

# UCLA

## UCLA Previously Published Works

### Title

Regional Differences in Seasonal Timing of Rainfall Discriminate between Genetically Distinct East African Giraffe Taxa

### Permalink

<https://escholarship.org/uc/item/7qp11270>

### Journal

PLOS ONE, 8(10)

### ISSN

1932-6203

### Authors

Thomassen, Henri A  
Freedman, Adam H  
Brown, David M  
[et al.](#)

### Publication Date

2013

### DOI

10.1371/journal.pone.0077191

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# Regional Differences in Seasonal Timing of Rainfall Discriminate between Genetically Distinct East African Giraffe Taxa

Henri A. Thomassen<sup>1\*</sup><sup>‡a</sup>, Adam H. Freedman<sup>1‡b</sup>, David M. Brown<sup>2</sup>, Wolfgang Buermann<sup>1,3‡c</sup>, David K. Jacobs<sup>2</sup>

**1** Center for Tropical Research, University of California Los Angeles, Los Angeles, California, United States of America, **2** Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, California, United States of America, **3** Department of Atmospheric and Oceanic Sciences, University of California Los Angeles, Los Angeles, California, United States of America

## Abstract

Masai (*Giraffa tippelskirchi*), Reticulated (*G. reticulata*) and Rothschild's (*G. camelopardalis*) giraffe lineages in East Africa are morphologically and genetically distinct, yet in Kenya their ranges abut. This raises the question of how divergence is maintained among populations of a large mammal capable of long-distance travel, and which readily hybridize in zoos. Here we test four hypotheses concerning the maintenance of the phylogeographic boundaries among the three taxa: 1) isolation-by-distance; 2) physical barriers to dispersal; 3) general habitat differences resulting in habitat segregation; or 4) regional differences in the seasonal timing of rainfall, and resultant timing of browse availability. We used satellite remotely sensed and climate data to characterize the environment at the locations of genotyped giraffes. Canonical variate analysis, random forest algorithms, and generalized dissimilarity modelling were employed in a landscape genetics framework to identify the predictor variables that best explained giraffes' genetic divergence. We found that regional differences in the timing of precipitation, and resulting green-up associated with the abundance of browse, effectively discriminate between taxa. Local habitat conditions, topographic and human-induced barriers, and geographic distance did not aid in discriminating among lineages. Our results suggest that selection associated with regional timing of events in the annual climatic cycle may help maintain genetic and phenotypic divergence in giraffes. We discuss potential mechanisms of maintaining divergence, and suggest that synchronization of reproduction with seasonal rainfall cycles that are geographically distinct may contribute to reproductive isolation. Coordination of weaning with green-up cycles could minimize the costs of lactation and predation on the young. Our findings are consistent with theory and empirical results demonstrating the efficacy of seasonal or phenologically dictated selection pressures in contributing to the reproductive isolation of parapatric populations.

**Citation:** Thomassen HA, Freedman AH, Brown DM, Buermann W, Jacobs DK (2013) Regional Differences in Seasonal Timing of Rainfall Discriminate between Genetically Distinct East African Giraffe Taxa. PLoS ONE 8(10): e77191. doi:10.1371/journal.pone.0077191

**Editor:** Maura Geraldine Chapman, University of Sydney, Australia

**Received:** November 12, 2012; **Accepted:** September 8, 2013; **Published:** October 23, 2013

**Copyright:** © 2013 Thomassen et al. 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.

**Funding:** The authors have no support or funding to report.

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

\* E-mail: hathomassen@ucla.edu

<sup>‡a</sup> Current address: Comparative Zoology, Institute of Evolution and Ecology, University of Tübingen, Tübingen, Germany

<sup>‡b</sup> Current address: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States of America

<sup>‡c</sup> Current address: School of Earth and Environment, Faculty of Environment, University of Leeds, Leeds, United Kingdom

## Introduction

Population divergence and speciation can result from genetic drift in geographic isolation, or from spatially variable natural selection [1,2], even when gene flow is not completely restricted [3,4]. Most studies of population differentiation focus on a single evolutionary mechanism, testing whether it has a significant effect on divergence. However, in order to assess the relative importance of neutral and adaptive processes, it is crucial to also consider the alternatives in a multi-model comparison. Here, we examine evolutionary processes that may maintain divergence in reproductively isolated East African giraffe taxa with abutting distributions. We will first introduce the problem of genetically distinct parapatric giraffe species, and then discuss four scenarios that might contribute to the maintenance of divergence, which we compared in this study.

Giraffes range from the Sahel to South Africa, living in scrub and savannah habitat in loose social groups with home range sizes between 5 and 992 km<sup>2</sup> [5,6]. They are highly mobile, capable of long-distance movements of 50–300 km [5]. Despite their mobility, giraffes are characterized by extreme genetic divergence amongst parapatric lineages [7]. Across Africa, at least six distinct groups can be identified, with little evidence of hybridization [7]. Recently, Groves and Grubb [8] treated these taxa as distinct species, and we will do the same in this paper. In addition, according to Groves and Grubb [8], there is little evidence to support a distinction between *Giraffa camelopardalis* and *Giraffa rothschildi*, and we will thus follow their suggestion by treating Rothschild's giraffe as *G. camelopardalis*. In East Africa, divergence between Masai (*G. tippelskirchi*), Reticulated (*G. reticulata*), and Rothschild's (*G. camelopardalis*) giraffe lineages is supported by

strong genetic structure in mtDNA and microsatellites [7], and occurs despite the facts that these taxa have parapatric distributions [9], they are able to travel long distances [5], and they live in continuous acacia woodland habitat where barriers that could prevent movements among their respective ranges were seemingly absent in historic times, prior to anthropogenic habitat fragmentation [10]. Even though the species have different pelage patterns, with the potential for pre-mating isolation due to pelage-based mate recognition, individuals of these taxa hybridize readily in zoos [11,12]. In contrast, cases of hybridization in the wild are rarely reported. The genetic evidence from mtDNA sequences indicates that the Masai giraffe has been separated from the Rothschild's and Reticulated giraffes since the early to middle Pleistocene (1.62 mya–0.54 mya) and the Rothschild's from the Reticulated giraffe since the middle Pleistocene (0.54 mya–0.18 mya), with minimal subsequent gene flow [7]. Consequently, it was suggested that the three giraffe taxa represent different species rather than subspecies [7]. In addition to clear genetic breaks between species, strong genetic subdivisions are also evident within species, particularly within the Masai giraffe [7].

Given the apparent absence of geographical barriers to dispersal, the striking genetic differentiation among these giraffes suggests that environmental or behavioural mechanisms limit gene flow. Here we consider four scenarios for the maintenance of divergence among the East African giraffe taxa, focusing on geographic and environmental parameters in a first-order assessment of their relative importance in discriminating between the three taxa. We make no claims concerning the environmental, orographic, or other conditions that initiated divergence among these taxa in the Pleistocene. Rather, we restrict our assessment to late Holocene to modern processes that contribute to the maintenance of the current, nearly complete reproductive isolation among the giraffe lineages. To this end, we conduct multivariate and spatially non-explicit as well as spatially explicit analyses to evaluate four hypothesized isolating scenarios: 1) isolation-by-distance; 2) the presence of barriers to dispersal, limiting gene flow; 3) spatial habitat differences that do not represent differences in timing of the seasons; and 4) differences in the seasonal timing of precipitation in relation to green-up.

Hypothesis I – Isolation-by-distance is the effect of diminishing genetic relatedness with increasing distance, and could potentially be important when dispersal is limited relative to the overall size of the range. Even though isolation-by-distance appears to be an unlikely force maintaining divergence between parapatric taxa, simulations suggest that under some circumstances parapatric speciation is possible solely due to limited dispersal distances and the accumulation of genetic incompatibilities [13,14].

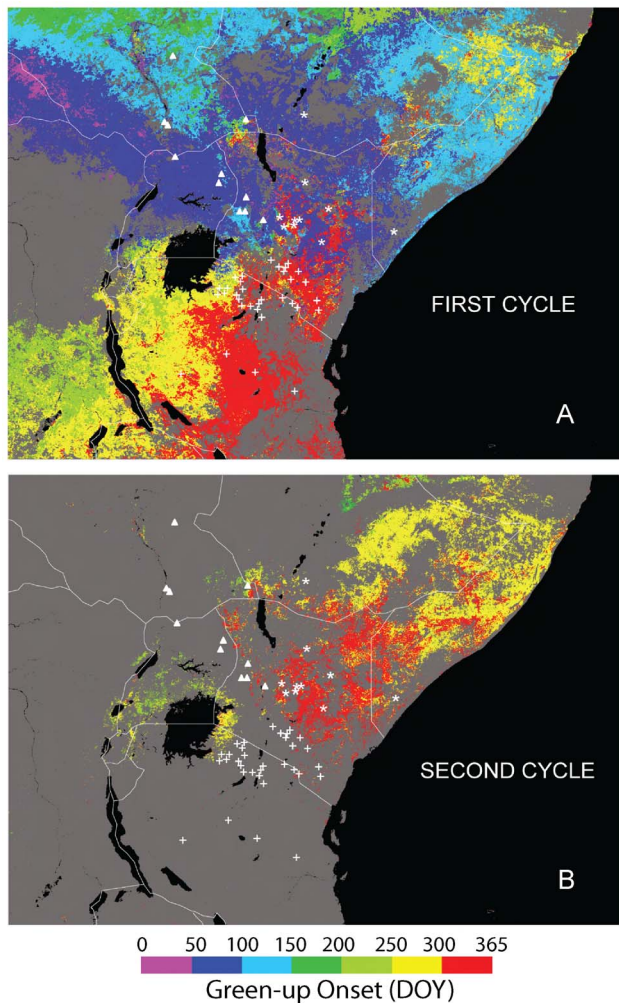
Hypothesis II – Geographic barriers to dispersal – and as a result gene flow – between the three giraffe species are not obvious, but they have been implicated – most notably the Rift Valley – in the divergence of other large mammals, including wildebeest (*Connochaetes taurinus*) [15] and impala (*Aepyceros melampus*) [16]. Thus, even though the ranges of giraffe species abut, dispersal might be limited by the steep topographical gradients of the Rift Valley and other habitat discontinuities associated with steep terrain. Dispersal limitation might be particularly strong among populations in the periphery of their distributions, and be present under either current or paleo-climate conditions. Given the mobility of giraffes, isolation-by-distance and geographic barriers are not strong candidates for the maintenance of divergence between the three giraffe species. For the sake of completeness, and to avoid bias by *a priori* ruling out any potential evolutionary process, we have nevertheless included both hypotheses in our analyses.

Hypothesis III – A third mechanism that might maintain reproductive isolation is divergent natural selection. Adaptation to local environmental conditions is increasingly viewed as a significant contributor to speciation (e.g., [2]). Habitat differences may reduce the fitness of dispersing individuals adapted to the habitat of the source population, resulting in population divergence, and ultimately leading to and maintaining reproductive isolation. Such divergence, often referred to as ecological speciation, may occur even in the face of ongoing gene flow [3,4,13,17,18]. Our third hypothesis focuses on spatially divergent general habitat conditions, but not differences in the timing of seasonal events. The latter is the focus of hypothesis IV.

Hypothesis IV – Finally, divergent natural selection can involve differential timing of reproduction [19,20]. Most known cases entailing temporal isolation are restricted to narrow biological interactions, such as evolutionary divergence through disparate timing of host plant phenology (e.g., [19]). It was previously hypothesized that temporally distinct regional rainfall cycles, which coincide with the availability of high-quality browse, impose divergent selection regimes on reproductive timing in giraffes [7]. The synchronization of weaning with the availability of fresh browse represents a possible means by which temporal reproductive isolation could be favoured. Such synchronization could benefit both offspring and mother by increasing growth rates, hastening weaning, limiting exposure of calves to predation, and offsetting the female's energy debt as a result of lactation.

In East Africa, three regionally distinct seasonal cycles of precipitation correlate with the timing of green-up [21] (Fig. 1), when fresh browse becomes available. Peaks in precipitation in this region follow the season(s) of maximal insolation, shifting latitudinally during the year with the intertropical convergence, and producing regionally distinct rainfall patterns [21,22]: 1) north of the equator, from northwestern Kenya through Uganda, July and August are the wettest months following the northern hemisphere summer solstice; 2) south of the equator, from southwestern Kenya through Tanzania, the rainy season occurs during southern hemisphere summer (December-March); 3) eastern Kenya, Somalia and Ethiopia experience bimodal precipitation, with maxima in spring (April-May) and fall (October-November), following maximal equatorial solar heating during the equinoxes. These regions generally correspond with the ranges of Rothschild's, Masai, and Reticulated giraffes respectively (Fig. 1). The Rothschild's giraffe was historically found in Uganda and Western Kenya [12]. The range of the Masai giraffe extends north through the Serengeti Plains and Masai lands up into Kenya, east to Mount Kilimanjaro, south to the Rufizi River, and west to Lake Rukwa and Lake Tanganyika. Finally, Reticulated giraffes occur from the Loroghi Mountains, the Barta Steppes, and Lake Turkana in the west to the Webi Shelbi River and the mountains of Ethiopia in the north, the dry coastal regions of Somalia in the east, and the Tana River in the south.

We tested how well each of the above hypotheses distinguishes between the three giraffe taxa using both non-spatially explicit and spatially explicit approaches. Because the more traditional methods to investigate associations between group membership and explanatory variables are non-spatial in nature, we start by focusing on environmental differences and differences in the timing of the seasons in a non-spatial context. Subsequently, we use more complex models that can specifically take into account the spatial relationships of populations as well as population connectivity.



**Figure 1. Spatial distribution of the day of the year (DOY) that green-up starts and giraffe point localities.** Colors represent the day of the year that green-up starts. In some areas there are two seasonal cycles of rainfall and associated green-up. The start of the first cycle is shown in panel (A), and of the second cycle in panel (B). Point localities of genotyped giraffe samples are plotted in triangles (Rothschild's), asterisks (Reticulated), and pluses (Masai). doi:10.1371/journal.pone.0077191.g001

## Materials and Methods

### Environmental variables

To capture the spatial distribution of parameters that are potentially useful in describing the giraffes' local habitat conditions, including those that relate to vegetation phenology, vegetation density, surface moisture, and topography, we used WorldClim climate data [23] as well as a suite of optical and microwave remote sensing data and derived products (Table 1). WorldClim bioclimatic metrics (WorldClim version 1.4 [23]) are derived from monthly temperature and rainfall climatologies [24] and are commonly used in characterizing habitat. They included eleven temperature and eight precipitation metrics, expressing spatial variations in annual means, standard deviations and extreme or limiting climatic factors. We checked for covariance among variables in our study area, and only included those with Pearson's correlations smaller than 0.9, resulting in a set of nine climate variables that were used in subsequent analyses (Table 1). We used this relatively high cutoff in order not to a priori rule out

potential small but significant additive effects of correlated variables. To study the effect of temporal differences in rainfall patterns in more detail, we used the monthly climatologies from the WorldClim database [23], and calculated monthly rainfall as percentages of total annual precipitation, which will be referred to by 'monthly rainfall' and the name of the month in the remainder of this paper.

Based on Moderate Resolution Imaging Spectroradiometer (MODIS) measurements on board of NASA's TERRA and AQUA satellites, we used the vegetation continuous field (VCF) product as a measure of the percentage of tree canopy cover [25], the Global Land Cover Dynamics product for vegetation phenology [21] and the leaf area index (LAI) product for vegetation density [26]. The spatial resolutions of these products based on optical passive measurements are 1km for leaf area index and vegetation phenology and 500 m for tree cover. To facilitate analysis, we aggregated the 500m native tree cover data to 1 km. The phenology fields capture the dates of onsets of green-up and dormancy of vegetation growing season cycles, and the algorithm was provided with the MODIS-based 16-day enhanced vegetation index (EVI) time series of the year 2001 to extract the respective dates [21]. To reduce processing and computation time, only one year of MODIS data was used. As a result, the vegetation phenology product has a considerable number of missing data points due to residual cloud cover. LAI is defined as the one-sided green leaf area per unit ground area. We averaged monthly LAI fields (Version 4) from the years 2000 to 2004 in order to reduce effects of residual cloud contamination [26] along with any natural inter-annual variability present in the data. The climatological monthly LAI composites were then used to generate three metrics: *LAI annual maximum* (LAI<sub>max</sub>), *LAI annual minimum* (LAI<sub>min</sub>), and *LAI annual range* (difference of maximum and minimum; LAI<sub>range</sub>). These LAI metrics provide spatial information on vegetation density.

In addition to these optical remote sensing products, we included microwave QSCAT data available in three-day composites at 2.25 km resolution [27]. Data of the year 2001 were used to create average monthly composites at 1 km resolution and then further processed to produce two metrics that included annual mean and standard deviation of radar backscatter at horizontal polarizations. The QSCAT radar measurements, at wavelengths of ~2 cm, are sensitive to surface canopy roughness, surface canopy moisture, and other seasonal attributes, such as deciduousness of vegetation [28]. For low density vegetation cover, such as woodlands, shrublands, and grassland savannas, the radar backscatter increases with increasing vegetation biomass and surface moisture [29]. Finally, for topography we used the Shuttle Radar Topography Mission (SRTM; <http://www2.jpl.nasa.gov/srtm/>) digital elevation data, aggregated from the native 90 m resolution to 1 km (available from the WorldClim group [23] at <http://www.worldclim.org>).

### Giraffe genetic and locality data

Giraffes ( $n = 429$ ) from 51 locations throughout the ranges of the three focal species (Fig. 1) were collected and genetically typed for 14 microsatellite loci for a previously published study [7]. Sample collection, DNA extraction, and microsatellite analyses are fully described in [7]. Briefly, DNA was extracted from skin biopsies for microsatellite typing on an ABI 377 or 3100 (Applied Biosystems, Inc; Foster City, CA, USA). Fragment lengths were scored using GeneScan and checked for errors using MICRO-CHECKER 2.2.3 [30] and MSA 4.0 [31]. Nei's D and F<sub>st</sub> between sampling sites were computed in Genalex 6 [32]. Genetic clusters were identified using Nei's D in POPULATIONS 1.2.28

**Table 1.** Overview of the predictor variables used in this study.

Data Record	Instrument	Variables derived	Ecological attributes
Leaf Area Index (LAI) ‡	Satellite-MODIS	LAlmax	Annual maximum
		LAlmin	Annual minimum
		LAlrange	Annual range (LAlmax – LAlmin)
		Treecover	Forest cover
Percent Tree Cover §	Satellite-MODIS	Treecover	Forest cover
Scatterometer-Backscatter †	Satellite-QSCAT	QScatMean	Annual mean surface moisture
		QScatstd	Standard deviation of surface moisture within a year
DEM	SpaceShuttle-SRTM	SRTM	Elevation
		SRTMsd	Elevation standard deviation (ruggedness)
		cost distances (CD)*	Permeability of habitat matrix based on elevation and ruggedness of the terrain
WorldClim ¶	Station-network	Bio1	Annual mean temperature
		Bio2	Mean diurnal temperature range
		Bio4	Temperature seasonality (standard deviation)
		Bio5	Maximum temperature of warmest month
		Bio6	Minimum temperature of coldest month
		Bio12	Annual mean rainfall
		Bio15	Rainfall seasonality (coefficient of variation)
		Bio16	Rainfall of driest quarter
		Bio17	Rainfall of wettest quarter
NDVI **	Satellite-AVHRR	Jan-Dec	Monthly rainfall as percentage of yearly total
		NDVImean	Annual mean vegetation greenness
		NDVIgreen	Greenness during greenest season
		Green-up	Day of year green-up starts
Distance			Geographic distance among sampling sites
Human population density	LandScan Global Population Database	Cost distances (CD)*	Permeability of habitat matrix based on human disturbance

Data at native resolutions smaller or larger than 1km have been aggregated to 1km.

†QSCAT annual mean and standard deviation are based on monthly data from the year 2001 with complete data coverage.

‡LAlmax, LAlmin, and LAlrange are derived from monthly mean values based on the first 5 year of MODIS data (2000–2004 [26]).

§Percent Tree Cover is based on MODIS data from 2001 [25].

¶WorldClim data are based on monthly climatologies from 1950–2000 [23].

\*Cost distances are computed either as Leas-Cost-Paths [48] or resistance distances [49].

\*\*See [21].

doi:10.1371/journal.pone.0077191.t001

(<http://bioinformatics.org/~typhon/populations>) and Bayesian clustering in STRUCTURE [33].

### Canonical variate analysis

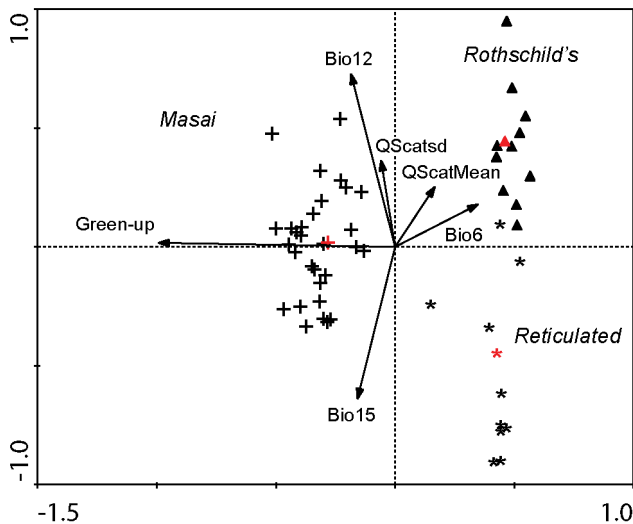
To test the hypothesis that the giraffe taxa occupy different habitats, a canonical variate analysis (CVA) was performed with CANOCO 4.5 [34]. We used genetic cluster membership (see also Figs. 2–3 for samples typed in [7]) to define the corresponding species at each sampling locality. For each species, a site was coded 0 if the species was absent and 1 if it was present. We then performed a canonical correspondence analysis (CCA), which is effectively a CVA with our data design [35]. A permutation test-based forward selection procedure was implemented to identify from the candidate set of environmental variables from WorldClim [23] and satellite remote sensing, those variables that best described habitat differences between the three taxa. We performed the forward selection procedure iteratively in order to exclude environmental variables that were highly correlated [36].

For a given iteration of the procedure, when a variable that was selected had a correlation coefficient  $r > 0.75$  with any of the previously entered variables, we excluded that variable, and re-ran the CVA. We re-ran the forward selection procedure in this manner until additional variables did not provide a significant improvement to the model. Bivariate correlations used to exclude variables were computed at 1000 random points throughout the study area. Significance tests on variables and on ordination axes employed 5,000 permutations each.

### Random forest models

To further test whether differences in the timing of rainfall could effectively differentiate among the three giraffe taxa as defined by genetic cluster membership, and to assess their importance relative to environmental variables that do not represent timing of seasons, we used random forest algorithms (randomForest v.4.5–30 [37]) as implemented in the R statistical framework (R Development Core 2009).





**Figure 2. CVA ordination plot.** Taxon centroids are in red; crosses = Masai; asterisks = Reticulated; triangles = Rothschild's; and vectors of environmental variables. Longer arrows indicate stronger contributions to the model, and their directions indicate degree of correlation with an axis. The first two axes explain 76.8% of taxon variation in environment. Bio6 = minimum temperature of the coldest month; Bio12 = annual precipitation; Bio15 = rainfall seasonality (coefficient of variation); green-up = the day of the year of the onset of green-up; QScatMean = surface moisture (QSCAT); QScatSd = QSCAT standard deviation. See Table 1 and Methods for a full description of the environmental variables.

doi:10.1371/journal.pone.0077191.g002

Classification tree models [38] implement binary recursive partitioning procedures to measure the amount of variation in a response explained by each predictor used in the model. *No a priori* assumptions are made about the relationship between predictor and response variables, allowing for the possibility of non-linear relationships with complex interactions. Homogeneity is measured by the Gini index [39], and splitting continues until further partitioning does not reduce the Gini index. Random forest methods incorporate a large number of these tree regressions, and for each tree constructed, use a random subset of the samples – the so-called bagging. Those samples not used in tree construction (the out-of-bag samples) are then tested against the random forest model, and error rates are computed across all runs to produce an estimate of classification error for the entire model [40]. Variable importance in random forest models is assessed by random subsampling of the variables and construction of new trees based on these predictor variable subsets.

First, we computed a random forest model using all predictor variables, including monthly rainfall (Jan-Dec) and bioclimatic and remote sensing variables. After verification that only monthly rainfall data were important contributors to the model, we computed a random forest model with only monthly rainfall variables, which was used in the subsequent predictive step. Random forest models were run with 20,000 trees (`ntree = 20,000`), variable importance was computed (`importance = TRUE`), and default settings for the remaining parameters were used.

An imbalance in the number of records within a class (here, the number of sites where each species was identified) can bias random forest predictions, and cause high error rates in the classification of the rare class (e.g., [41,42]). This phenomenon can be seen most frequently when imbalances of several orders of magnitude exist.

Our dataset is imbalanced at a 1:1.1:3 (Reticulated: Rothschild's: Masai) ratio, but we nevertheless explored the iterative down-sampling approach developed by Evans and Cushman [41]. We generated 1000 random forest models with random subsamples of the largest class to a sample size of 10 and compared the average out-of-bag error rate and variable importance scores to the model run without subsampling.

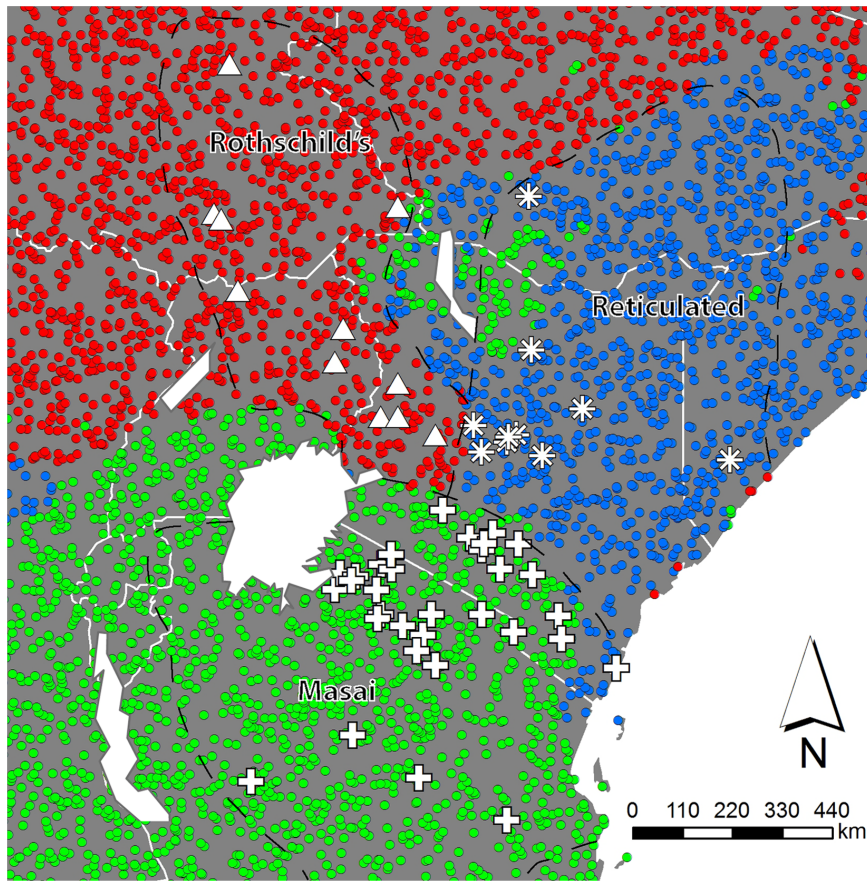
Although random forests can be used with a large number of predictor variables, the out-of-bag error rate can be increased when many variables are included that do not contribute to explaining the response variable. To minimize the out-of-bag error rate while at the same time minimizing the number of predictor variables included, Murphy et al. [43] developed a new variable selection procedure that uses the variable importance scores ( $I$ ) given as an output from random forests. For each variable  $n$ , its variable importance score  $I_n$  is divided by the maximum variable importance score  $I_{max}$ , resulting in a model improvement ratio (MIR) ranging between 0 and 1. In subsequent random forest models, variables with MIR smaller than a set threshold are withheld from the predictor variable set, and the resulting out-of-bag error rate is compared to that of the full model. We iteratively removed variables below MIR thresholds of 0-1at increments of 0.1.

Highly correlated predictor variables can potentially bias random forest results (e.g., [44–46]). To evaluate whether such a bias might be present in our results using the randomForest package, we confirmed variable importance using a conditional inference random forest algorithm implemented in the R package 'party v. 1.0–6'. Conditional variable importance randomly shuffles the values of the predictor variable and computes a new model with the shuffled values. This new model is then compared to the one where the values were not shuffled. The difference in accuracy of the two models is indicative of the variable importance. Conditional random forest runs were run for 20,000 trees; remaining parameters were left at their default values.

To create a spatially explicit prediction of taxon distributions based on our random forest model, we extracted the values of the twelve monthly rainfall variables for 10,000 random points in a wide range of our study area. We used these points and our random forest model for only monthly rainfall data to predict ('predict' function in randomForest package) which taxon would be present at each of the 10,000 randomly drawn locations. The results were plotted on a map of the study area and compared visually to species distribution maps of the three taxa to evaluate the concordance between predicted and observed ranges. Because the available species distribution maps are rather crude approximations of the true ranges, we did not compute a percent overlap between predicted and observed ranges.

### Generalized dissimilarity models

*A priori* assignment of populations into one of the three study taxa would appear justified, because genetic evidence suggests clear divergence among those groups [7]. However, to further assess a potential bias of such an approach, and to explicitly assess the effects of isolation-by-distance or dispersal barriers as well as those of environmental heterogeneity, we also implemented a distance-based methodology using generalized dissimilarity modelling (GDM [47]). GDM is an expansion on matrix regression techniques to relate dissimilarities in predictor variables to dissimilarities in response variables, and make spatially explicit predictions of the predictor-response relationship into areas that have not been sampled. An advantage of GDM over other modelling methodologies is that it can explicitly take into account the influence of geographic distance and dispersal barriers on



**Figure 3. Results for random forest prediction.** A random forest model based on taxon discrimination by monthly rainfall (Jan-Dec) was used to predict which taxon occurs at each of 10,000 randomly selected locations in the study area (coloured dots). Observed localities of the giraffe taxa are plotted in triangles (Rothschild's), asterisks (Reticulated), and pluses (Masai). Predicted taxon localities are indicated in red (Rothschild's), blue (Reticulated), and green (Masai). Approximate species ranges are indicated by dashed lines and their respective names (after [56]). doi:10.1371/journal.pone.0077191.g003

explaining biological variation, and allows for modelling variables that are difficult to define at individual sampling locations, such as genetic markers [47]. It can fit non-linear relationships of environmental variables to biological variation through the use of I-spline basis functions [47]. It is a two-step method: first, dissimilarities of a set of predictor variables are fitted to the genetic or phenotypic dissimilarities (the response variables). In an iterative process, predictor variables are added to and removed from the model, and only the variables that significantly improve the model are retained. Specifically, predictor variables are introduced to the model in random order and the variation in the response variable explained by the inclusion of that variable is compared to that without the variable ( $\Delta D$ ). Next, over many iterations the predictor variable is added again, but with the values randomized among sampling sites, resulting in a random distribution of  $\Delta D_{\text{rand}}$ .  $\Delta D$  is compared to  $\Delta D_{\text{rand}}$ , based on which the predictor variable is either retained or dropped. Generalized dissimilarity models were run using an Avenue script in ArcView v 3.2 in conjunction with a SPlus v 4 script obtained from the authors of GDM [47].

To assess the level of population divergence, we used genetic distances (both  $F_{st}$  and Nei's  $D$  values) among the sixteen populations from our study taxa that were typed for fourteen microsatellite loci [7]. To characterize the regional timing of rainfall, we used the dissimilarity in monthly rainfall variables among sampling sites. In addition, to test for the influence of

isolation-by-distance and dispersal barriers, we included geographic distance and least-cost-paths [48] or resistance distances [49]. Least-cost distances take into account spatial heterogeneity in permeability of habitats for dispersal. Least-cost-path and resistance distances were computed in Pathmatrix 1.1 [48] and Circuitscape 2.2 [49] respectively from friction surfaces that represented two types of barriers. First, giraffes generally do not occur higher than 2000 m above sea level [9], or in steep terrain. A friction surface representing potential ancient barriers was, therefore, based upon altitude and ruggedness (SRTM and SRTMsd respectively; see Table 1) of the terrain, which captured potential dispersal barriers formed by mountainous areas (e.g. regions in and along the Rift Valley). Values for SRTMsd ranged between  $\sim 1$  and  $\sim 400$ , and were directly used as friction values in computations of cost distances. We similarly coded areas above 2000 m in altitude as 400, and those below as 1. Cost distances were then computed for altitude and ruggedness separately. We also added the values of the two friction surfaces for computation of a single cost distance matrix. Thus, areas above 2000 m in altitude, with the maximum level of ruggedness were  $\sim 400$  times as difficult for dispersal by giraffes as level areas below 2000 m. Because the assignment of costs is relatively arbitrary, we also computed cost surfaces for altitude and altitude+ruggedness where all values  $> 1$  (i.e. the minimum cost assigned to a grid cell) were divided by 10 and where those values were multiplied by 10. Thus, we computed the following cost distances: 1) ruggedness (untrans-

formed cost surface); and 2) altitude and altitude+ruggedness for cost surfaces where the difference between the minimum and maximum cost was one, two, or three orders of magnitude.

The second friction surface represents more recent habitat changes by humans, and was computed directly from human population densities in East Africa (LandScan™ Global Population Database. Oak Ridge, TN: Oak Ridge National Laboratory. Available from <http://www.ornl.gov/landscan>). Although the relation between human population density and anthropogenic pressure varies from one region to the other (e.g., [50]), expanding human populations and increasing population densities may be proxies for land use changes (e.g., [51]) and other human-wildlife conflicts, such as cattle grazing (e.g., [52]), that affect the dispersal ability of giraffes. Because microsatellites evolve relatively rapidly, they may contain a signal of population divergence mediated by potential recent dispersal barriers resulting from anthropogenic land-use changes under the assumption of a generation time of approximately four years [7].

As a means to contrast the hypotheses regarding seasonal timing of rainfall, dispersal barriers, and geographic distance, we ran six models with different sets of predictor variables for both *Fst* and Nei's *D*. In two of these models all predictor variables were entered (full model), with cost distances based on either altitude or population density. Importance of any of those variables in a model would implicate its role in maintaining divergence among the three giraffe taxa. To evaluate cross-correlations among predictor variables, in the remaining four models the following subsets of the predictor variables were entered: only monthly precipitation variables, only geographic distance, or only one of the two cost distances. The percentages of the variation explained by each model were compared to assess which parameter set best explained the observed genetic variation.

### Frequency of giraffe births

To further investigate hypothesis IV (regional differences in the timing of rainfall), we assessed whether regional differences in the timing of green-up are related to reproductive timing in giraffes. Unfortunately, data on calving times in giraffes is largely lacking for our study taxa, and only available for two distinct genetic groups of Masai giraffe [7]: those in Nairobi National Park [53] and those in the Serengeti [54]. We first tested whether calving times conformed to a uniform distribution in each area, using Shapiro Wilks and Anderson-Darling tests. In addition, we tested whether the timing of births differs between these two regions using an autocorrelation analysis (*acf* function in the R Statistical Package), and compared this to the timing of maximum rainfall in each region.

## Results

### Tests of associations between predictor variables and giraffe divergence

We first assessed the roles of differences in the timing of green-up as well as general habitat differences in differentiating the three giraffe taxa in East Africa, by means of a canonical variate analysis (CVA) on a set of environmental variables pertaining to 51 locations where giraffe samples have been genetically typed (Fig. 1) [7]. The regional differences in rainfall maxima result in opposing seasons of green-up in the north and south of the region (Fig. 1a) and a second green-up in October/November in the north-eastern Kenya, eastern Ethiopia, and Somalia (Fig. 1b). Timing of first green-up, climate variables (Bio1-Bio17), and a number of satellite-derived ecological variables were included in the analysis (see Material and Methods and Table 2). The first and second

CVA axes explained 47.4% and 29.4% of the variation among taxa respectively ( $F = 24.315$ ,  $p = 0.0002$ ). The first axis discriminated the Masai giraffe on the basis of first green-up (axis 1 vs. green-up, weighted  $r = -0.97$ ; Table 3; Fig. 2). The second axis differentiated the Rothschild's and Reticulated giraffes, and was positively correlated with annual precipitation, and negatively correlated with Bio15 (precipitation seasonality not indicating timing, i.e. coefficient of variation; Table 3). This result strongly implicated the first green-up in differentiating the Masai giraffe. However, missing satellite data (Fig. 1) precluded analysis of the second/autumn green-up in much of Somalia, Ethiopia, and Kenya where we predicted it should differentiate the Reticulated giraffe.

To extend the analysis to the second green-up, and to further investigate the timing component of the annual precipitation cycle that drives green-up in the three regions, we generated a monthly precipitation dataset, and used these variables (Jan-Dec) in addition to the satellite remote sensing and climate variables that capture general habitat characteristics to construct a random forest model [40,55] (Table 1 and 2). Under this random forest model, most sampling localities were classified in their expected taxonomic group. The out-of-bag error rate was 3.8%, meaning that on average <1 locality showed a mismatch between observed and predicted taxonomic grouping. Out-of-bag error rate increased to 6.45% when we applied down-sampling of the largest class (Masai giraffe), where all Masai sites were correctly classified, and one Rothschild's and one Reticulated site were misclassified. However, after applying MIR to select the smallest set of variables that minimized the out-of-bag error rate, all sites were correctly classified (out-of-bag error rate = 0%). The most important variables in explaining differentiation among taxa were qualitatively similar between runs where we did and did not apply down-sampling, and consisted of monthly rainfall in February and October, followed by March, August, July, and April (Table 4). These were also the variables retained after applying MIR. Moreover, the first five of those variables (February, October, March, August, July) were also the most important variables under the conditional inference variable importance criterion (Table 4). Rainfall measures in February and March, and in July and August are highly correlated ( $R^2 < 0.95$ ), but this is of little to no influence on our random forest models. In each regression tree only one of the two correlated variables is picked as the most important variable. The presented importance scores are a summary of many tree regressions, and are an indication of how often each variable is used in a regression tree. Months known to be important in discriminating regional climate proved to be informative in the random forest model: February is associated with maximum precipitation in southern Kenya and Tanzania; July and August with maximum precipitation in north-western Kenya and Uganda; and March and October correspond to the post-equinoxal precipitation in north-eastern Kenya, Somalia, and eastern Ethiopia. Remote sensing and climate variables that do not capture the timing of seasons were relatively unimportant in the random forest model (Table 4), suggesting that general habitat differences alone cannot explain the observed taxonomic differentiation among giraffes in East Africa. A predictive map of the spatial distribution of our study taxa based on our random forest model corresponds with known taxon distributions [56] (Fig. 3). The one major inaccuracy is a prediction of Masai further north, in between Reticulated and Rothschild's predictions. This is an area in which many species distribution maps show a gap in giraffe occurrence (e.g., [56]). In an additional random forest model, we also considered the subdivision of Masai giraffe into two distinct units in the region, as suggested by molecular data [7], totalling



**Table 2.** Overview of analyses conducted and hypotheses tested.

Analysis	Response variable	Predictor variables entered	Hypotheses tested
CVA	Taxon membership	LAI, Treecover, QSCAT, SRTM, Bio1-17, NDVI, Green-up	General habitat vs timing of green-up
RF	Taxon membership	Step 1: Same as CVA plus Jan-Dec Step 2: Jan-Dec	General habitat vs timing of rainfall Timing of rainfall
GDM F	Genetic distance §	Jan-Dec, cost distances*, distance	Timing of rainfall vs barriers vs distance
GDM E	Genetic distance §	Jan-Dec	Timing of rainfall
GDM D	Genetic distance §	Distance	Distance
GDM CD	Genetic distance §	Cost distances*	Barriers

§Genetic distances were computed as  $F_{st}$  and Nei's D from microsatellite data.

\*Cost distances include those based on elevation+ruggedness and human population density (see Table 1 and Material and Methods).

doi:10.1371/journal.pone.0077191.t002

four genetic entities in East Africa. The results of this model are comparable to that for the three giraffe taxa: an out-of-bag error rate of 3.7% and high importance of monthly rainfall variables, suggesting that seasonal timing of rainfall can also distinguish between smaller genetic entities.

To explicitly assess the effects of isolation-by-distance or dispersal barriers as well as those of environmental heterogeneity, we also implemented a distance-based methodology using generalized dissimilarity modeling (GDM [47]). The full models for  $F_{st}$  values explained about 60% of the total observed variation, and those for Nei's D approximately 80% (Table 5). Monthly precipitation variables, in particular July and February, were consistently the most important in our models (Table 5; Fig. S1). The cost distances and geographic distance were also significant, but generally contributed far less to explaining the observed variation than precipitation variables. The only exception to this was observed in the model for  $F_{st}$  values with the cost distance based on human population density. Here, the cost distance was the second most important variable in the model, after July (Table 5; Fig. S1). In addition, models based only on geographic distance or the cost distances explained approximately 22%–78% less of the total genetic variation than the full models (Table 5). Whether the models with altitude or altitude+ruggedness cost distances were based on friction surfaces with one, two, or three orders of magnitude difference between low and high cost grid cells, made only negligible difference for the full models. However, for the cost-distance-only models (CD), the total variation explained ranged between 0.3% and 31.7%, the latter approximating that of the geographic-distance-only model (D) (Table 5).

In comparison, models based only on precipitation variables performed nearly as well as the full models (Table 5). While rainfall values in some subsequent months are correlated, cross-correlated months are only included in the models if they have additive explanatory power. The interpretation of the selected months should, however, be general with respect to the timing of seasons, without assignment of any individual weight to cross-correlated months. The results from generalized dissimilarity models suggest that: 1) differences in timing of rainfall are important in discriminating among the three studied taxa; 2) dispersal barriers—in particular those imposed by human habitation—may have resulted in recent differentiation; and 3) isolation-by-distance played a relatively minor role in divergence among taxa.

### Frequency of giraffe births

For differences in the timing of seasons to have biological meaning with respect to reproduction, calving times should also show differences between taxa. Tests of the null model of a uniform distribution of births across the year using the frequency of calving times from two distinct genetic groups within Masai giraffe [7] in Nairobi National Park [53] and the Serengeti [54] rejected a uniform distribution in both areas (Nairobi NP birth peak in August, September and one in January: Shapiro Wilks test,  $p=0.0019$ ; Anderson-Darling test,  $p=0.0026$ ; Serengeti birth peak in May-July: Shapiro Wilks test,  $p=0.0205$ ; Anderson-Darling test,  $p=0.0191$ ). In addition, an autocorrelation analysis (acf function in the R Statistical Package), testing the lag time

**Table 3.** Correlations between environmental variables used in the CVA analysis and the first two taxon ordination axes.

	Ax 1	Ax 2	Green-up	Bio6	Bio12	Bio15	QScatMean	QScatsd
Ax 1	1							
Ax 2	0	1						
Green-up	−0.97	0.01	1					
Bio6	0.34	0.14	−0.38	1				
Bio12	−0.18	0.56	0.17	−0.20	1			
Bio15	−0.15	−0.49	0.19	0.13	−0.50	1		
QScatMean	0.16	0.19	−0.16	−0.00	0.39	−0.04	1	
QScatsd	−0.06	0.28	0.13	0.14	0.09	0.20	−0.42	1

Bio6 = minimum temperature of the coldest month; Bio12 = annual precipitation; Bio15 = rainfall seasonality (coefficient of variation); Green-up = day of the year of the onset of green-up; QScatMean = surface moisture (QSCAT); QScatsd = QSCAT standard deviation. See Table 1 and Materials and Methods for a full description of the environmental variables.

doi:10.1371/journal.pone.0077191.t003

**Table 4.** Random forest model results.

Predictor variable	Mean decrease in accuracy	Mean decrease in Gini index	Conditional variable importance
<b>Feb</b>	<b>16.14</b>	<b>4.2127</b>	0.0219*
<b>Oct</b>	<b>15.87</b>	<b>4.2656</b>	0.0719*
<b>Mar</b>	<b>14.58</b>	<b>3.4089</b>	0.0273*
<b>Aug</b>	<b>12.78</b>	<b>2.0622</b>	0.0422*
<b>Jul</b>	<b>12.43</b>	<b>2.0542</b>	0.0422*
<b>Apr</b>	<b>11.45</b>	<b>1.4400</b>	0.0124
Dec	11.07	1.6265	0.0129
Jan	11.02	1.6585	0.0182
Nov	9.19	0.9415	0.0056
Jun	8.42	0.9905	0.0209
LAlrange	8.13	1.0316	0.0143
Sep	8.03	0.8364	0.0084
NDVIgreen	7.13	0.7951	0.0158
Bio12	6.84	0.6681	0.0031
Bio16	6.18	0.6359	0.0048
NDVImean	5.44	0.5143	0.0051
Bio15	5.29	0.3559	0.0006
LAlmax	4.91	0.4031	0.0065
Bio5	4.19	0.2783	0.0018
Bio4	2.26	0.0801	0.0000
May	1.74	0.1258	0.0000
Bio1	1.64	0.1204	0.0002
LAlmin	1.39	0.1287	0.0065
Bio6	1.37	0.1285	0.0000
Bio17	0.89	0.0621	0.0000
QScatsd	0.82	0.1080	0.0001
QScatMean	0.42	0.0443	0.0000

Higher values of the “mean decrease in accuracy” and the “mean decrease in Gini index” indicate higher predictor variable importance. Variables in bold are the ones included in the random forest model that minimizes the number of variables used as well as the out-of-bag error rate after applying the model improvement ratio approach (see Material and Methods). Conditional inference variable importance is shown for a conditional inference random forest model, which corrects for potential biases due to correlations between predictor variables. Variables marked by \* are the five most important variables according to the conditional inference. The variables Jan-Dec represent the seasonal timing of rainfall; the remaining variables are representative of spatial differences in habitat. Also see Tables 1 and 2. doi:10.1371/journal.pone.0077191.t004

where maximum correlation is observed between the frequency distribution of calving times in populations in Nairobi National Park and the Serengeti, showed that the highest correlation was observed at a lag time of 2–3 months, suggesting that these genetic units show differences in the timing of births. This corresponds to a similar lag time in peak rainfall and green-up in these areas.

## Discussion

We have investigated the potential current geographic and environmental factors that may contribute to maintaining divergence between giraffe taxa in East Africa. The results of our simultaneous tests of hypotheses presented here suggest that, among the factors investigated, regional differences in timing of maximum rainfall are of primary importance. Even though general habitat differences, dispersal barriers, and isolation-by-distance also appear to contribute to inter-taxon differentiation, our analyses suggest they play less important roles.

Previous studies have shown that parapatric speciation is possible when dispersal distances are limited and genetic incompatibilities accumulate, resulting in reproductive isolation

of parapatric populations [13,14]. Indeed, even though giraffes are able to travel large distances [5], dispersal may be limited due to small home range sizes and responses to limited resource availability [54]. Moreover, genetic structure is apparent not only between, but also within species [7]. Nevertheless, generalized dissimilarity models in which geographic distance was entered as a predictor variable suggested that it did not play a major role in explaining genetic variation in a spatial context (Table 5).

Studies using a resistance surface to compute cost distances are often subject to an oversimplification of the relation between the habitat matrix and gene flow, and our study is no exception. To obtain a better model of gene flow as a function of the habitat matrix, Shirk et al. [57] developed a novel approach in which the functional relationship between habitat characteristics and resistance are varied, resulting in a series of cost distances that are correlated to a measure of gene flow. For multiple habitat characteristics, this is first done in a univariate procedure. Subsequently, the univariate optimal functions are used to start a multivariate optimization procedure to find the resistance surface best describing the observed genetic divergence between

**Table 5.** Generalized dissimilarity modelling results.

Genetic distance	Model	Cost distance entered in model	Percent of total variation explained (1/2/3 orders of magn.)*		Variables included in full model
Fst	F	alt+ruggedness	58.8/59.3/59.3		Jul, Feb, Jun, May, D, Oct, Sept, CD, Nov, Aug
		altitude	58.7/58.7/58.7		
		ruggedness	59.4		
		pop dens	60.9	Jul, CD, Feb, May, D, Aug, Oct, Nov	
	E		58.6		
	D		21.3		
	CD	alt+ruggedness	9.9/2.5/0.8		
		altitude	22.7/22.7/22.1		
		ruggedness	3.0		
		pop dens	38.8		
Nei's D	F	alt+ruggedness	79.7/80.1/79.8		Jul, Feb, Oct, CD, Nov, May, D, Aug
		altitude	79.6/79.5/79.5		
		ruggedness	79.9		
		pop dens	79.9	Jul, Feb, Oct, CD, Aug, D, Nov, May	
	E		79.3		
	D		31.2		
	CD	alt+ruggedness	13.4/2.1/0.3		
		altitude	31.7/30.1/30.0		
		ruggedness	2.7		
		pop dens	38.6		

Results shown are for six different models each on Fst and Nei's D genetic distances with monthly precipitation variables and geographic distance or cost-distances based on either altitude + ruggedness of the terrain or human population density. Variables entered in the six models were: full model (F: environment + distance + cost distance; E/D/CD); environment only (E); distance only (D); cost distance only (CD). Variables included in the models are only shown for the full models. CD = cost distance; D = geographic distance; E = environmental variables (i.e. monthly precipitation); pop dens = human population density.

\*For models where altitude and/or ruggedness of the terrain were entered, results are shown for cost distances based on one, two, or three orders of magnitude difference between suitable and unsuitable habitat. See Material and Methods for further details.

doi:10.1371/journal.pone.0077191.t005

populations. Even though the approach proposed by Shirk et al. [57] is a major improvement over approaches assigning arbitrary costs to certain habitat conditions, it depends heavily on detailed expert knowledge. First, the starting parameters in the optimization procedure are set based on expert knowledge, and second, the optimized parameter values should be evaluated to what extent they are biologically meaningful. Unfortunately, such expert knowledge is sparse for giraffes, which is why we chose not to follow the procedure outlined by Shirk et al. [57]. Our analyses of dispersal barriers resulting in a reduction of gene flow are, therefore, subject to limitations. Yet, our results are in line with those of Arctander et al. [15], who found low levels of divergence for several species across the Rift Valley, suggesting that the Rift Valley may not be a major topographical barrier to large mammals in general. It remains unclear to what extent human-induced barriers may influence divergence among giraffe populations and species. Generalized dissimilarity models of Fst values where human population density was included as a cost distance suggest that dispersal might be limited across areas with high human population densities. However, similar models for Nei's D were less conclusive in this respect. Given that human disturbance of the magnitude currently seen is a recent phenomenon, and giraffes have long generation times, the microsatellite data may not show a signature of reduced gene flow yet. Nevertheless, human occupancy does not appear to be related to habitat characteristics unfavourable to giraffe dispersal over longer evolutionary times.

Locally adaptive responses to divergent habitat could provide selective pressures favouring endemic populations over migrants or hybrids. The most significant habitat variables that are not related to the timing of the seasons in canonical variate analyses were Bio12 (annual precipitation) and Bio15 (coefficient of variation in annual rainfall). Rothschild's giraffe occupy habitat that generally gets more rainfall during the year than the regions occupied by Masai (medium levels of rainfall) and Reticulated giraffe (lowest levels of rainfall). Similarly, Rothschild's giraffe habitat is characterized by smaller seasonal differences in rainfall than that of Reticulated and Masai giraffe. The low levels of rainfall in Reticulated habitat result in the lowest observed vegetation cover during the greenest time of the year (NDVIgreen), as well as small differences in greenness between the seasons (LAIrange). Nevertheless, there is large overlap in these variables between the three giraffe species, and they are therefore not effective predictors of genetic divergence.

Although the association between patterns of green-up and giraffe taxa suggests a role for geographic differences in the seasonal timing of rainfall in the maintenance of differentiation, the historical processes of initial giraffe divergence remain an open question. The ~ 23,000 year precession cycle and associated ~ 100,000 year climate cycle strongly influence African precipitation and is known to have resulted in changes in precipitation intensity multiple times in the region since the early to middle Pleistocene [58–60]. This corresponds to the approximate time of the first split

between these giraffe taxa, that of the Masai and Reticulated+Rothschild's giraffes [7]. Discrete rainfall regimes with comparable seasonal attributes likely persisted throughout the late Pleistocene and early Holocene, albeit with significant variation in precipitation intensity and geographic extents [61]. Spatially and temporally differentiated, these rainfall patterns could have contributed to initial divergence of the three giraffe taxa. Alternatively, as a result of the precession cycles, savannah habitat has repeatedly expanded and contracted [62,63], a process that has been hypothesized to be the primary cause of divergence in several savannah mammals [9,64–71]. Under the hypothesis of periodic isolation in savannah refugia, followed by range expansions tracking the expansion of savannah during more favourable arid periods, it is plausible that the three giraffe taxa in East Africa initially diverged in allopatry, and remained distinct through one of the mechanisms described above. Studies of paleoclimatic conditions and habitat suitability, as well as detailed genetic and demographic studies should provide insight into the causal mechanisms of initial divergence of giraffes as well as other African savannah species. Finally, even though least-cost-path analyses that consider orographic features suggest that they do not impose significant dispersal barriers causing reduced gene flow under current conditions, they may have contributed to initial giraffe divergence under historical conditions, when these features may have been more severe, or – combined with paleo-climate conditions – may have harboured less suitable habitat conditions.

The striking correlation between seasonal timing of rainfall and genetic divergence among giraffe taxa might be explained by different, but not necessarily mutually exclusive mechanisms. First, genetic divergence might be related to a synchronization of the reproductive cycle with regional timing of rainfall. For available data, a uniform distribution of births across the year was rejected, which is supported by independent data from the Serengeti [72]. This suggests that calving times in giraffes display seasonality. This is substantiated by observations of seasonal peaks in births among various giraffe taxa throughout Africa [73–78]. For East Africa, data is only available for Masai giraffe in the two above mentioned areas. These areas represent genetically distinct entities within the Masai giraffe [7] that also experience distinctly different rainfall regimes. Our finding that populations in Nairobi National Park and the Serengeti show differences in birth timing, is consistent with their unique signatures of seasonal timing of rainfall, the latter of which was confirmed by the random forest model on the four genetically distinct groups. Similar differences in seasonal timing of births between Masai, Reticulated, and Rothschild's giraffe seem to be supported: the preponderance of births appears to occur in the dry season (Fig. 2 in [54], [79]) (Fig. S2), which is January through March for Rothschild's [9] and May through August for Masai giraffes [54,79].

Synchronization of the reproductive cycle with the timing of rainfall may involve selective advantages related to the condition of the female at the time of conception, increased growth rates and predation avoidance for calves, and quick recovery of the female at time of weaning. As a result of births peaking late in the dry season, weaning of giraffe young occurs at a time that fresh browse becomes available [80,81]. This synchronization may be beneficial to both offspring and mother. First, consumption of high-quality browse by weaning calves during green-up may result in increased growth rates, hasten weaning and thus limit exposure of calves to predation, which is a major cause of mortality in wild giraffes [80]; lion and hyena predation on calves can approach 50% [54,82–84]. Second, lactating females experience substantial costs minimizing predation risk [54], which could be offset by the abundance of browse. Females with young prefer open settings [85], sacrificing

foraging opportunities for lowered predation risk to their calves [54]. This form of predation avoidance likely contributes to the maternal energy debt experienced by lactating females late in the dry season [86]. This energy debt does not occur in non-lactating adults [79]. Hastened weaning of offspring should reduce these energetic costs for females, and the availability of high-quality browse for females during and following the most demanding phase of reproduction could mitigate the impact of reduced foraging time. The importance of the availability of high-quality browse is consistent with observations that lactating females show stronger preferences for tannin-free, high-protein-level browse associated with green-up than adult males or non-lactating females [72,79,81].

Peaking birth rates during the dry season [9,54,72,79] suggest that conception, or possibly implantation, may be most frequent during a narrow time interval [78]. Conception may be influenced by female condition at the time of mating, and related to resource availability, yet is probably also under some form of genetic or clocklike control, as is often the case in mammalian reproduction [87]. In giraffes, gestation time is roughly 14–16 months (e.g., [88–91]). Whereas giraffes reportedly mate throughout the year, based on the limited data available on births, the bulk of conceptions probably occur late in the wet season [78]. This may also be the time of optimal female condition due to the abundance of browse during the preceding period [78,92], allowing reproductive females to recover from the energy deficit generated during gestation and lactation. Thus, reproductive timing may provide benefits with regard to parturition and subsequent weaning, but also with respect to conception, and maximizes the condition of both female and offspring. This notion is supported by the fact that giraffe calf recruitment is positively correlated with late dry season precipitation (i.e. earlier than normal green-up) over the preceding five-year period [93]. Similar situations seem to occur in elephants [94] and African buffalo [95], where conception is tightly linked to higher levels of NDVI – a measure of vegetation greenness indicating the availability of browse (Table 1).

An alternative mechanism that might explain the relation between genetic divergence and seasonal timing of rainfall could involve seasonal variation in habitat use as a response to differences in the timing of maximum rainfall and the associated availability of browse. Resource tracking has clear selective advantages and is one of the likely underlying causes of migration on both small and large scales. Although giraffes are capable of travelling large distances [5], they often have small ranges and exhibit localized responses to seasonal variation in resource availability [54]. For instance, when rainfall peaks in one area, giraffes in that area may be able to travel large distances because of the widespread availability of browse. However, this will coincide with a dry period in the adjacent regions, where giraffes may be confined to small patches of habitat with sufficient resources. Such an effect could render populations geographically isolated. In addition, habitat preferences – which are suggested to be different among males and females [85] – may limit the effective ranges of individual giraffes or populations in a given season. Thus, small home ranges that track the availability of browse associated with local to regional differences in maximum rainfall could facilitate the isolation of giraffe populations through neutral evolutionary processes. Finally, exposure to specific rainfall regimes could increase a given individual's preference for those natal cues through natal habitat preference induction (NHPI; e.g., [96] and references therein). Dispersing individuals might preferentially disperse to areas with habitat characteristics similar to those in the natal habitat. Such a scenario is independent of selection, but relies on imprinting of habitat cues during early stages of development.

In our hypothesis of ecologically mediated maintenance of population divergence, differences in reproductive timing need not act alone. For instance, mate recognition mechanisms may also contribute to isolation. In this context, differences in pelage pattern may serve as visual cues in mate choice, possibly through imprinting on the conspecific pelage pattern during the early stages of life [7]. However, to our knowledge no field data exist on the use of pelage patterns in mate recognition. To better understand the detailed mechanisms of isolation in contact zones, further studies of mate choice and habitat use are needed.

## Conclusions

We have shown that among the predictive variables considered, regional differences in the seasonal timing of rainfall and the associated timing of green-up best discriminate among the three East African giraffe taxa, and that general habitat differences, dispersal barriers, and geographic distance do so less effectively. One explanation for this striking relation may be related to reproductive asynchrony, suggesting regional adaptation of the reproductive cycle to the differential timing of green-up. This scenario might represent a form of ecologically-mediated reproductive isolation consistent with a growing body of work that suggests that selection can produce or maintain the divergence between ecologically distinct groups [2,97]. Theory shows the efficacy of differential timing of the seasons or phenology on driving reproductive isolation of parapatric and sympatric populations [98], and studies of natural systems suggest that selection on timing of host plant flowering can lead to sympatric or micro-allopatric speciation of insects [19]. Similarly, our results might be explained by selection associated with timing of annual events, facilitating the maintenance of genetic and phenotypic divergence on regional scales in large, highly mobile animals. However, alternative mechanisms are also plausible. These might be related to differences in seasonal timing of rainfall, such as resource tracking and resulting seasonal allopatry, or to other factors, such as mate recognition based on pelage patterns. We

have described a striking correlation between spatially divergent timing of maximum rainfall and giraffe divergence, warranting further research to better understand the exact nature of the relation and its potential role in maintaining giraffe population divergence.

## Supporting Information

**Figure S1 Response curves for variables entered in generalized dissimilarity models that were selected in the model as important in explaining the observed variation.** The maximum value of each variable is indicative of its importance in the model. Response curves are shown for models of Nei's D (A, B) and Fst (C, D) genetic distances with a set of predictor variables consisting of monthly precipitation, geographic distance, and a cost distance based on either altitude and ruggedness (A, C) or human population density (B, D). (TIF)

**Figure S2 Monthly calving frequencies of giraffe in the Serengeti (solid line) and Nairobi National Parks (broken line).** Adapted from [54]. (TIF)

## Acknowledgments

We acknowledge Russ Seymour for compiling giraffe locality data. We thank John Pollinger for help in extraction and quality control of locality data, Ryan Harrigan for help in random forest methods, and Klaus Koepfli and Michael Patten for discussions on this manuscript. We thank S. Cushman and two anonymous reviewers for constructive comments on previous versions of this manuscript.

## Author Contributions

Conceived and designed the experiments: HAT DMB AHF DKJ. Performed the experiments: HAT DMB AHF. Analyzed the data: HAT AHF WB. Contributed reagents/materials/analysis tools: WB. Wrote the paper: HAT AHF DKJ.

## References

- Coyne JA, Orr AH (2004) Speciation. Sinauer Associates, Sunderland, MA. 545 p.
- McKinnon JS, Mori S, Blackman BK, David L, Kingsley DM, et al. (2004) Evidence for ecology's role in speciation. *Nature* 429: 294–298.
- Niemiller ML, Fitzpatrick BM, Miller BT (2008) Recent divergence-with-gene-flow in Tennessee cave salamanders (Plethodontidae; *Gyrinophyllus*) inferred from gene genealogies. *Mol Ecol* 17: 2258–2275.
- Nosil P (2008) Speciation with gene flow could be common. *Mol Ecol* 17: 2103–2106.
- Le Pendu Y, Ciofolo I (1999) Seasonal movements of giraffes in Niger. *J Tropical Ecol* 15: 341–353.
- van der Jeugd HP, Prins HHT (2000) Movements and group structure of giraffe (*Giraffa camelopardalis*) in Lake Manyara National Park, Tanzania. *J Zool London* 251: 15–21.
- Brown DM, Brenneman R, Koepfli K-P, Pollinger JP, Milá B, et al. (2007) Extensive population genetic structure in the giraffe. *BMC Biol* Doi:10.1186/1741-7007-5-57.
- Groves C, Grubb P (2011) Ungulate Taxonomy. Baltimore: The John Hopkins University Press. 336 p.
- Kingdon J (1979) East African mammals: an atlas of evolution in Africa Volume 3B. The University of Chicago Press, Chicago, IL.
- White F (1983) The vegetation of Africa: a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa. UNESCO.
- Ansell WF (1971) Family Giraffidae. – In: Meester J, Setzer HW, editors. The mammals of Africa: an identification manual. Washington, DC: Smithsonian Institution Press. 1–84.
- Dagg AL, Foster JB (1982) The giraffe: its biology, behavior, and ecology. Melbourne, FL: Krieger Publishing Company.
- Gavrilets S, Li H, Vose MD (2000) Patterns of parapatric speciation. *Evolution* 54: 1126–1134.
- Hoelzer GA, Drewes R, Meier J, Doursat R (2008) Isolation-by-distance and outbreeding depression are sufficient to drive parapatric speciation in the absence of environmental influences. *PLoS Computational Biol* 4: e1000126.
- Arctander P, Johansen C, Coutellec-Vreto M-A (1999) Phylogeography of three closely related African bovines (Tribe Alcelaphini). *Mol Biol Evol* 16: 1724–1739.
- Templeton AR, Georgiadis NJ (1996) A landscape approach to conservation genetics: conserving evolutionary processes in the African Bovidae. In: Avise JC, Hamrick JL, editors. Conservation genetics: case histories from nature. New York: Chapman and Hall. 398–430.
- Gavrilets S (2000) Waiting time to parapatric speciation. *Proc Roy Soc B* 267: 2483–2492.
- de León LF, Bermingham E, Podos J, Hendry AP (2010) Divergence with gene flow as facilitated by ecological differences: within-island variation in Darwin's finches. *Phil Trans Roy Soc B* 365: 1041–1052.
- Feder JL, Hunt TA, Bush L (1993) The effects of climate, host plant phenology and host fidelity on the genetics of apple and hawthorn infesting races of *Rhagoletis pomonella*. *Entomol Exp Appl* 69: 117–135.
- Yamamoto S, Sota T (2009) Incipient allochronic speciation by climatic disruption of the reproductive period. *Proc Roy Soc B* 276: 2711–2719.
- Zhang XY, Friedl MA, Schaaf CB, Strahler AH, Liu Z (2005) Monitoring the response of vegetation phenology to precipitation in Africa by coupling MODIS and TRMM instruments. *J Geophys Res (Atmospheres)* 110: D12103, doi:10.1029/2004JD005263.
- Griffiths JF (1972) Eastern Africa. In: Griffiths JF, editor. *Climates of Africa*. Amsterdam: Elsevier. 313–347.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high-resolution interpolated climate surfaces for global land areas. *Internat J Climatol* 25: 1965–1978.
- Nix H (1986) A biogeographic analysis of Australian elapid snakes. Atlas of Elapid Snakes of Australia. Australian Government Publishing Service.
- Hansen MC, DeFries RS, Townshend JRG, Sohlberg RA, Dimicelli C, et al. (2002) Towards an operational MODIS continuous field of percent tree cover



- algorithm: Examples using AVHRR and MODIS data. *Remote Sens Environ* 83: 303–319.
26. Myneni RB, Hoffman S, Knyazikhin Y, Privette JL, Glassy J, et al. (2002) Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sens Environ* 83: 214–231.
  27. Long DG, Drinkwater MR, Holt B, Saatchi S, Bertoia C (2001) Global Ice and Land Climate Studies Using Scatterometer Image Data. *EOS Transac AGU* 82: 503.
  28. Frolking S, Milliman T, McDonald K, Kimball J, Zhao M, et al. (2006) Evaluation of the SeaWinds Scatterometer for Regional Monitoring of Vegetation Phenology. *J Geophys Res* 111: doi: 10.29/2005JD006588.
  29. Saatchi SS, Houghton RA, Dos Santos Alvalá RC, Soares JV, Yu Y (2007) Distribution of aboveground live biomass in the Amazon basin. *Global Change Biol* 13: 816–837.
  30. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4: 535–538.
  31. Dieringer D, Schlötter C (2003) Microsatellite analyzer (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol Ecol Notes* 3: 167–169.
  32. Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295.
  33. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
  34. ter Braak CJF, Šmilauer P (2002) CANOCO Reference Manual and CanoDraw for Windows User's Guide: Software for Canonical Community Ordination (version 4.5). Microcomputer Power.
  35. Lepš J, Šmilauer P (2003) *Multivariate Analysis of Ecological Data using CANOCO*. 282 p. Cambridge University Press, Cambridge, UK.
  36. Pease KM, Freedman AH, Pollinger JP, McCormack JE, Buermann W, et al. (2009) Landscape genetics of California mule deer (*Odocoileus hemionus*): the roles of ecological and historical factors in generating differentiation. *Mol Ecol* 18: 1848–1862.
  37. Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* 2: 18–22.
  38. Ripley BD (1996) *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press. 416 p.
  39. Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) *Classification and regression trees*. MontereyCA, USA: Wadsworth and Brooks/Col. 368 p.
  40. Breiman L (2001) *Random forests*. *Machine Learning* 45: 5–32.
  41. Evans JS, Cushman SA (2009) Gradient modeling of conifer species using random forests. *Landscape Ecol* 24: 673–683.
  42. Chen C, Liaw A, Breiman L (2004) Using random forest to learn imbalanced data. <http://http://www.stat.berkeley.edu/tech-reports/666.pdf>
  43. Murphy MA, Evans JS, Sorfer A (2010) Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology* 91: 252–261.
  44. Strobl C, Boulesteix AL, Zeileis A, Hothorn T (2007) Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinf* 8: 25.
  45. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A (2008) Conditional variable importance for random forests. *BMC Bioinf* 9: 307.
  46. Strobl C, Malley J, Tutz G (2009) An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods* 14: 323–348.
  47. Ferrier S, Manion G, Elith J, Richardson K (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity Distrib* 13: 252–264.
  48. Ray N (2005) PATHMATRIX: a geographical information system tool to compute effective distances among samples. *Mol Ecol Notes* 5: 177–180.
  49. McRae BH (2006) Isolation by resistance. *Evolution* 60: 1551–1561.
  50. Homewood K, Lambin EF, Coast E, Kariuki A, Kikula I, et al. (2001) Long-term changes in Serengeti-Mara wildebeest and land cover: pastoralism, population, or policies? *Proc Natl Acad Sci U S A* 98: 12544–12549.
  51. Riggio J, Jacobson A, Dollar L, Bauer H, Becker M, et al. (2012) The size of savannah Africa: a lion's (*Panthera leo*) view. *Biodivers Conserv* 22: 17–35.
  52. Groen TA, van Langevelde F, van de Vijver CADM, de Raad AL, de Leeuw J, et al. (2011) A continental analysis of correlations between tree patterns in African savannas and human and environmental variables. *J Arid Environ* 75: 724–733.
  53. Foster JB, Dagg AI (1972) Notes on the biology of giraffe. *E Afr Wildl J* 10: 1–16.
  54. Pellew RA (1983) The giraffe and its food resources in the Serengeti II. Response of the giraffe population to changes in the food supply. *Afr J Ecol* 21: 269–283.
  55. Cutler DR, Edwards TC Jr., Beard KH, Cutler A, Hess KT, et al. (2007) Random forests for classification in ecology. *Ecology* 88: 2783–2792.
  56. R East (1998) ASG co-chair, African Antelope Database, IUCN/SSC Antelope Specialist Group Report.
  57. Shirk AJ, Wallin DO, Cushman SA, Rice CG, Warheit KI (2010) Inferring landscape effects on gene flow: a new model selection framework. *Mol Ecol* 19: 3603–3619.
  58. Jacobs DK, Sahagian DL (1993) Climate-induced fluctuations in sea-level during non-glacial times. *Nature* 361: 710–712.
  59. Rossignol Strick M, Paterne M (1999) A synthetic pollen record of the eastern Mediterranean sapropels of the last 1 Ma: implications for the time-scale and formation of sapropels. *Marine Geol* 153: 221–237.
  60. Teunter E, Weber SL, Hilgen EJ, Lourens LJ (2003) The response of the African summer monsoon to remote and local forcing due to precession and obliquity. *Global Planet Change* 36: 219–235.
  61. Liu Z, Harrison SP, Kutzbach J, Otto-Bliesner B (2004) Global monsoons in the mid-Holocene and oceanic feedback. *Clim Dynam* 22: 157–182.
  62. Hamilton AC (1982) *Environmental history of East Africa: A study of the Quaternary*. London: Academic Press. 328 p.
  63. Hamilton AC, Taylor D (1991) History of climate and forests in tropical Africa during the last 8 million years. *Clim Change* 19: 65–78.
  64. Flagstad Ø, Syversten PO, Stenseth NC, Jakobsen KS (2001) Environmental change and rates of evolution: the phylogeographic pattern within the hartebeest complex as related to climate variation. *Proc Roy Soc London B* 268: 667–677.
  65. Muwanika VB, Nyakaana S, Siegmund HR, Arctander P (2003) Phylogeography and population structure of the common warthog (*Phacochoerus africanus*) inferred from variation in mitochondrial DNA sequences and microsatellite loci. *Heredity* 91: 361–372.
  66. Alpers DL, Van Vuuren BJ, Arctander P, Robinson TJ (2004) Population genetics of the roan antelope (*Hippotragus equinus*) with suggestions for conservation. *Mol Ecol* 13: 1771–1784.
  67. Barnett R, Yamaguchi N, Barnes I, Cooper A (2006) The origin, current diversity and future conservation of the modern lion (*Panthera leo*). *Proc Roy Soc B* 273: 2119–2125.
  68. Lorenzen ED, Simonsen BT, Kat TW, Arctander P, Siegmund HR (2006) Hybridization between subspecies of waterbuck (*Kobus ellipsiprymnus*) in zones of overlap with limited introgression. *Mol Ecol* 15: 3787–3799.
  69. Antunes A, Troyer JL, Roelke ME, Pecon-Slattery J, Packer C, et al. 2008. The evolutionary dynamics of the lion *Panthera leo* revealed by host and viral population genomics. *PLoS Genet* 11: e10000251.
  70. Lorenzen ED, Arctander P, Siegmund HR (2008) Three reciprocally monophyletic mtDNA lineages elucidate the taxonomic status of Grant's gazelles. *Conserv Genet* 9: 593–601.
  71. Zinner D, Groeneveld LF, Keller C, Roos C (2009) Mitochondrial phylogeography of baboons (*Papio* spp.) – Indication for introgressive hybridization? *BMC Evol Biol* 9: 83.
  72. Sinclair ARE, Mduma SAR, Arcese P (2000) What determines phenology and synchrony of ungulate breeding in Serengeti? *Ecology* 81: 2100–2111.
  73. Stevenson-Hamilton J (1947) *Wild life in South Africa*. London: Cassell and Co.
  74. Brynard AM, Pienaar U de V (1960) Annual report of the biologist, 1958/1959. *Koedoe* 3: 1–205.
  75. Dasmann RF, Mossman AS (1962) Reproduction in some ungulates in Southern Rhodesia. *J Mammal* 43: 533–637.
  76. Pienaar U de V (1963) The large mammals of the Kruger National Park – their distribution and present-day status. *Koedoe* 6: 1–37.
  77. Fairall N (1968) The reproductive seasons of some mammals in the Kruger National Park. *Zoologica Africana* 3: 189–210.
  78. Hall-Martin AJ, Skinner JD, Van Dyk JM (1975) Reproduction in the giraffe in relation to some environmental factors. *E Afr Wildl J* 13: 237–248.
  79. Pellew RA (1984) Food consumption and energy budgets of the giraffe. *Journal of Applied Ecology* 21: 141–159.
  80. Pratt DM, Anderson VF (1979) Giraffe cow-calf relationships and social development of the calf in the Serengeti. *Zeitschrift für Tierpsychologie* 51: 233–251.
  81. Caister LE, Shields WM, Gosser A (2003) Female tannin avoidance: a possible explanation for habitat and dietary segregation of giraffes (*Giraffa camelopardalis peralta*) in Niger. *Afr J Ecol* 41: 210–210.
  82. Dagg AI (1971) *Giraffa camelopardalis*. *Mammal Species* 5: 1–8.
  83. Kingdon J (1997) *The Kingdon field guide to African mammals*. San Diego, CA: Academic Press.
  84. Owen-Smith N (2008) Changing vulnerability to predation related to season and sex in an African ungulate assemblage. *Oikos* 117: 602–610.
  85. Ginnett TF, Denment MW (1999) Sexual segregation by Masai giraffes at two spatial scales. *Afr J Ecol* 37: 93–106.
  86. Pellew RA (1984) The feeding ecology of a selective browser, the giraffe (*Giraffa camelopardalis tippelskirchi*). *J Zool London* 202: 57–81.
  87. Malpoux B, Migaud M, Tricoire H, Chemineau P (2001) Biology of mammalian photoperiodism and the critical role of pineal gland and melatonin. *J Biol Rhythms* 16: 336–347.
  88. Dagg AI, Foster JB (1976) *The giraffe: its biology, behavior, and reproduction*. New York: Van Nostrand.
  89. Reason R (2000) Reproductive parameters in female giraffe (*Giraffa camelopardalis*) at Brookfield Zoo. *Animal Keeper's Forum* 27: 120–123.
  90. Bercovitch FB, Bashaw MJ, Penny CG, Rieches RG (2004) Maternal investment in captive giraffes. *J Mammal* 85: 428–431.
  91. del Castillo SM, Bashaw MJ, Patton ML, Rieches RR, Bercovitch FB (2005) Fecal steroid analysis of female giraffe (*Giraffa camelopardalis*) reproductive condition and the impact of endocrine status on daily time budgets. *Gen Comp Endocr* 141: 271–281.
  92. Field CR, Blankenship LH (1973) Nutrition and reproduction of Grant's and Thomson's gazelles, Coke's hartebeest and giraffe in Kenya. *J Reprod Fertil, Supplement* 19: 287–301.

93. Ogotu JO, Piepho H-P, Dublin HT, Bholá N, Reid RS (2008) Rainfall influences on ungulate population abundance in the Mara-Serengeti ecosystem. *J Anim Ecol* 77: 814–829.
94. Wittemyer G, Rasmussen HB, Douglas-Hamilton I (2007) Breeding phenology in relation to NDVI variability in free-ranging African elephant. *Ecography* 30: 42–50.
95. Ryan SJ, Knechtel CU, Getz WM (2007) Ecological cues, gestation length, and birth timing in African buffalo (*Syncerus caffer*). *Behav Ecol* 18: 635–644.
96. Stamps JA, Krishnan VV, Willits NH (2009) How different types of natal experience affect habitat preference. *Am Nat* 174: 623–630.
97. Doebeli M, Dieckmann U, Metz JA, Tautz D (2005) What we have also learned: adaptive speciation is theoretically plausible. *Evolution* 59: 691–695.
98. Gavrilets S, Vose A (2007) Case studies and mathematical models of ecological speciation. 2. Palms on an oceanic island. *Mol Ecol* 16: 2910–2921.