zealand) and subsequently verified by eye. All alignments generated for this study were deposited in Open Science Framework at https://osf.io/9wv86/.

Phylogenetic analyses

The mtDNA-GH82O3 and nuDNA-G137O3 datasets were analysed with probabilistic methods. Bayesian inferences were conducted in MrBayes v3.2.6 [46] by calculating the posterior probabilities (PP) after 10^7 Metropolis-coupled MCMC generations with tree sampling every 1000 generations and a burn-in of 25%. Maximum Likelihood (ML) analyses were performed with PhyML v3.1 [47] and Bootstrap percentages (BP) were calculated after 1000 replicates. The GTR+I+G substitution model was applied for both methods, as suggested by the Likelihood calculations in jModeltest [48] based on the Akaike information criterion [49].

Bayesian analyses were also performed for each of the 21 introns using the model of DNA substitution selected under jModeltest (Table 2).

SuperTRI analyses

The lists of bipartitions obtained from the Bayesian analyses (.parts and .tstat files) for each nuclear marker were transformed into a weighted binary matrix (MRP, matrix representation with parsimony) for supertree construction using SuperTRI v57 [20]. Here, each binary character corresponds to a node, which was weighted according to its frequency in one of the 21 lists of bipartition. Thereby, the SuperTRI method accounts for principal as well as secondary signals, given that all phylogenetic hypotheses found during the Bayesian analyses are represented in the weighted binary matrix used for supertree construction. The reliability of the nodes was assessed using three measures: supertree bootstrap percentages (SBPs) were obtained from PAUP* v4b10 [50] after 1000 BP replicates of the MRP matrix of 24749 binary characters generated by SuperTRI v57; mean posterior probabilities (MPP) and reproducibility indices (Rep) were directly calculated on SuperTRI v57. In the nuclear tree (Fig 1A), we chose to indicate the number of markers supporting each node of interest (NRep) rather than the Rep value, which represents the ratio of the number of markers supporting the node to the total number of markers [20].

STRUCTURE analyses

Giraffe haplotypes were reconstructed from the nuDNA-G137O3 dataset for each of the 21 introns by applying the PHASE v2.1 algorithm implemented in the software DNASP v5.0 [45],

Table 2. Characteristics of the nuclear alignments used for Bayesian phylogenetic analyses and posterior probabilities obtained for the different giraffe taxa.

Alignments	Substitution Model*	Length*	IS*	C	R	T	G	CR	CRG	CRT	GT	Th
ACP5	F81+G	640	9	-	0.03	-	0.04	0.75	-	0.52	-	-
C1orf74	F81	870	9	-	-	-	1.00	-	-	-	1.00	-
CCT2	HKY	812	6	-	-	-	-	-	-	-	-	-
COL5A2	HKY	878	7	-	-	0.96	-	-	-	-	0.04	-
CTAGE5	HKY	839	13	-	-	0.64	0.10	1.00	-	-	-	-
CWF19L1	HKY	668	5	-	-	-	-	1.00	-	-	-	-
DDX1	HKY	734	15	-	-	-	-	-	-	-	0.43	-
DHX36	HKY	815	11	-	-	-	0.98	-	-	1.00	-	-
IGF2B1	HKY+G	801	7	-	-	-	0.90	-	-	-	-	1.00
MACF1	HKY	718	9	-	-	-	-	-	-	-	-	-
NOTCH2	F81	854	5	-	-	-	0.50	-	-	-	-	-
NUP155	HKY	657	5	-	-	-	-	-	-	-	-	-
OTOF	K80	741	7	-	-	-	-	-	-	-	-	-
PLCE1	HKY	836	9	-	-	-	-	-	-	-	-	-
RASSF4	SYM+G	645	10	-	-	-	-	-	-	-	-	-
RFC5	HKY+G	825	9	-	-	0.75	-	-	0.70			-
SAP130	HKY+G	888	11	-	-	-	-	-	-	-	-	-
SOS1	F81	760	4	-	-	-	-	1.00	-	-	-	-
UBN2	F81	719	11	-	-	1.00	-	-	1.00	-	-	-
USP33	HKY+G	937	13	-	-	-	1.00	-	-	-	0.98	-
USP54	HKY	1329	12	-	-	-	-	-	-	-	0.98	-
nuDNA	GTR+I+G	16966	187	1.00	1.00	1.00	1.00	1.00	-	-	0.82	-

IS: Informative sites for parsimony within *Giraffa*; C: *G. camelopardalis* sensu stricto B (subspecies *camelopardalis*, *antiquorum*, *rothschildi*, *peralta*); R: *G. reticulata*; T: *G. tippelskirchi* (subspecies *tippelskirchi* and *thornicrofti*); G: *G. giraffa*; Th: *G. thornicrofti*; "-": not found; "*": outgroups excluded.

https://doi.org/10.1371/journal.pone.0217956.t002

allowing for recombination and reducing the output probability threshold of conserved regions (CT) from 0.9 by default to 0.6.

Bayesian analyses of genetic admixture were run in STRUCTURE v.2.3.4 [26] to identify genetically homogeneous groups of individuals (populations of origin, K). The haplotype information deduced by PHASE for each of the 21 introns was used to code individuals sharing the same genotype with a unique integer in the input file [51]. The analyses were done as recommended by Gilbert et al. [52], i.e., number of MCMC generations = 200 000 and burn-in = 100 000 generations for K = 1–10 clusters. We applied several combinations of ancestry model, allele frequency and supporting information (Popdata) like the assignment of the subspecies (population identity/ POPID) or sampling location (LOCPRIOR model) for each individual. We tested two ancestry models, since we do not know whether studied populations were discrete or had an admixed ancestry. Moreover, the identification of the most probable number of clusters (K) might be further affected by the choice of the allele frequency model. By default, the software assumes correlated allele frequency among populations caused by migration and shared ancestry [53]. Since past admixture was expected between giraffe populations, this model may represent the appropriate choice. However, several runs were conducted under the independent allele frequency model, as it might be more powerful to detect highly distinct populations [54].

We also tested two settings for lambda (λ), the parameter specifying the distribution of allelic frequencies in each population: the default setting (λ = 1) and an estimated value of λ (λ = 0.45), calculated during a run comprising 20 iterations for K = 1. Runs were performed

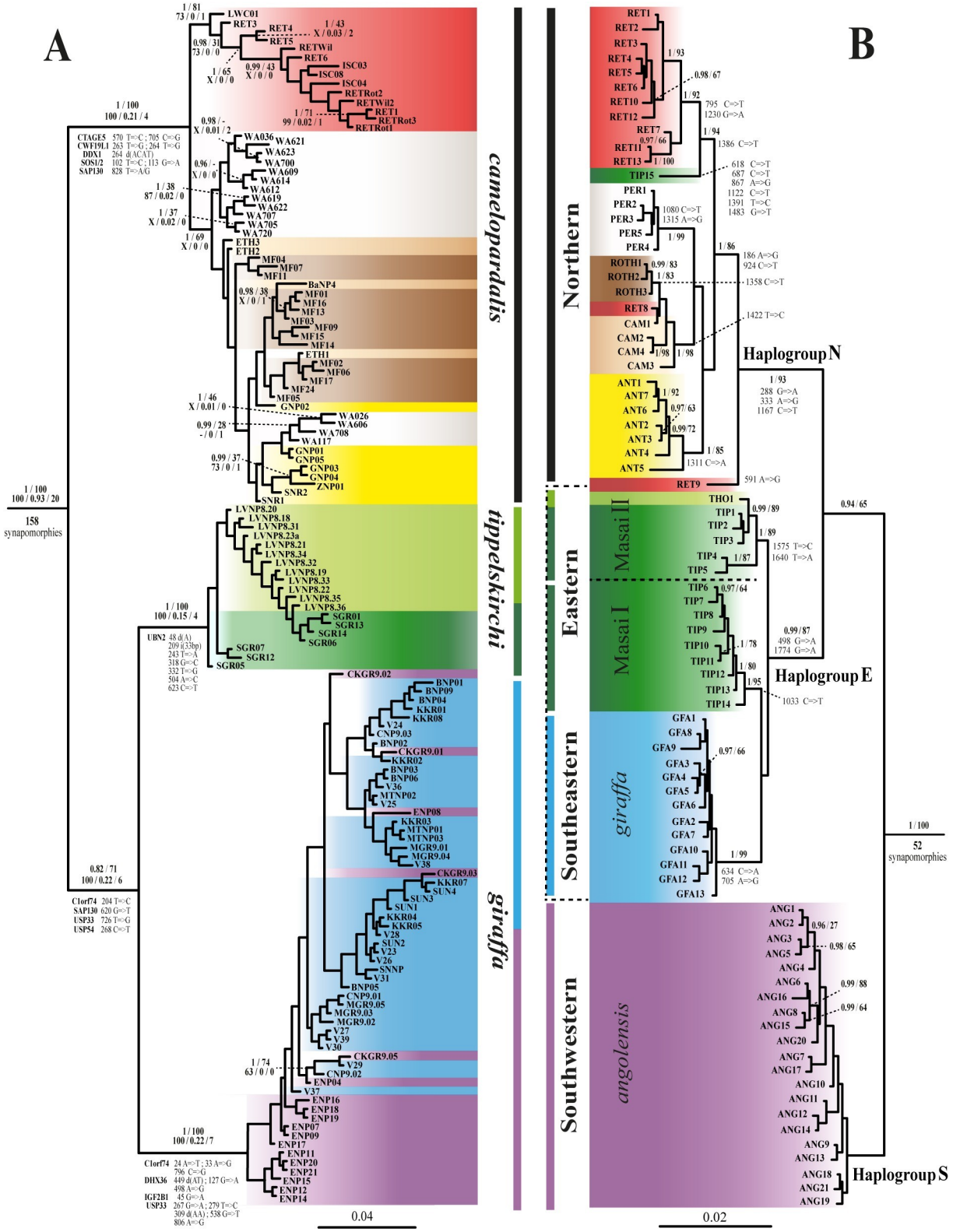


Fig 1. Comparative phylogeny of nuclear and mitochondrial datasets. The nine subspecies are differentiated by the following colours: red: *reticulata*, white: *peralta*, brown: *rothschildi*, beige: *camelopardalis*, yellow: *antiquorum*, blue: *giraffa*, purple: *angolensis*, light green: *thornicrofti* and dark green: *tippelskirchi*. The three outgroup species are not shown. (**A**) Bayesian tree inferred from the nuclear dataset, named nuDNA-G137O3, including the sequences of 21 introns for 137 giraffes. The tree was rooted with *Bos*, *Ovis*, and *Okapia* (not shown). For each

node recovered with significant support in the Bayesian analysis (PP ≥ 0.9), as well as for other nodes discussed in the text, the two values above indicate the Posterior Probability with MrBayes (PP) and the Bootstrap Percentage obtained from the Maximum Likelihood analysis (BP). The three values below were obtained from the SuperTRI analyses of the 21 introns: from left to right: Supertree Bootstrap Percentage (SBP), Mean Posterior Probability (MPP) and the number of markers supporting the node (NRep). The symbol "–" indicates that the node was not found monophyletic in the analysis, and the letter "X" indicates that an alternative hypothesis was supported by SBP > 50. The exclusive synapomorphies (including indels; i: insertion; d: deletion), representing fixed substitutions among members of a group, are listed for the nodes discussed in the text. (**B**) Bayesian tree of the 82 mitochondrial haplotypes detected for *Giraffa* reconstructed from a fragment covering the complete *Cytb* gene and the 5' part of the control region (1776 characters) and rooted with *Bos*, *Ovis* and *Okapia* (not shown). For each node supported by PP ≥ 0.95, the BP value obtained from the Maximum Likelihood analysis is indicated. Fixed substitutions among members of a group (exclusive synapomorphies) are listed for the nodes supported by PP ≥ 0.95 and for uncommon mitochondrial haplotypes.

https://doi.org/10.1371/journal.pone.0217956.g001

without any assignation of individuals, or by assigning individuals to either a POPID representing the designated subspecies or to their sampling location (LOCPRIOR, national parks where the giraffes were sampled; see Table B in S2 Appendix), as this option is recommended when only a weak signal is present in the markers [55]. All analyses were replicated 20 times.

The most likely number of distinct groups for each run was identified by means of STRUCTURE HARVESTER [56]. Thereby, the optimal K was determined using two approaches: (1) the ΔK method of Evanno et al. [57], which recognizes the most likely number of distinct clusters by the largest ΔK value, calculated by the rate of change in the log probability of data between successive K values; and (2) the "plateau "method of Pritchard et al. [51], where the log probability of the data (ln Pr (X|K) was plotted against a range of K values, and the optimal K was selected as the point at which the plot curvature plateaus. A regression curve and gridlines were added to the diagrams generated by STRUCTURE HARVESTER to help in determining the point of plateau.

To assess the reliability of the results, CLUMPAK [58] was used to display the barplots from K = 1 to 10 for each of the 20 iterations by means of the implemented software DISTRUCT [59].

Analyses of nuclear haplotypes

The nuDNA-GH274 dataset (nuclear haplotypes inferred from the concatenated matrix of 21 introns and 137 giraffes) was used to construct a median-joining network [60] using PopART v1.7 [61]. The robustness of haplotype clusters was evaluated by bootstrapping (1000 replicates) under PAUP* v4b10 [50] using either the Maximum Parsimony (MP) method (heuristic search, faststep option) or the Neighbor-Joining (NJ) method (GTR+I+G model), and under the Maximum Likelihood (ML) criterion using RAxML on CIPRES [62] (http://www.phylo.org). PopART was also used to construct a median-joining network for each of the 21 introns. For six introns (i.e. *CTAGE5*, *NUP155*, *OTOF*, *PLCE1*, *RASSF4* and *SOS1*), individuals with missing allele(s) were not considered in the analysis in order to avoid any distortion of the results.

Allele sharing among individuals was investigated for each of the 21 introns using the haploweb approach [30] by building raw haplowebs with the online tool HaplowebMaker (https://eeg-ebe.github.io/HaplowebMaker/). Haplowebs were constructed without the consideration of singletons (alleles encountered only once in the whole dataset, "remove singletons" option) and default settings for the other parameters. In the networks, curves are illustrating alleles found co-occurring in heterozygous individuals. Thereby, a group of alleles linked together by heterozygotes represents an exclusive allele pool, the corresponding groups of individuals is called a field for recombination (FFR) [63]. Based on the allele-sharing information per marker provided by HaplowebMaker, a conspecificity matrix [31] was built using the online tool CoMa (https://eeg197ebe.github.io/CoMa/) with no calculation on heterospecific pairs (option 1), i.e. no value was given to absence of sharing. In this matrix, the conspecificity score for

each pair of individuals is the number of markers for which these individuals belong to the same FFR. The conspecificity matrix was illustrated as a heatmap using the hierarchical clustering method implemented in the BioVinci data visualization software (BioTuring, San Diego, CA) and the UPGMA method for the associated dendrogram.

Multispecies coalescent analyses

Three coalescent-based approaches were applied to infer species boundaries within the genus *Giraffa*: (1) the "Species Tree Ancestral Reconstruction" template (*BEAST [19]), (2) the extension of the *BEAST model called "Species Tree and Classification Estimation, Yarely" (STACEY) [24], and (3) the Bayesian Phylogenetics and Phylogeography program (BPP v.3.2 [23, 64]) (see specifications for each program in Table 1).

We estimated the species-tree phylogeny using the coalescent algorithm implemented in BEAST v.2.4.4 [65] in order to consider an alternative to the traditional concatenated phylogenetic approach (see Kubatko and Degnan [66] for caveats concerning concatenation). Inferences were based on the nuDNA-G274O6 dataset using an *a priori* assignment at the level of individuals, i.e. by assigning for each of the 137 giraffes two alleles for the 21 introns. We assumed an uncorrelated lognormal molecular clock for all 21 loci. For each marker, we selected the best suited substitution model inferred in jModeltest [48] (Table 2). Analyses were run with 2x 10^8 generations, with trees sampled every 5000 steps. The .log files were analysed with Tracer v1.7 [67] to assess the convergence of model parameters (effective sample size [ESS] $>$ 200). The species tree was summarized as a Maximum Clade Credibility tree in TreeAnnotator v.1.10 [68] after discarding 25% as burn-in.

The nuDNA-G274 and nuDNA-G274O6 datasets were further used for species delimitation analyses using the STACEY template implemented in BEAST v2.4.4. STACEY represents an improvement to the DISSECT model [69] to infer a "species or minimal clusters tree" (SMC) under the birth-death-collapse tree prior and without the requirement of a guide tree. The tips of the SMC tree represent minimal clusters of individuals that may be collapsed to a single putative species, if branches are shorter than a specified length (collapse height) [24]. A first run was conducted without taxonomic a priori assumptions by assigning two alleles per gene and individual. For the other run, each individual was assigned to one of the six taxa (6S hypothesis) that were found monophyletic with at least one of our phylogenetic analyses: *G. camelopardalis* sensu stricto C (including only the three subspecies *camelopardalis*, *antiquorum* and *rothschildi*), *G. peralta*, *G. reticulata*, *G. giraffa* (including the two subspecies *angolensis* and *giraffa*), *G. tippelskirchi* sensu stricto and *G. thornicrofti*. Analyses were done as suggested in the manual, i.e. using a relative death rate of 0.0 for the tree prior, a lognormal distribution with a mean of 4.6 and a standard deviation of 2 to the growth rate prior and a uniform distribution for the relative death rate prior with a lower bound of -0.5 and an upper bound of 0.5. The dataset was partitioned by the 21 genes, with independent strict clock models and individual assignment of the best suited substitution model to each gene (Table 2). Each analysis was run for 2.5 x 10^8 generations and convergence of parameters was assessed in Tracer v1.7 [67]. Subsequently, the most supported number of distinct clusters was estimated using SpeciesDelimitationAnalyser v1.8.0 [24] by analysing the species trees with a burn-in of 25% and the default collapse height of 0.0001.

Species delimitation analyses with BPP v3.2 were based on a reduced dataset comprising only 66 giraffes due to software limitation. Seventy-one individuals were excluded from the original dataset using the three following criteria: (1) 14 individuals with missing data, (2) 39 individuals sharing the same haplotype and (3) 18 individuals characterized by a long terminal branch in the Bayesian tree. We analysed the support for five taxonomic hypotheses: two

species, with two possible geographic patterns (2Sa and 2Sb hypotheses), three species (3S hypothesis), i.e. *G. camelopardalis* sensu stricto A, *G. giraffa* and *G. tippelskirchi*, four species (4S hypothesis), i.e. *G. camelopardalis* sensu stricto B, *G. giraffa*, *G. reticulata*, and *G. tippelskirchi*, and five species (5S hypothesis), i.e. *G. camelopardalis* sensu stricto C, *G. giraffa*, *G. peralta*, *G. reticulata*, and *G. tippelskirchi* (see results for more details). First, we applied the A00 algorithm, the simple MSC model with the species tree fixed to explain the acceptance proportions of MCMC moves [23] under the default gamma prior values *G* (2, 2000) for the tau (τ, root divergence time) and theta (θ, genetic difference among taxa). Then, we assessed the support for each putative species using the A11 algorithm [64]. The three species model priors (SMP 1, 2 and 3) were tested. The analyses were run for 500 000 generations followed by a burn-in of 10%. Convergence between runs was checked for fine tune acceptance proportions between 0.15 and 0.7, as well as ESS $>$ 200.

Nuclear and mitochondrial pairwise distances

The nuDNA-G137 dataset (16966 nt) and mtDNA-GH82 dataset (1742 nt) were used to calculate pairwise distances in PAUP* v4b10 [50] (see Tables A and B in S4 Appendix). For the nuclear dataset, we performed calculations considering the five taxonomic hypotheses mentioned above. For the mtDNA dataset, we primarily performed calculations based on the three main mitochondrial haplogroups, named N, E and S depicted in Fig 1B, but considered also the five possible hypotheses of species delimitation described in the Multispecies coalescent analyses section.

Results

Phylogenetic analyses of the nuclear dataset

The 21 nuclear introns were analysed independently and in combination. The phylogenetic trees obtained from the separate analyses of the 21 independent introns are detailed in S2 Appendix and the Bayesian nuclear tree of the concatenated dataset (17276 nt) is depicted in Fig 1A. The results of other analyses (ML bootstrap [BP] and SuperTRI indices [SBP/MPP/NRep]) are indicated only for the nodes supported by posterior probability (PP) values $\geq$ 0.9, as well as for nodes discussed in the text (e.g., subspecies).

The monophyly of *Giraffa* is supported by all analyses and almost all markers separately (NRep = 20), and the genus is diagnosed by 158 exclusive synapomorphies in the nuclear genes. Within *Giraffa*, 19 nodes are supported by PP $\geq$ 0.9 in the Bayesian tree of the nuDNA supermatrix (Fig 1A; Fig Y in S2 Appendix), but only three of them are associated with BP $>$ 90: (1) the clade here named *G. camelopardalis* sensu stricto A, which groups together all members of the subspecies *camelopardalis*, *antiquorum*, *peralta*, *reticulata*, and *rothschildi* (PP = 1; BP = 100); (2) *G. giraffa*, including all members of the subspecies *angolensis* and *giraffa* (PP = 1; BP = 100); and (3) *G. tippelskirchi*, comprising all members of the subspecies *thornicrofti* and *tippelskirchi* (PP = 1; BP = 100). The monophyly of other taxa was less supported in the ML analysis: BP = 69 for *G. camelopardalis* sensu stricto B (*G. camelopardalis* s.s. A excluding *reticulata*) and BP = 81 for *G. reticulata*.

The results of separate analyses of the 21 introns showed that none of them supports the monophyly of *G. camelopardalis* s.s. B and that *G. reticulata* is found monophyletic only for *ACP5*, but with insignificant support (PP = 0.03). By contrast, *G. tippelskirchi* is independently supported by four genes: *COL5A2* (PP = 0.96), *CTAGE5* (PP = 0.64), *RFC5* (PP = 0.75) and *UBN2* (PP = 1); and all individuals of this taxon share seven molecular signatures in the *UBN2* gene (Fig 1A). The taxa corresponding to *G. camelopardalis* s.s. A and *G. giraffa* are the most robust and reliable nodes within *Giraffa* (Fig 1A, Fig Y in S2 Appendix): *G. camelopardalis* s.s.

A is supported by the separate analyses of 4 introns, i.e. *ACP5* (PP = 0.75), *CTAGE5* (PP = 1), *CWF19L1* (PP = 1) and *SOS1* (PP = 1), and members of this group share eight molecular signatures detected in five markers; *G. giraffa* is found monophyletic with PP $\geq$ 0.5 in the separate analyses of 5 introns, i.e. *C1orf74* (PP = 1), *DHX36* (PP = 0.98), *IGF2B1* (PP = 0.9), *NOTCH2* (PP = 0.5) and *USP33* (PP = 1), and members of this group share 12 molecular signatures detected in four markers.

SuperTRI analyses

The SuperTRI analyses of the 21 introns are highly informative for relationships within *Giraffa* (S2 Appendix). Indeed, only six nodes are supported by MPP > 0.1 and NRep $\geq$ 2 (Fig 1A): *Giraffa* + *Okapia* (MPP = 1; NRep = 21); *Giraffa* (MPP = 0.93; NRep = 20); *G. giraffa* + *G. tippelskirchi* (MPP = 0.22; NRep = 6); *G. giraffa* (MPP = 0.22; NRep = 7); *G. camelopardalis* s.s. A (MPP = 0.21; NRep = 4); and *G. tippelskirchi* (MPP = 0.15; NRep = 4). All these nodes are also characterized by several exclusive synapomorphies detailed in Fig 1A. By contrast, SuperTRI analyses did not provide support for the two other taxa: *G. camelopardalis* s.s. B (MPP/NRep = 0) and *G. reticulata* (MPP = 0; NRep = 1). Particularly relevant is the fact that SuperTRI results also show no support (i.e. MPP $\leq$ 0.05 and NRep $\leq$ 1; Figs Aa and Ab in S2 Appendix) for all interpopulational or interindividual relationships within the three species *G. camelopardalis* s.s. A, *G. giraffa* and *G. tippelskirchi*.

STRUCTURE analyses

Our Bayesian population structure analyses were carried out on alleles inferred for 21 introns and 137 giraffes (0.5% of missing data). We tested different models (admixture versus no admixture, independent versus correlated allele frequency), with and without supporting priors on the subspecies (POPID) or on the geographic origins of the individuals (LOCPRIOR), as well as two values of lambda, fixed ($\lambda = 1$) or estimated ($\lambda = 0.45$) (Table 3). For each run, the most likely number of distinct groups (K) was determined using both ΔK and "plateau" methods [57, 51].

Using the ΔK method of Evanno et al. [57], 58% of the STRUCTURE analyses (14 / 24) resulted in the highest ΔK value for the separation into two clusters (K) corresponding to a North/ South dichotomy and the comparisons between DISTRUCT barplots indicated differences in the affiliation of both *tippelskirchi* and *thornicrofti* giraffes to either the northern or the southern group (Table 3; 2Sa and 2Sb hypotheses). The highest ΔK value for three distinct clusters was obtained for 25% (6 / 24) of the analyses (Table 3), supporting the 3S hypothesis. Finally, the separation into four K clusters was supported by four analyses (17%, Table 3).

Using the "plateau" method of [51], we found that K = 3 is the most probable number of clusters for 12 STRUCTURE HARVESTER diagrams (50% of the 24 analyses), whereas the highest support for four clusters could only be found in 8% of the analyses (2 / 24) (S3 Appendix). For other diagrams, it was difficult to determine at which K the plateau is reached: for 29% of the analyses (7 / 24), it was not possible to choose between K = 3 and 4; for 8% of the analyses (2 / 24) it was not possible to choose between K = 2 or 3; and for 4% of the analyses (1 / 24) it was not possible to choose between K = 2, 3 or 4.

Analyses of nuclear haplotypes

The haplotype network and bootstrap values obtained from the ML, MP and NJ analyses of the 274 nuclear haplotypes of 137 individuals are shown in Fig 2. All analyses support a division into three divergent haplogroups (separated by a minimum of 36 mutations) corresponding to (1) *G. camelopardalis* s.s. A ($BP_{MP/NJ/ML}$ = 71/100/99), which includes the subspecies *camelopardalis*, *antiquorum*, *rothschildi*, *reticulata* and *peralta;* (2) *G. tippelskirchi* ($BP_{MP/NJ/ML}$ = 100), which

Table 3. STRUCTURE analyses based on 21 introns and associated ΔK values (highest in bold) calculated using the method of Evanno et al. [57], as well as the optimal K value(s) deduced from the "plateau" method of Pritchard et al. [51] (underlined) (Our conclusions based on the results of both methods are highlighted in grey).

Ancestry Model	Popdata	Allele Frequency	ΔK K = 2 2Sa/b hypotheses		ΔK K = 3 3S hypothesis		ΔK K = 4 4S hypothesis	
			λ = 1.0	λ = 0.45	λ = 1.0	λ = 0.45	λ = 1.0	λ = 0.45
Admixture	-	correlated	31.5 [a,b]	27.3 [b]	2.2	**116.5**	**49.7**	18.3
Admixture	POPID	correlated	28.7 [b]	29.9 [a, b]	**1736.4**	4.24	17.1	**691.2**
Admixture	LOCPRIOR	correlated	30.6 [a]	31.0 [b]	**614.4**	**465.9**	10.9	81.3
Admixture	-	independent	24.8 [a]	**27.1** [b]	4.6	4.4	**154.9**	14.7
Admixture	POPID	independent	**25.1** [b]	**26.6** [a]	4.8	4.5	0.8	5.9
Admixture	LOCPRIOR	independent	**28.8** [a, b]	**28.8** [b]	1.7	3.7	13.2	13.7
No Admixture	-	correlated	**12.7** [a, b]	**14.7** [a]	0.3	0.3	0.3	2.3
No Admixture	POPID	correlated	28.0 [b]	26.7 [a]	**1556.3**	4.2	797.8	**703.9**
No Admixture	LOCPRIOR	correlated	**30.8** [a]	**53.0** [a]	1.3	1.1	4.4	3.1
No Admixture	-	independent	24.0 [a]	**25.8** [a]	**585.9**	2.3	14	22.9
No Admixture	POPID	independent	**25.5** [b]	**25.3** [b]	4.4	3.0	5.2	11.7
No Admixture	LOCPRIOR	independent	**28.5** [a]	**27.8** [a b]	1.0	1.2	2.6	2.3

POPID = subspecies assignment for each individual; LOCPRIOR = consideration of sampling location; "XX": K = 2 or 3; "XX": K = 3; "XX": K = 4; "XX": K = 3 or 4; "XX": K = 2, 3 or 4

[a]: affiliation of the subspecies *tippelskirchi* and *thornicrofti* to *G. camelopardalis* (2Sa hypothesis)

[b]: affiliation of the subspecies *tippelskirchi* and *thornicrofti* to *G. giraffa* (2Sb hypothesis).

https://doi.org/10.1371/journal.pone.0217956.t003

includes the subspecies *tippelskirchi* and *thornicrofti*, and (3) *G. giraffa* ($BP_{MP/NJ/ML}$ = 100) containing the southern subspecies *giraffa* and *angolensis*.

The haplotype network shows a separation between *reticulata* and other subspecies of *G. camelopardalis* s.s. A (a taxon named *G. camelopardalis* s.s. B), as well as a separation between the two subspecies of *G. tippelskirchi*, i.e. *tippelskirchi and thornicrofti*. None of these additional clusters are however supported by $BP_{MP/NJ/ML} > 50$, except *G. camelopardalis* s.s. B (BP_{ML} = 72) and *thornicrofti* (BP_{ML} = 54) in the RAxML analysis. By contrast, no subspecies can be distinguished within *G. giraffa*.

The haplotype networks constructed for each of the 21 nuclear introns are shown in Fig 3. Only five taxa show allelic clustering: (1) *G. camelopardalis* s.s. A and (2) the group *G. giraffa* + *G. tippelskirchi* in seven networks (*C1orf74*, *CTAGE5*, *CWF19L1*, *SAP130*, *SOS1*, *USP33* and *USP54*); (3) *G. giraffa* in six networks (*ACP5*, *C1orf74*, *DHX36*, *IGF2B1*, *RFC5*, and *USP33*); (4) *G. tippelskirchi* in six networks (*C1orf74*, *COL5A2*, *CTAGE5*, *RFC5*, *UBN2*, and *USP33*); and (5) *thornicrofti* in one network (*IGF2B1*).

We detected incomplete clustering (i.e., 1–3 "foreign" alleles in the cluster, or less than three alleles not included into the cluster) for the following taxa: *G. camelopardalis* s.s. A (*ACP5*: one *thornicrofti* allele; *DDX1* and *RFC5*: two alleles outside); *G. giraffa* (*DDX1*: two *reticulata* alleles; *NOTCH2*: one *tippelskirchi* allele; *SOS1*: two alleles outside; *USP54*: two *thornicrofti* alleles); *G. tippelskirchi* (*ACP5*: one allele outside; *SOS1*: two *giraffa* alleles); *G. camelopardalis* s.s. B (*USP54*: three *reticulata* alleles); and *G. reticulata* (*ACP5* and *USP54*: three alleles outside). The patterns found for the six other introns (*CCT2*, *MACF1*, *NUP155*, *OTOF*, *PLCE1*, *RASSF4*) do not fit any of the tested taxonomic hypotheses.

The haplowebs constructed for each of the 21 nuclear introns are shown in Fig 4. For the six introns *CCT2*, *MACF1*, *NUP155*, *OTOF*, *PLCE1*, and *RASSF4*, the co-occurrence of alleles

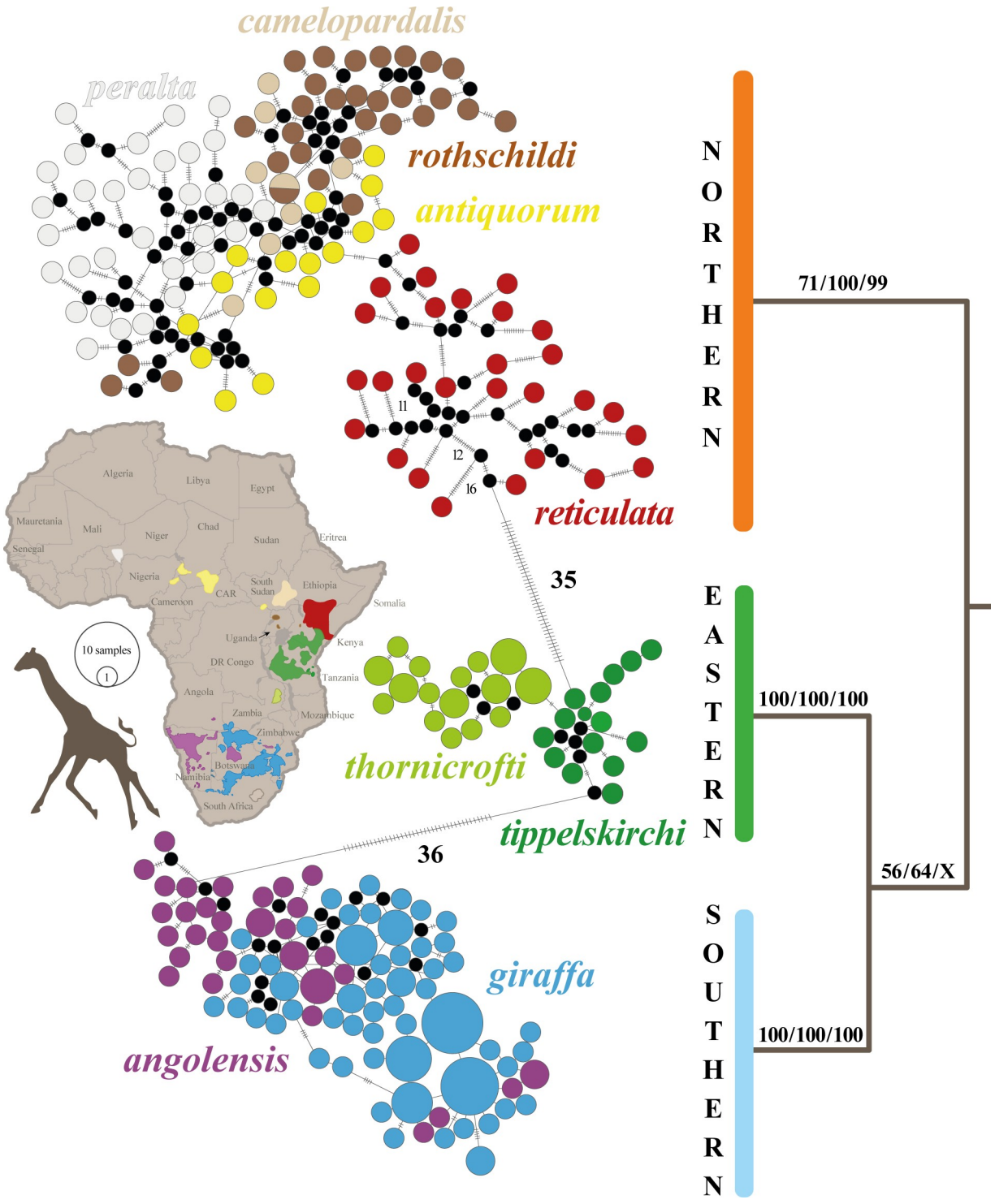


Fig 2. Current distribution of giraffe subspecies and population genetic analyses of nuclear haplotypes. The nine subspecies currently recognized are distinguished by different colours on the map (modified from https://giraffeconservation.org/giraffe-species/). At the left, the median-joining network was constructed under PopART based on 274 nuclear haplotypes for 137 giraffes. The black circles correspond to unsampled haplotypes and the number of mutations between haplotypes are indicated on the branches. At the right, the 50% majority-rule bootstrap consensus tree was reconstructed under PAUP using the nuDNA-G274O6 dataset (see Material and Methods for more details). The values at the nodes represent Bootstrap percentages $\geq$ 50 calculated with maximum parsimony, distance and maximum likelihood methods (from left to right). Relationships within subspecies are not shown here.

https://doi.org/10.1371/journal.pone.0217956.g002

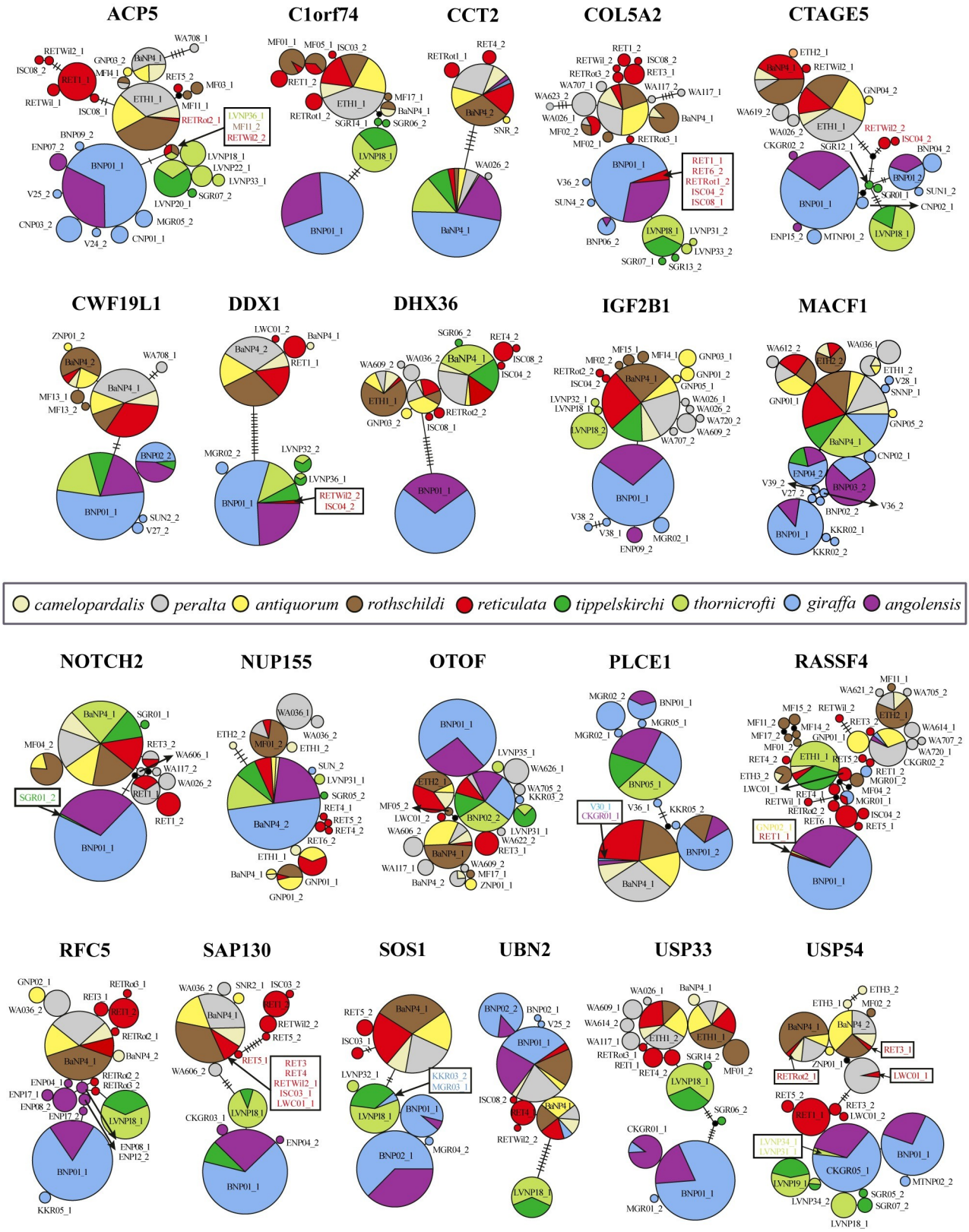


Fig 3. Allelic networks for 21 nuclear introns. The circles represent alleles with sizes proportional to their frequency in the populations. Each allele is designated with one representative individual (the list of all individuals is provided in Tables A-U in S6 Appendix). The nine subspecies currently

recognized are distinguished by different colours. Individuals characterized by a rare allele (in the subspecies) are highlighted with a black frame. The numbers of mutations between alleles are indicated on the branches.

https://doi.org/10.1371/journal.pone.0217956.g003

does not corroborate any taxonomic hypotheses tested within this study. However, other introns provided some support for the three following species: (1) *G. camelopardalis* s.s. A, which is recovered in nine haplowebs (*C1orf74*, *CTAGE5*, *CWF19L1*, *DDX1*, *RFC5*, *SAP130*, *SOS1*, *UBN2* and *USP33*); (2) *G. giraffa*, which is also found in nine haplowebs (*ACP5*, *C1orf74*, *CTAGE5*, *DHX36*, *IGF2B1*, *NOTCH2*, *RFC5*, *UBN2* and *USP33*); and (3) *G. tippelskirchi*, which is shown in six haplowebs (*C1orf74*, *COL5A2*, *CTAGE5*, *RFC5*, *UBN2*, and *USP33*). In addition, *G. tippelskirchi* is grouped to *G. giraffa* in five haplowebs (*CWF19L1*, *DDX1*, *SAP130*, *SOS1* and *USP54*) or to *G. camelopardalis* s.s. A in two haplowebs (*DHX36* and *NOTCH2*).

The conspecificity matrix built from the conspecificity scores obtained through the construction of the haploweb for each of the 21 introns (Fig 4; Tables V-Ap in S6 Appendix) is shown in Fig 5. Here, the number of independent markers supporting the hypothesis of the conspecificity for a respective pair of individuals is visualized by various nuances of red: the highest score (21 out of 21) is shown in dark red, whereas the lowest score (0 of 21) is illustrated in white. The conspecificity matrix depicts three dark red rectangles corresponding to the three species *G. camelopardalis* s.s. A, *G. giraffa* and *G. tippelskirchi* (3S hypothesis).

Multispecies coalescent analyses

We constructed a MSC species-tree from the nuDNA-G274-O6 dataset using *BEAST. The topology is similar to the supermatrix topology of Fig 1A, with maximal support (PP = 1) for *G. camelopardalis* s.s. A, *G. giraffa* and *G. tippelskirchi*. However, the monophyly of *G. camelopardalis* s.s. B, *G. camelopardalis* s.s. C and four subspecies (*antiquorum*, *peralta*, *reticulata*, and *thornicrofti*) was also highly supported (PP = 1) in the MSC tree (Fig A in S5 Appendix). The subspecies *tippelskirchi* was found monophyletic, but with low PP support (= 0.39).

The analyses based on STACEY showed highest support for five distinct giraffe species, i.e., *G. camelopardalis* s.s. C, *G. giraffa*, *G. peralta*, *G. reticulata* and *G. tippelskirchi*, a pattern found in 87% of the trees. Other hypotheses of species delimitation were less supported: the 4S hypothesis (*G. camelopardalis* s.s. B, *G. giraffa*, *G. reticulata* and *G. tippelskirchi*) was found in 7% of the trees; whereas the 6S hypothesis, which recognizes *G. camelopardalis* s.s. C, *G. giraffa*, *G. peralta*, *G. reticulata*, *G. tippelskirchi* sensu stricto, and *G. thornicrofti*, was found in 6% of the trees. Similar results were obtained when outgroup sequences were excluded (data not shown).

Species delimitation analyses based on BPP provided maximal support (PP = 1) for all species recognized according to the 3S, 4S, and 5S hypotheses (depicted in Fig 6). The same results were found with the three species model priors (Table A in S5 Appendix). The further division of *G. tippelskirchi* into two separate taxa, i.e. *G. tippelskirchi* sensu stricto and *G. thornicrofti* (6S hypothesis) was only weakly supported ($PP_{SMP1} = 0.26$; $PP_{SMP2} = 0.4$; $PP_{SMP3} = 0.34$).

Phylogenetic analyses of the mitochondrial fragment

The Bayesian tree reconstructed from the mtDNA-GH82O3 dataset (1776 nt) is shown in Fig 1B. It shows the existence of three major geographic haplogroups: northern (N), eastern + southeastern (E), and southwestern (S) giraffes.

The N haplogroup is supported by both Bayesian and bootstrap analyses (PP = 1; BP = 93). It includes all haplotypes detected for *G. camelopardalis* s.s. A, as well as one divergent

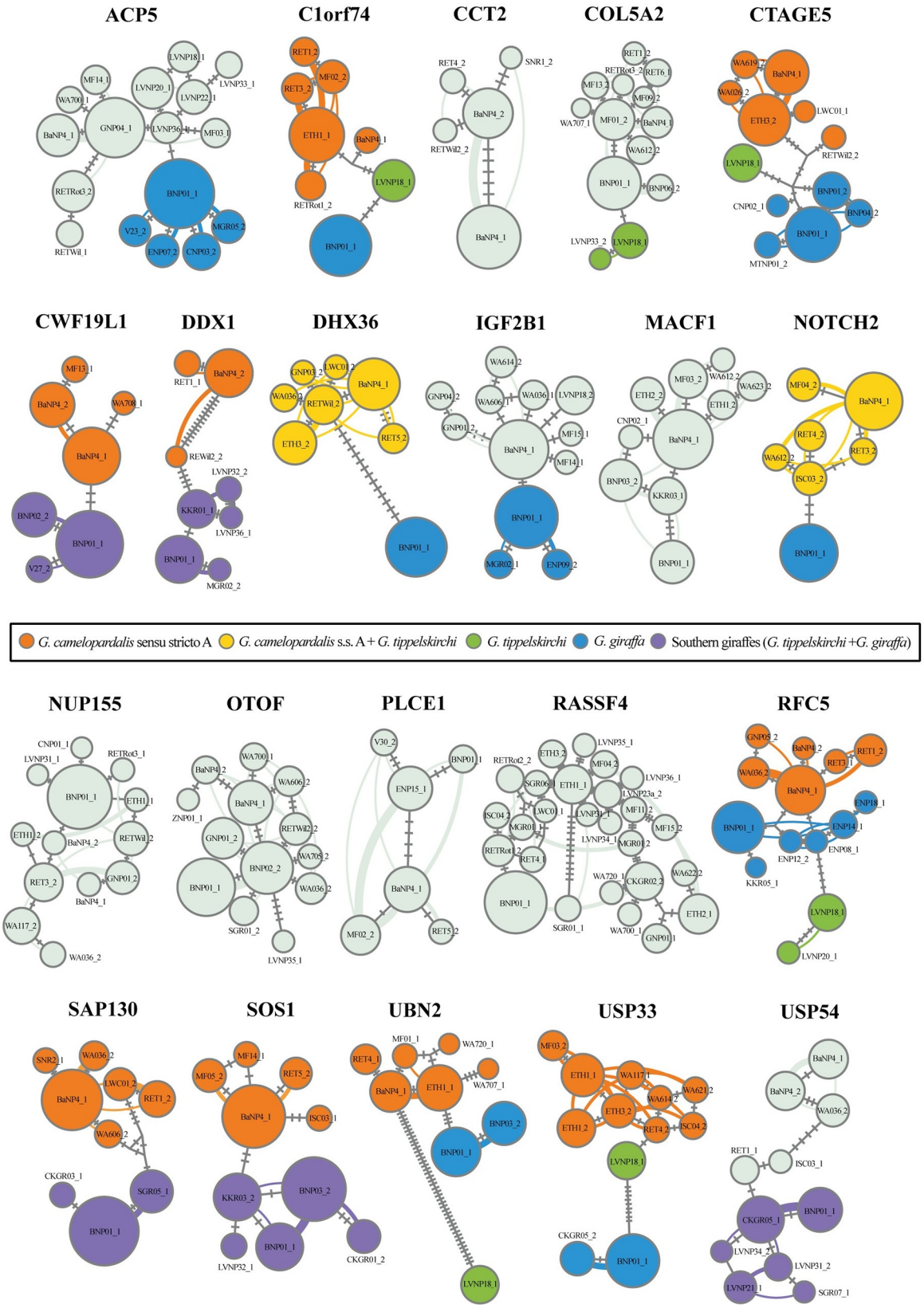

ACP5
C1orf74
CCT2
COL5A2
CTAGE5
CWF19L1
DDX1
DHX36
IGF2B1
MACF1
NOTCH2
G. camelopardalis sensu stricto A
G. camelopardalis s.s. A + G. tippelskirchi
G. tippelskirchi
G. giraffa
Southern giraffes (G. tippelskirchi + G. giraffa)
NUP155
OTOF
PLCE1
RASSF4
RFC5
SAP130
SOS1
UBN2
USP33
USP54

Fig 4. Haplowebs for 21 nuclear introns. The circles represent alleles with sizes proportional to their frequency in the populations. Each allele is designated with one representative individual (the list of all individuals is provided in Tables V-Ap in S6 Appendix). The investigated taxonomic hypotheses are distinguished by different colours. The numbers of mutations between alleles are indicated on the branches. The curves connect individuals sharing one unique pool of alleles.

https://doi.org/10.1371/journal.pone.0217956.g004

haplotype of *G. tippelskirchi* (TIP15, EU088334) sequenced by Brown et al. [29] for nine individuals from Kenya (Athi River Ranch) (see details in Table A in S2 Appendix). Three subspecies of *G. camelopardalis* are monophyletic: *antiquorum* (PP = 1; BP = 85), *peralta* (PP = 1;

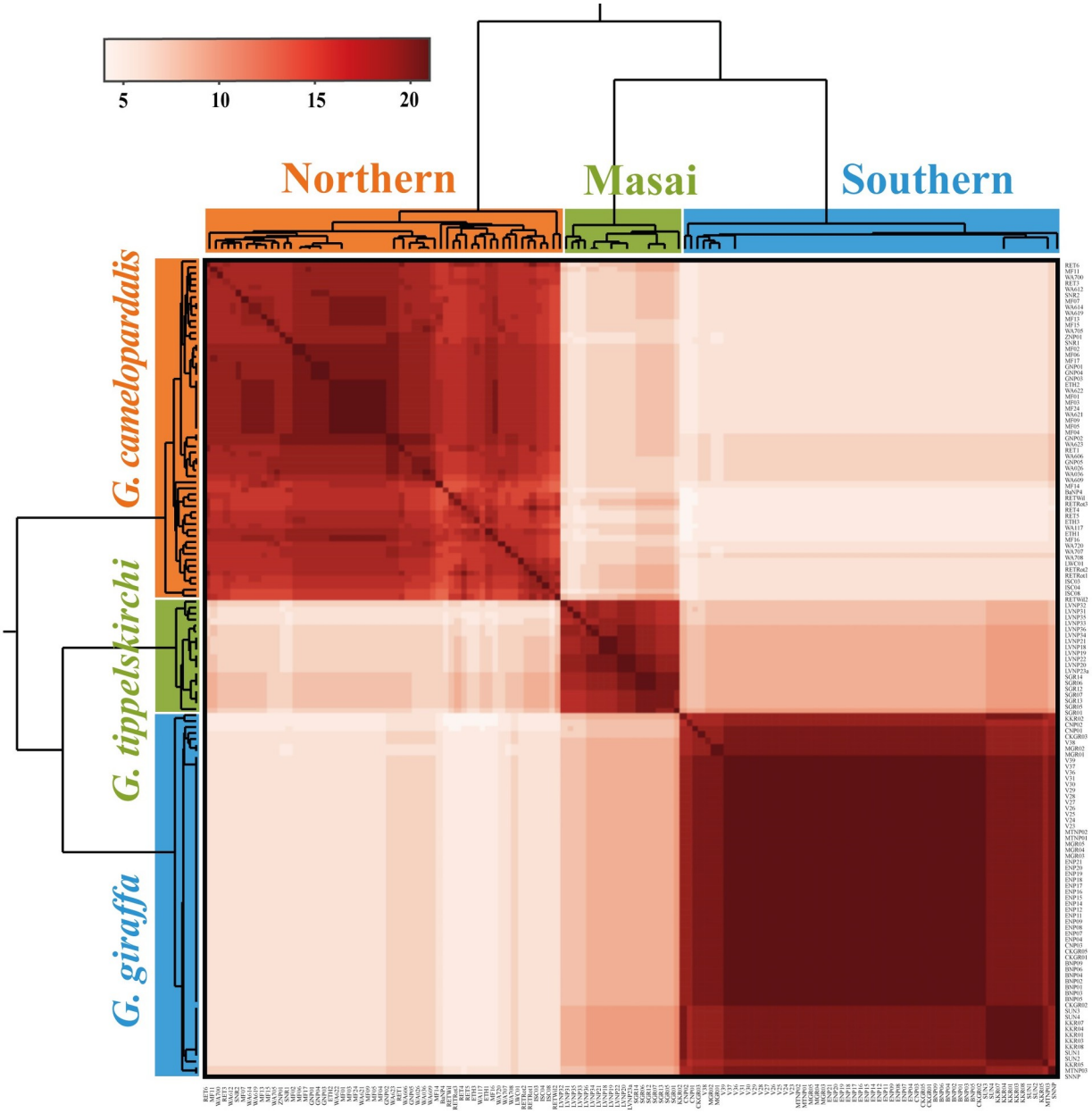


Fig 5. Conspecificity matrix of giraffe taxa. The conspecificity matrix was generated by calculating for each pair of individuals a conspecificity score equal to the number of independent markers supporting the hypothesis of conspecificity in haploweb analyses of the 21 introns. The conspecificity matrix was visualized as a heat map in which the highest scores (21 out of 21) are shown in dark red, the lowest scores (0 of 21) are shown in white and intermediate scores are depicted in various nuances of red.

https://doi.org/10.1371/journal.pone.0217956.g005

BP = 99) and *rothschildi* (PP = 1; BP = 83). The subspecies *camelopardalis* is found polyphyletic. The reticulated giraffes constitute a polyphyletic assemblage: although most of them are grouped together (PP = 1; BP = 92) as the sister group of the divergent haplotype TIP15 (EU088334) of *G. tippelskirchi* (PP = 1; BP = 94), the haplotype RET8 sequenced by Fennessy et al. [38] is closely related to *rothschildi* (PP = 0.89; BP = 46), and the haplotype RET9 (EU088321) sequenced by Brown et al. [29] appears as the sister group of all other northern haplotypes.

The E haplogroup comprises giraffes from eastern and southeastern Africa (PP = 0.99; BP = 87). It contains members of two putative species, *G. tippelskirchi* and *G. giraffa*, and can be further divided into three subgroups corresponding to "Masai I", "Masai II", and the subspecies *giraffa*. The interrelationships between the three subgroups are unresolved. The Masai I subgroup (PP = 1; BP = 95) contains Masai giraffes (subspecies *tippelskirchi*) from Kenya and Tanzania. The Masai II subgroup (PP = 1; BP = 89) includes Masai giraffes (subspecies *tippelskirchi*) from Kenya and Tanzania, as well as giraffes of the subspecies *thornicrofti* from northern Zambia (Luangwa Valley National Park). The third subgroup represents the subspecies *giraffa* (PP = 1; BP = 99) and includes giraffes from southern Zambia, northern Botswana, northeastern Namibia, Zimbabwe and South Africa.

The S haplogroup contains exclusively individuals of the subspecies *angolensis* from Namibia and central Botswana. Its monophyly is less supported than the two other mitochondrial haplogroups (PP = 0.37; BP = 60). Our analyses provide a moderate support (PP = 0.94; BP = 65) for an early divergence of the S haplogroup.

Nuclear and mitochondrial pairwise distances

The alignment of 21 nuclear introns was used to calculate pairwise distances between giraffes (Fig 6 and Table B in S4 Appendix). The results show that the mean distance between *G. camelopardalis* s.s. B and *G. reticulata* is 0.14% and the mean distance between *G. camelopardalis* s.s. C and *G. peralta* is 0.07%, which is significantly smaller than other interspecific distances involving *G. camelopardalis* s.s. A, *G. giraffa* and *G. tippelskirchi* (comprised between 0.35 and 0.51%).

For the mtDNA alignment, we calculated pairwise distances between 82 haplotypes. Three haplotypes (TIP15, RET8 and RET9) were excluded from the analysis due to their grouping outside of their assigned taxon in the phylogenetic tree (Fig 1B). The distances between the haplogroups identified in Fig 1B are summarized in Table A in S4 Appendix and Fig 6. There are three major haplogroups: haplogroup N = northern (= *G. camelopardalis* s.s. A); haplogroup E = Masai I, Masai II, and southeastern (= subspecies *giraffa*); and haplogroup S = southwestern (= subspecies *angolensis*). The mean distances between these three haplogroups are comprised between 3.07 and 4.16%. Within haplogroup N, the distances between *G. camelopardalis* s.s. B and *reticulata* range from 1.29% (ROTH3 *versus* RET3) to 2.19% (PER2 *versus* RET13). Within haplogroup E, we found similar distances between Masai I, Masai II and southeastern haplotypes, i.e., between 1.17% (TIP1 *versus* GFA7) and 2.12% (TIP5 *versus* GFA9). Within haplogroup S, the distances range from 0 to 0.96% (ANG12 *versus* ANG16).

Discussion

Population genetic analyses support the 3S hypothesis

The assessment of population genetic structure has become indispensable in evolutionary biology and conservation to reveal hidden biodiversity. Among freely accessible software provided for this task, STRUCTURE [26] is the most commonly used program, with 17473 citations in

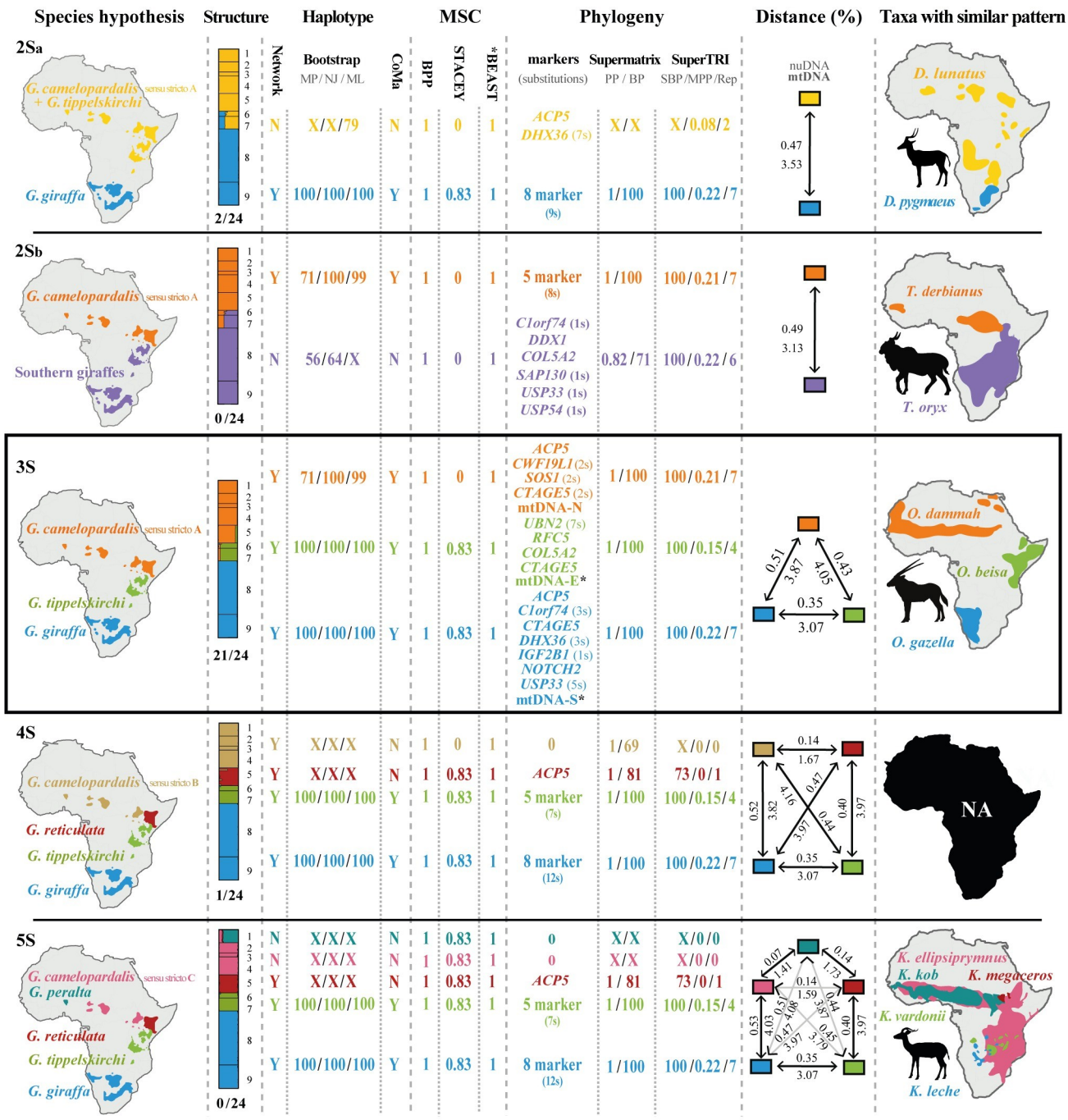


Fig 6. The five molecular hypotheses for giraffe taxonomy. The five taxonomic hypotheses that received some support from our analyses on giraffes show the existence of two species, with two possible geographic patterns (2Sa and 2Sb hypotheses), three species (3S hypothesis), i.e. *G. camelopardalis* sensu stricto A, *G. giraffa* and *G. tippelskirchi*, four species (4S hypothesis), i.e. *G. camelopardalis* sensu stricto B, *G. giraffa*, *G. reticulata*, and *G. tippelskirchi*, or five species (5S hypothesis), i.e. *G. camelopardalis* sensu stricto C, *G. giraffa*, *G. peralta*, *G. reticulata*, and *G. tippelskirchi*. In the first column are drawn the geographic distributions of giraffe species for each of the five taxonomic hypotheses. In the second column are summarized the results obtained from STRUCTURE analyses. Barplots were illustrated with DISTRUCT (1 = *peralta*, 2 = *antiquorum*, 3 = *camelopardalis*, 4 = *rothschildi*, 5 = *reticulata*, 6 = *tippelskirchi*, 7 = *thornicrofti*, 8 = *giraffa*, 9 = *angolensis*) and number of analyses supporting each taxonomic hypothesis (in total 24, see Table 3) is indicated beneath barplots. In the third column are illustrated the results obtained in the different haplotype analyses, including the network analysis (Y = yes, the species represents a cluster; N = no, the species is not found as a cluster), the bootstrap values obtained with the phylogenetic analyses based on the Maximum Parsimony (MP), Distance (NJ) and Maximum Likelihood (ML) criterion ("X": support < 50) and the conspecificity matrix (CoMa) (Y = yes, the species is supported by the analysis; N = no, the species is not supported by the analysis). In the fourth column are shown the support values provided by the three Multispecies coalescent (MSC) methods, i.e. BPP, STACEY and *BEAST. In the fifth column are listed the results

obtained in the phylogenetic analyses including the markers supporting each taxonomic hypothesis in the separate analyses of 21 introns and mtDNA, as well as the support values obtained from supermatrix (PP = posterior probability; BP = bootstrap percentage) and SuperTRI analyses (SBP = SuperTri Bootstrap Percentage; MPP = Mean Posterior Probability; Rep = Reproducibility Index) ("-": not found). In the sixth column are detailed the mean pairwise distances between individuals of the same taxon calculated using either nuDNA data (concatenation of 21 introns, above) or mtDNA (below) (Since all the mitochondrial sequences of the subspecies *giraffa* belong to haplogroup E, they were considered as *tippelskirchi* for distance comparisons; see Discussion for more details on mtDNA introgression). In the seventh column are shown the distribution maps of bovid genera with a similar geographic pattern of speciation.

https://doi.org/10.1371/journal.pone.0217956.g006

Web of Science (January 2019). Using Bayesian inference, STRUCTURE is a model-based clustering method to detect population structure and assign individuals to K populations [26]. However, many published results based on STRUCTURE are not reproducible because the genotypes were not available or the parameters used for the analyses were not fully detailed by the authors [52, 70].

The program STRUCTURE was previously used to infer genetic structure in giraffe populations, using either genotypes from 14 microsatellite loci of 381 individuals [29] or phased alleles of seven introns for 105 giraffes [38] or rather the extended dataset of 21 introns for 137 individuals [12]. Brown et al. [29] suggested the existence of at least six species, but the optimal K was not determined using either the method of Evanno et al. [57] or that of Pritchard et al. [51], and their results are not reproducible, because the microsatellite data were not made available. According to Winter et al. [12], "K = 4 shows four well resolved groups and is supported as best fitting number of clusters by several statistical methods", but they did not provide any details on the model and method used for their STRUCTURE analyses. Using the same dataset, comprising allelic information of 21 nuclear introns for 137 giraffes, we tested 16 different models under STRUCTURE in order to shed more light on giraffe population structure. Considering the method of Evanno et al. [57], 58% of the analyses provided support for two distinct populations of origin (K = 2), 25% for three distinct clusters (K = 3), and only 17% confirmed the result obtained by Winter et al. [12], i.e. K = 4.

The selection of the appropriate K using the method of Pritchard et al. [51] partly confirmed previously mentioned difficulties to determine the point of plateau [53, 57]. We clearly recognize K = 3 as the optimal clustering for 50% of the analyses. For other analyses, it was difficult to identify at which K the plateau is reached (K = 2 or 3?; K = 2, 3 or 4?; K = 3 or 4?; K = 3, 4 or 5?; Table 3, see Figs A-X in S3 Appendix).

Selecting the best suitable model for STRUCTURE is far from simple, especially for taxa with a wide distribution range like giraffes. The choice of an admixture model with correlated allele frequency seems appropriate for populations of East Africa, where hybrids between individuals from divergent populations were previously described (see below). However, such a model may be more questionable for isolated populations, such as the subspecies *peralta*. In order to better estimate the optimal value of K under STRUCTURE, we recommend therefore for future users of the program to test different combinations of model parameters, to estimate the value of λ, and to make comparison between optimal K estimated with either the ΔK method [57] or the "plateau" method [51]. Using this approach and taking into account that the ΔK method can be biased towards K = 2 [70] and that the smallest value of K is preferred when several values of K give similar estimates of log Pr (X | K) [51], we concluded that K = 3 is the most likely hypothesis for 88% of the analyses (highlighted in grey in Table 3).

Our network and bootstrap analyses of the 274 nuclear giraffe haplotypes (21 introns, 137 giraffes), the networks and haplowebs of the 21 introns, as well as the conspecificity matrix also highly support a division into three divergent haplogroups, representing the three species *G. camelopardalis* s.s. A, *G. giraffa*, and *G. tippelskirchi* (Figs 2, 3, 4 and 5).

Phylogenetic analyses support the 3S hypothesis

In the nuclear tree reconstructed from the concatenation of 21 introns (Fig 2), four putative species were found to be monophyletic: *G. giraffa*, *G. tippelskirchi*, *G. camelopardalis* s.s. A and *G. reticulata*. However, the two latter mentioned taxa obtained weak ML bootstrap support (BP = 69 and 81, respectively). To further investigate phylogenetic relationships, we conducted separate Bayesian analyses for all markers and summarized the results with the SuperTRI method [20]. Within *Giraffa*, the analyses showed that only four nodes can be considered as reliable (SBP = 100; MPP $>$ 0.15; Nrep $>$ 4): *G. camelopardalis* s.s. A (grouping together northern and reticulated giraffes), *G. giraffa* (southern giraffes), *G. tippelskirchi* (southeastern giraffes), and *G. giraffa* + *G. tippelskirchi* (Fig 2). All these nodes are supported by the separate analyses of several independent introns (between four and seven), which explain why MPP values are significantly higher than for all intraspecific relationships (between 0.15 and 0.22 versus between 0 and 0.03). By contrast, the SuperTRI analyses provided no support (MPP = 0; Nrep $\leq$ 1) for the existence of both *G. camelopardalis* s.s. B and *G. reticulata*. The monophyly of *G. reticulata* was found by only *ACP5*, but with insignificant support (PP = 0.03).

Multispecies coalescent approaches show further geographic structure

Two MSC methods, *BEAST and BPP, showed strong support (PP = 1) for the 3S hypothesis, in which three species can be distinguished, i.e., *G. camelopardalis* s.s. A, *G. giraffa*, and *G. tippelskirchi*. However, STACEY analyses provided support for further species delimitation, i.e., the 5S hypothesis (87%). The five taxa, *G. camelopardalis* s.s. C, *G. giraffa*, *G. peralta*, *G. reticulata*, and *G. tippelskirchi*, are also highly supported by both *BEAST and BPP analyses (PP = 1). As recently pointed by Sukumarana and Knowles [71] and Jackson et al. [72], it appears that multispecies coalescent methods delimit structure, not species. In agreement with that, it is important to note that only two of the five putative MSC species can be diagnosed by molecular signatures (Fig 1), i.e. the ones assumed by the 3S hypothesis: *G. tippelskirchi* is characterised by seven exclusive synapomorphies (ES), all found in the *UBN2* gene, which are shared by 19 individuals; and *G. giraffa* is characterised by 12 ES detected in four independent genes and shared by 61 individuals. For the three other taxa of the MSC 5S hypothesis, we did not detect any fixed mutation in the 21 nuclear introns. This means that the populations of *G. camelopardalis* s.s. C, *G. peralta*, and *G. reticulata* have never been completely isolated genetically. Their grouping into *G. camelopardalis* s.s. A is however supported by eight ES detected in five independent genes and shared by 57 individuals. The 3S hypothesis is therefore strengthened by the criterion of genetic isolation, as the detection of ES in the three species *G. camelopardalis* s.s. A, *G. giraffa*, and *G. tippelskirchi* indicates that their populations were reproductively isolated during enough time, allowing for the fixation of diagnostic mutations in all individuals.

Interspecies relationships within *Giraffa*

According to the fossil record, contemporary giraffes first appeared during the Pleistocene around 1 Mya [73], a hypothesis also supported by molecular dating estimates [74]. All candidate species to root the tree of giraffes are highly distant taxa: *Okapia*, which is the only other extant genus of the family Giraffidae, separated from *Giraffa* during the Middle Miocene (around 15.2 Mya); other ruminant families, such as Bovidae, Cervidae, Moschidae and Antilocapridae, diverged from Giraffidae at the transition between Oligocene and Miocene (around 23.4 Mya) [74]. The rooting of the giraffe tree can be therefore misleading due to a long branch attraction (LBA) artefact (for a review see Bergsten [75]) between the distant outgroup and one of the longest branches of the ingroup. This problem explains the highly variable root

position in our mitochondrial analyses: with MrBayes, the first haplogroup to diverge is either S (Fig 1B, PP = 0.37; BP = 60) or E (if the two bovid species are excluded as outgroup taxa, data not shown, PP = 0.55); with BEAST, haplogroups E and S are found to be sister-groups (PP = 0.74), as in the mitochondrial tree of Fennessy et al. [38].

The nuclear dataset provided more signal for resolving basal relationships within *Giraffa*. As indicated in Fig 1, our phylogenetic analyses supported a sister-group relationship between *G. giraffa* and *G. tippelskirchi* (PP = 0.82; BP = 71). This node was found monophyletic with 6 independent markers (*C1orf74*, *DDX1*, *COL5A2*, *SAP130*, *USP33* and *USP54*). By comparison, SuperTRI analyses clearly showed that the two other hypotheses (either *G. camelopardalis* s.s. A + *G. giraffa* or *G. camelopardalis* s.s. A + *G. tippelskirchi*) are less supported (MPP $\leq$ 0.09; NRep $\leq$ 2 markers). All these results agree therefore with a deep North/ South dichotomy within *Giraffa*.

Evidence for introgressive hybridization between giraffe species

The comparison between the mtDNA tree based on 82 giraffe haplotypes and the nuclear tree reconstructed from 21 introns sequenced for 137 giraffes reveals a robust conflict for the evolutionary history drawn from maternal and biparental markers (Fig 1). Some mito-nuclear conflicts can be simply explained by recent hybridization between sympatric or parapatric taxa (species or subspecies), resulting in the transfer of the mitochondrial genome from one taxa to the other, a process referred to as mitochondrial introgression [6–11].

A first case of potential hybridisation is represented by the mitochondrial haplotype TIP15, which constitutes the sister-group of the main haplogroup of reticulated giraffes (Fig 2B), from which it differs by a distance of only 1%. The nine Masai giraffes possessing this haplotype were collected in southern Kenya (Athi River Ranch) [29], where wild populations of *tippelskirchi* and *reticulata* can sometimes hybridize [76]. We suggest therefore that introgressive hybridization can account for the transfer of the mitochondrial haplotype TIP15 from *reticulata* to *tippelskirchi*. The allelic networks of the 21 nuclear introns suggest also past nuclear introgression, this time from *tippelskirchi* to *reticulata*, as two individuals of *reticulata*, ISC04 and RETWil2, are characterized by several rare alleles identical or similar to those found in *tippelskirchi*: in *ACP5* (only for RETWil2), *COL5A2* (only for ISC04), *CTAGE5* and *DDX1* (both individuals) (Fig 3).

The second case of mitochondrial introgression concerns the haplotype RET8 detected in one reticulated giraffe from the Nürnberg Zoo [38]. Its grouping with Rothschild's giraffes may be explained by interbreeding between *reticulata* and *rothschildi* either in zoos [77] or in the wild, as field observations have documented the occurrence of *reticulata* X *rothschildi* hybrid phenotypes in Kenya [78]. Unfortunately, these hybrid individuals or populations were not yet studied for nuclear genes.

The mitochondrial haplotype RET9, which was detected by Brown et al. [29] in a single reticulated giraffe (accession number: EU08821), is intriguing because it is divergent from all other sequences of haplogroup N. We propose two hypotheses to explain its divergence. The first hypothesis assumes the retention of ancestral haplotypes in wild populations of reticulated giraffes; it will be confirmed if identical or similar haplotypes are discovered in other reticulated giraffes. Another hypothesis implies that the sequence EU088321 is problematic, either because it contains multiple sequencing errors or because it is a nuclear sequence of mitochondrial origin (Numt) [79]. Obviously, further investigations are needed to solve this issue.

The most important and interesting mito-nuclear discordance concerns giraffes from eastern and southern Africa. In the nuclear tree (Fig 2A), these giraffes are divided into two geographic groups corresponding to two different species: giraffes from southern Africa (South

Africa, Namibia, Botswana, and southern Zambia) belong to *G. giraffa*, whereas eastern giraffes (southern Kenya, Tanzania, and northern Zambia) belong to *G. tippelskirchi*. These two species are not monophyletic in the mitochondrial tree: *G. giraffa* is polyphyletic, because members of the two subspecies *giraffa* and *angolensis* are not grouped together; whereas *G. tippelskirchi* is paraphyletic, due to the inclusive position of the subspecies *giraffa* (southeastern giraffes). To interpret these conflicting results, it is crucial to remember that basal relationships within *Giraffa* are not reliable in the mitochondrial tree, due to a high genetic distance towards outgroup taxa (see above for explanations). Taken this in mind, it can be hypothesized that the three species identified with nuclear data were characterized by three different ancestral mitochondrial haplogroups: N for *G. camelopardalis* s.s. A, E for *G. tippelskirchi*, and S for *G. giraffa*. According to this hypothesis, we can further propose that the common ancestor of southeastern populations of *G. giraffa* (subspecies *G. g. giraffa*) acquired a mitochondrial genome from *G. tippelskirchi* (haplogroup E) by introgressive hybridization between parapatric populations. Using a calibration at 1 ± 0.1 Mya for the common ancestor of giraffes [73, 74], we estimated that the introgressive event occurred around 420 kya (see Fig C in S2 Appendix), i.e. during one of the most important glacial periods of the Pleistocene. In sub-Saharan Africa, glacial periods were generally characterized by the contraction of forest areas and the concomitant extension of open areas, such as savannahs and deserts. In addition, river levels were lower, facilitating dispersals and the colonization of new areas. Since Pleistocene environments were more stable in subtropical southern East Africa than in tropical East Africa [80], we suggest that some Masai giraffes migrated around 420 kya from East Africa to southern East Africa, promoting secondary contacts between *G. tippelskirchi* and *G. giraffa*, and therefore the mitochondrial introgression of haplotype E into *G. g. giraffa*. In the latter subspecies, the ancestral haplotype S has been completely replaced by the new haplotype E. By contrast, the ancestral haplotype S has been maintained in southwestern populations of the subspecies *G. g. angolensis*. The absence of haplotype E in southwestern giraffes suggests that female giraffes were not able to disperse from East to West and reciprocally. Important biogeographical barriers may have been the Kalahari Desert during glacial periods of the Pleistocene, and the Okavango Delta associated with Palaeo-lake Makgadikgadi during interglacial periods. However, nuclear data support gene flow mediated by dispersing males between eastern (*G. g. giraffa*) and western populations (*G. g. angolensis*) of southern giraffes. Female philopatry and male biased dispersal are classically observed in mammal species [81]. In giraffes, such different sexual behaviours can be explained by nursery herds, which consist of several females and their offspring [82], and by solitary males, which spend a lot of time to find receptive females. Thereby, males may often have to migrate over long distances to successfully pass on their genes [83]. In this regard, we can assume that males are generally more willing than females to take the risk of overcoming biogeographic barriers, such as deep rivers or large deserts. Markers from the Y chromosome should be sequenced to further address our biogeographic scenario involving a better dispersal capacity for males than females.

Conclusion for giraffe conservation management

The species is the most important taxonomic unit for conservation assessments and for the establishment of justified management plans [84, 85]. Giraffes are currently considered as a single species by the IUCN [37], but its status has recently moved from Least Concern to Vulnerable due to a population decline of 36–40% over three generations. Even though, the situation seems to have improved for some populations (e.g. *giraffa* [86]; *peralta* [87]) in the course of enhanced conservation management, population numbers of most subspecies continue to decrease [37].

Our taxonomic study indicates that the conservation status should be separately assessed for the three species *G. camelopardalis* s.s. A (northern giraffes), *G. giraffa* (southern giraffes) and *G. tippelskirchi* (Masai giraffes). According to population estimations of the IUCN [37], the southern species *G. giraffa*, has recently increased by 168% and hence fall into the category "Least Concern"; the East African species *G. tippelskirchi* has decreased by $\geq$ 50% over a period of three generations and hence should be listed as "Vulnerable"; the northern species *G. camelopardalis* s.s. A has decreased by $\geq$ 70% over the past 30 years and with only 20 000 individuals left in the wild, it should be listed under the category "Endangered" (according to Criterion A1 [88]).

Supporting information

S1 Appendix. Classifications of *Giraffa*.
(PDF)

S2 Appendix. Phylogenetic analyses.
(PDF)

S3 Appendix. Population Structure analyses.
(PDF)

S4 Appendix. Genetic distances.
(PDF)

S5 Appendix. Multispecies coalescent approach.
(PDF)

S6 Appendix. Analyses of nuclear haplotypes.
(PDF)

Acknowledgments

AH would like to thank Klaus-Peter Koepfli and Robert K. Wayne, who provided additional information and clarification on their sequences published in 2007. We would also like to acknowledge Jean-François Flot and one anonymous reviewer for constructive comments on the manuscript.

Author Contributions

Conceptualization: Alice Petzold, Alexandre Hassanin.

Formal analysis: Alice Petzold, Alexandre Hassanin.

Funding acquisition: Alexandre Hassanin.

Methodology: Alice Petzold, Alexandre Hassanin.

Supervision: Alexandre Hassanin.

Writing – original draft: Alice Petzold, Alexandre Hassanin.

Writing – review & editing: Alice Petzold, Alexandre Hassanin.

References

1. Rabosky DL. Reproductive isolation and the causes of speciation rate variation in nature. Biol J Linn Soc. 2016; 118(1): 13–25.

2. Fišer C, Robinson CT, Malard F. Cryptic species as a window into the paradigm shift of the species concept. Mol Ecol. 2018; 27(3): 613–635. https://doi.org/10.1111/mec.14486 PMID: 29334414
3. Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, et al. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst Biol. 2006; 55(4): 595–609. https://doi.org/10.1080/10635150600852011 PMID: 16967577
4. Ratnasingham S, Hebert PDN. A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PloS One 2013; 8(7): e66213. https://doi.org/10.1371/journal.pone.0066213 PMID: 23861743
5. Petit RJ, Excoffier L. Gene flow and species delimitation. Trends Ecol Evol. 2009; 24(7):386–393. https://doi.org/10.1016/j.tree.2009.02.011 PMID: 19409650
6. Hassanin A, Khouider S, Gembu GC, Goodman SM, Kadjo B, Nesi N, et al. The comparative phylogeography of fruit bats of the tribe Scotonycterini (Chiroptera, Pteropodidae) reveals cryptic species diversity related to African Pleistocene forest refugia. C R Biol. 2015; 338(3): 197–211. https://doi.org/10.1016/j.crvi.2014.12.003 PMID: 25636226
7. Ropiquet A, Hassanin A. Hybrid origin of the Pliocene ancestor of wild goats. Mol Phylogenetics Evol. 2006; 41(2): 395–404.
8. Hassanin A, Ropiquet A. Resolving a zoological mystery: the kouprey is a real species. Proc Royal Soc. B 2007; 274(1627): 2849–2855.
9. Hassanin A. The role of Pleistocene glaciations in shaping the evolution of polar and brown bears. Evidence from a critical review of mitochondrial and nuclear genome analyses. C R. Biol. 2015; 338(7): 494–501. https://doi.org/10.1016/j.crvi.2015.04.008 PMID: 26026577
10. Hassanin A, Houck ML, Tshikung D, Kadjo B, Davis H, Ropiquet A. Multi-locus phylogeny of the tribe Tragelaphini (Mammalia, Bovidae) and species delimitation in bushbuck: Evidence for chromosomal speciation mediated by interspecific hybridization. Mol Phylogenet. Evol. 2018; 129: 96–105. https://doi.org/10.1016/j.ympev.2018.08.006 PMID: 30121341
11. Hassanin A, Colombo R, Gembu GC, Merle M, Tu VT, Görföl T, et al. Multilocus phylogeny and species delimitation within the genus *Glauconycteris* (Chiroptera, Vespertilionidae), with the description of a new bat species from the Tshopo Province of the Democratic Republic of the Congo. J Zool Syst Evol Res. 2018; 56(1): 1–22.
12. Winter S, Fennessy J, Janke A. Limited introgression supports division of giraffe into four species. Ecol Evol. 2018; 8(20): 10156–10165. https://doi.org/10.1002/ece3.4490 PMID: 30397455
13. Fischer A, Prüfer K, Good JM, Halbwax M, Wiebe V, André C, et al. Bonobos fall within the genomic variation of chimpanzees. PLoS One 2011; 6: e21605. https://doi.org/10.1371/journal.pone.0021605 PMID: 21747915
14. Hassanin A, An J, Ropiquet A, Nguyen TT, Couloux A. Combining multiple autosomal introns for studying shallow phylogeny and taxonomy of Laurasiatherian mammals: Application to the tribe Bovini (Cetartiodactyla, Bovidae). Mol Phylogenet. Evol. 2013; 66(3): 766–775. https://doi.org/10.1016/j.ympev.2012.11.003 PMID: 23159894
15. Flot JF. Species delimitation's coming of age. Syst Biol. 2015; 64(6): 897–899. https://doi.org/10.1093/sysbio/syv071 PMID: 26420142
16. Mello B, Vilela JF, Schrago CG. Conservation phylogenetics and computational species delimitation of Neotropical primates. Biol Conserv. 2018; 217: 397–406.
17. De Queiroz K. Species concepts and species delimitation. Syst Biol. 2007; 56(6): 879–886. https://doi.org/10.1080/10635150701701083 PMID: 18027281
18. De Queiroz A, Gatesy J. The supermatrix approach to systematics. Trends Ecol Evol. 2007; 22(1): 34–41. https://doi.org/10.1016/j.tree.2006.10.002 PMID: 17046100
19. Heled J, Drummond AJ. Bayesian inference of species trees from multilocus data. Mol Biol Evol. 2009; 27(3): 570–580. https://doi.org/10.1093/molbev/msp274 PMID: 19906793
20. Ropiquet A, Li B, Hassanin A. SuperTRI: a new approach based on branch support analyses of multiple independent data sets for assessing reliability of phylogenetic inferences. C R Biol. 2009; 332(9): 832–847. https://doi.org/10.1016/j.crvi.2009.05.001 PMID: 19748458
21. Knowles LL, Carstens BC. Delimiting species without monophyletic gene trees. Syst Biol. 2007; 56(6):887–895. https://doi.org/10.1080/10635150701701091 PMID: 18027282
22. Kingman JFC. The coalescent. Stoch Process Their Appl. 1982; 13(3): 235–248.
23. Yang Z. The BPP program for species tree estimation and species delimitation. Curr Zool. 2015; 61(5): 854–865.

24. Jones G. Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. J Math Biol. 2017; 74(1–2): 447–467. https://doi.org/10.1007/s00285-016-1034-0 PMID: 27287395

25. Fujisawa T, Aswad A, Barraclough TG. A rapid and scalable method for multilocus species delimitation using Bayesian model comparison and rooted triplets. Syst Biol. 2016; 65(5): 759–771. https://doi.org/10.1093/sysbio/syw028 PMID: 27055648

26. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000; 155(2): 945–959. PMID: 10835412

27. Olave M, Sola E, Knowles LL. Upstream analyses create problems with DNA-based species delimitation. Syst Biol. 2014; 63(2): 263–271. https://doi.org/10.1093/sysbio/syt106 PMID: 24361591

28. Kalinowski ST. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity 2011; 106(4): 625–632. https://doi.org/10.1038/hdy.2010.95 PMID: 20683484

29. Brown DM, Brenneman RA, Koepfli KP, Pollinger JP, Milá B, Georgiadis NJ, et al. Extensive population genetic structure in the giraffe. BMC Biol. 2007; 5:1–13. https://doi.org/10.1186/1741-7007-5-1

30. Flot JF, Couloux A, Tillier S. Haplowebs as a graphical tool for delimiting species: a revival of Doyle's" field for recombination" approach and its application to the coral genus Pocillopora in Clipperton. BMC Evol. Biol. 2010; 10: 372–386. https://doi.org/10.1186/1471-2148-10-372 PMID: 21118572

31. Debortoli N, Li X, Eyres I, Fontaneto D, Hespeels B, Tang CQ, et al. Genetic exchange among bdelloid rotifers is more likely due to horizontal gene transfer than to meiotic sex. Curr. Biol. 2016; 26: 723–732. https://doi.org/10.1016/j.cub.2016.01.031 PMID: 26948882

32. Saint-Hilaire G. Quelques considérations sur la Girafe. Ann Sci Nat. 1827; 210–237.

33. Thomas O. On the five-horned Giraffe obtained by Sir Marry Johnston near Mount Elgon. J Zool. 1901; 71: 474–483.

34. Lydekker R. Catalogue of the ungulate mammals in the British Museum (Natural History). 3rd ed. London: British Museum Trustees; 1914. pp. 234–257.

35. Dagg AI, Foster JB. The giraffe: its biology, behaviour, and ecology. New York: Van Nostrand Reinhold Company; 1976.

36. Wilson DE, Reeder DM. Mammal species of the world: a taxonomic and geographic reference. 1st ed. Maryland: John Hopkins University Press; 2005.

37. IUCN. The IUCN Red List of Threatened Species. Available from: http://www.iucnredlist.org (accessed 22 March 2019).

38. Fennessy J, Bidon T, Reuss F, Kumar V, Elkan P, Nilsson MA, et al. Multi-locus analyses reveal four giraffe species instead of one. Curr Biol. 2016; 26: 2543–2549. https://doi.org/10.1016/j.cub.2016.07.036 PMID: 27618261

39. Bercovitch FB, Berry PS, Dagg A, Deacon F, Doherty JB, Lee DE, et al. How many species of giraffe are there? Curr. Biol. 2017; 27:136–R137.

40. Donoghue MJ. A critique of the biological species concept and recommendations for a phylogenetic alternative. Bryologist 1985; 172–181.

41. Baker RJ, Bradley RD. Speciation in mammals and the genetic species concept. J. Mammal 2006; 87: 643–662. https://doi.org/10.1644/06-MAMM-F-038R2.1 PMID: 19890476

42. Baum DA, Shaw KL. Genealogical perspectives on the species problem. In: Experimental and molecular approaches to plant biosystematics. Saint Louis: Botanical Garden; 1995.

43. Isaac NJ, Mallet J, Mace GM. Taxonomic inflation: its influence on macroecology and conservation. Trends Ecol. Evol. 2004; 19: 464–469. https://doi.org/10.1016/j.tree.2004.06.004 PMID: 16701308

44. Stanton DW, Frandsen P, Waples RK, Heller R, Russo IRM, Orozco-terWengel PA, et al. More grist for the mill? Species delimitation in the genomic era and its implications for conservation. Conserv. Genet 2019; 20: 101–113.

45. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 2009; 25: 1451–1452. https://doi.org/10.1093/bioinformatics/btp187 PMID: 19346325

46. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012; 61: 539–542. https://doi.org/10.1093/sysbio/sys029 PMID: 22357727

47. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010; 59: 307–321. https://doi.org/10.1093/sysbio/syq010 PMID: 20525638

48. Darriba D, Taboada GL, Doalla R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 2012; 9: 772–776.

49. Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control 2017; 19:716–723.
50. Swofford DL. PAUP*v4.0.b10: phylogenetic analysis using parsimony and other methods. Sinauer Associates, Sunderland 2003.
51. Pritchard JK, Wen X, Falush D. Documentation for STRUCTURE software, version 2.3. University of Chicago 2010.
52. Gilbert KJ, Andrew RL, Bock DG, Franklin MT, Kane NC, Moore JS, et al. Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program STRUCTURE. Mol Ecol. 2012; 21: 4925–4930. https://doi.org/10.1111/j.1365-294X.2012.05754.x PMID: 22998190
53. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 2003; 164: 1567–1587. PMID: 12930761
54. Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu MV. An overview of STRUCTURE: applications, parameter settings, and supporting software. Front Genet. 2013; 4: 1–13.
55. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. Mol Ecol Res. 2009; 9: 1322–1332.
56. Earl DA, von Holdt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 2011; 4: 359–361.
57. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005; 14: 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x PMID: 15969739
58. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Res. 2015; 15: 1179–1191.
59. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. Mol Ecol Resour. 2004; 4: 137–138.
60. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 1999; 16: 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036 PMID: 10331250
61. Leigh JW, Bryant D. PopART: full-feature software for haplotype network construction. Methods Ecol Evol. 2015; 6(9): 1110–1116.
62. Miller MA, Pfeiffer W, Schwartz T. The CIPRES science gateway: enabling high-impact science for phylogenetics researchers with limited resources. Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the extreme to the campus and beyond, ACM Chicago; 2012.
63. Etoundi E, Marescaux J, Vastrade M, Debortoli N, Hedtke SM, Pigneur LM, et al. Distinct biogeographic origins of androgenetic Corbicula lineages followed by genetic captures. BioRxiv 2019: 590836.
64. Yang Z, Rannala B. Unguided species delimitation using DNA sequence data from multiple loci. Mol Biol Evol. 2014; 31(12): 3125–3135. https://doi.org/10.1093/molbev/msu279 PMID: 25274273
65. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput Biol. 2014; 10: e1003537. https://doi.org/10.1371/journal.pcbi.1003537 PMID: 24722319
66. Kubatko LS, Degnan JH. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst Biol. 2007; 56(1): 17–24. https://doi.org/10.1080/10635150601146041 PMID: 17366134
67. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst Biol. 2018; 67(5): 901–904. https://doi.org/10.1093/sysbio/syy032 PMID: 29718447
68. Rambaut A, Drummond AJ. TreeAnnotator v1. 4.8. Published by the author 2007.
69. Jones G, Aydin Z, Oxelman B. DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. Bioinformatics 2014; 31(7): 991–998. https://doi.org/10.1093/bioinformatics/btu770 PMID: 25422051
70. Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI, et al. The K = 2 conundrum. Mol Ecol. 2017; 26: 3594–3602. https://doi.org/10.1111/mec.14187 PMID: 28544181
71. Sukumaran J, Knowles LL. Multispecies coalescent delimits structure, not species. Proc Natl Acad Sci USA. 2017; 114(7): 1607–1612. https://doi.org/10.1073/pnas.1607921114 PMID: 28137871
72. Jackson ND, Carstens BC, Morales AE, O'Meara BC. Species delimitation with gene flow. Syst Biol. 2017; 66(5): 799–812. https://doi.org/10.1093/sysbio/syw117 PMID: 28003535

73. Mitchell G, Skinner JD. On the origin, evolution and phylogeny of giraffes *Giraffa camelopardalis*. Trans R Soc S Afr. 2003; 58: 51–73.

74. Hassanin A, Delsuc F, Ropiquet A, Hammer C, van Vuuren BJ, Matthee C, et al. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. C R Biol. 2012; 335: 32–50. https://doi.org/10.1016/j.crvi.2011.11.002 PMID: 22226162

75. Bergsten J. A review of long-branch attraction. Cladistics 2005; 21: 163–193.

76. Kingdon J, Happold D, Butynski T, Hoffmann M, Happold M, Kalina J. Mammals of Africa. New York: Bloomsbury Publishing; 2013.

77. Krumbiegel I. Die Giraffe (*Giraffa camelopardalis*). Lutherstadt Wittenberg: A. Ziemsen Verlag; 1971.

78. Stott KW, Selsor CJ. Further remarks on giraffe intergradation in Kenya and unreported marking variations in reticulated and Masai giraffes. Mammalia 1981; 45: 261–263.

79. Hassanin A, Bonillo C, Nguyen BX, Cruaud C. Comparisons between mitochondrial genomes of domestic goat (*Capra hircus*) reveal the presence of numts and multiple sequencing errors. Mitochondrial DNA 2010; 21: 68–76. https://doi.org/10.3109/19401736.2010.490583 PMID: 20540682

80. Castañeda IS, Caley T, Dupont L, Kim JH, Malaizé B, Schouten S. Middle to Late Pleistocene vegetation and climate change in subtropical southern East Africa. Earth Planet Sci Lett 2016; 450: 306–316.

81. Greenwood PJ. Mating systems, philopatry and dispersal in birds and mammals. Animal Behav. 1980; 28: 1140–1162.

82. Carter KD, Seddon JM, Frère CH, Carter JK, Goldizen AW. Fission–fusion dynamics in wild giraffes may be driven by kinship, spatial overlap and individual social preferences. Animal Behav. 2013; 85: 385–394.

83. Bercovitch FB, Bashaw MJ, del Castillo SM. Sociosexual behavior, male mating tactics, and the reproductive cycle of giraffe *Giraffa camelopardalis*. Horm Behav. 2006; 50: 314–321. https://doi.org/10.1016/j.yhbeh.2006.04.004 PMID: 16765955

84. Mace GM. The role of taxonomy in species conservation. Phil Trans R Soc Lond. B 2004; 359: 711–719.

85. Groves CP, Cotterill FPD, Gippoliti S, Robovský J, Roos C, Taylor PJ, et al. Species definitions and conservation: a review and case studies from African mammals. Conserv Genet. 2017; 18: 1247–1256.

86. Deacon F, Tutchings A. The South African giraffe *Giraffa camelopardalis giraffa*: a conservation success story. Oryx 2019; 53(1): 45–48.

87. Suraud JP, Fennessy J, Bonnaud E, Issa AM, Fritz H, Gaillard JM. Higher than expected growth rate of the Endangered West African giraffe *Giraffa camelopardalis peralta*: a successful human–wildlife cohabitation. Oryx 2012; 46(4): 577–583.

88. IUCN. Guidelines for Using the IUCN Red List Categories and Criteria. 2011. Available from: https://cmsdata.iucn.org/downloads/redlistguidelines.pdf. (accessed 18 March 2019)