

RESEARCH ARTICLE

A continent-wide high genetic load in African buffalo revealed by clines in the frequency of deleterious alleles, genetic hitchhiking and linkage disequilibrium

Pim van Hooft¹*, Wayne M. Getz^{2,3}, Barend J. Greyling⁴, Bas Zwaan⁵, Armanda D. S. Bastos⁶

1 Wildlife Ecology and Conservation Group, Wageningen University, Wageningen, The Netherlands, **2** Department of Environmental Science Policy & Management, University of California, Berkeley, California, United States of America, **3** School of Mathematical Sciences, University of KwaZulu-Natal, Durban, South Africa, **4** Agricultural Research Council, Irene, South Africa, **5** Laboratory of Genetics, Wageningen University, Wageningen, The Netherlands, **6** Department of Zoology & Entomology, University of Pretoria, Hatfield, South Africa

* pim.vanhooft@wur.nl



OPEN ACCESS

Citation: van Hooft P, Getz WM, Greyling BJ, Zwaan B, Bastos ADS (2021) A continent-wide high genetic load in African buffalo revealed by clines in the frequency of deleterious alleles, genetic hitchhiking and linkage disequilibrium. PLoS ONE 16(12): e0259685. <https://doi.org/10.1371/journal.pone.0259685>

Editor: Randeep Singh, Amity University, INDIA

Received: May 12, 2021

Accepted: October 24, 2021

Published: December 9, 2021

Copyright: © 2021 van Hooft et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The dataset is entitled: "Microsatellite data from various African buffalo (*Syncerus caffer*) populations throughout Africa", with a unique digital object identifier (DOI): [doi:10.5061/dryad.s7h44j14c](https://doi.org/10.5061/dryad.s7h44j14c).

Funding: The author(s) received no specific funding for this work.

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

Abstract

A high genetic load can negatively affect population viability and increase susceptibility to diseases and other environmental stressors. Prior microsatellite studies of two African buffalo (*Syncerus caffer*) populations in South Africa indicated substantial genome-wide genetic load due to high-frequency occurrence of deleterious alleles. The occurrence of these alleles, which negatively affect male body condition and bovine tuberculosis resistance, throughout most of the buffalo's range were evaluated in this study. Using available microsatellite data (2–17 microsatellite loci) for 1676 animals from 34 localities (from 25°S to 5°N), we uncovered continent-wide frequency clines of microsatellite alleles associated with the aforementioned male traits. Frequencies decreased over a south-to-north latitude range (average per-locus Pearson $r = -0.22$). The frequency clines coincided with a multilocus-heterozygosity cline (adjusted $R^2 = 0.84$), showing up to a 16% decrease in southern Africa compared to East Africa. Furthermore, continent-wide linkage disequilibrium (LD) at five linked locus pairs was detected, characterized by a high fraction of positive interlocus associations (0.66, 95% CI: 0.53, 0.77) between male-deleterious-trait-associated alleles. Our findings suggest continent-wide and genome-wide selection of male-deleterious alleles driven by an earlier observed sex-chromosomal meiotic drive system, resulting in frequency clines, reduced heterozygosity due to hitchhiking effects and extensive LD due to male-deleterious alleles co-occurring in haplotypes. The selection pressures involved must be high to prevent destruction of allele-frequency clines and haplotypes by LD decay. Since most buffalo populations are stable, these results indicate that natural mammal populations, depending on their genetic background, can withstand a high genetic load.

Introduction

The presence of deleterious alleles in wildlife populations can result in genetic load: a decrease of the fitness of the average genotype relative to the fittest genotype present in a population [1]. A high genetic load can strongly increase local extinction risk via inbreeding depression, particularly in small populations [2, 3]. Inbreeding depression, which is characterized by lower fertility, survival and growth rates, is generally caused by increased homozygosity for partially recessive deleterious mutations (dominance hypothesis) [4]. However, a high genetic load does not necessarily result in a reduced population growth rate [3, 5]. It has been argued that in many cases deleterious alleles have a limited impact because of density-dependent processes [6]. According to this argumentation selection is typically soft, where selective death due to deleterious alleles would otherwise be replaced by nonselective death due to environmental and ecological conditions (e.g., droughts and intraspecific resource competition). Consequently, population size and persistence often do not seem to be affected by the presence of deleterious alleles, even when they are clearly expressed [7].

Even in the absence of clear effects on population viability, increased expression of deleterious alleles via inbreeding can reduce immune function and body condition in wildlife, thereby increasing susceptibility to pathogens [8–10]. Examples are gastrointestinal nematodes in Soay sheep (*Ovis aries*), hookworm, helminth and bacterial infections in California sea lions (*Zalophus californianus*), feather lice parasitism in wild lesser kestrel (*Falco naumanni*) and Galapagos hawk (*Buteo galapagoensis*), avian malaria in crows (*Corvus brachyrhynchos*) and blue tits (*Cyanistes caeruleus*), West Nile virus in crows (*Corvus brachyrhynchos*), bovine tuberculosis (BTB) in badgers (*Meles meles*), red deer (*Cervus elaphus*) and lions (*Panthera leo*), and porcine circovirus type 2 in wild boar (*Sus scrofa*) [8, 10–19].

African buffalo (*Syncerus caffer*) is a suitable species for studying wildlife population viability in the presence of high genetic load. Most buffalo populations in southern Africa are seemingly thriving despite the observation of high frequencies of expressed deleterious alleles in Kruger National Park (KNP) and Hluhluwe-iMfolozi Park (HiP), South Africa [20–22]. High frequencies of deleterious alleles in these two populations, however, are a result of a sex-ratio meiotic gene-drive system rather than inbreeding [20, 21]. Meiotic drivers are selfish genetic elements that, by distorting meiosis, favour transmission of the chromosome on which they reside. In the case of sex chromosomes, this results in distorted primary sex ratios [23].

Several microsatellite studies on African buffalo from KNP and HiP presented multiple independent lines of evidence, summarized in S1 Table, for the occurrence of deleterious alleles with a negative effect in males on body condition and BTB resistance [20–22, 24, 25]. Two types of microsatellites were observed: one type containing alleles associated with negative phenotypic effects in both sexes, and another containing alleles associated with negative phenotypic effects in males but positive phenotypic effects in females. As these microsatellite alleles are likely linked to male-deleterious alleles at protein-coding genes we will henceforth refer to these as male-deleterious-trait-associated (MDTA) microsatellite alleles. Considering the notable frequency differences among year-cohorts and between unhealthy (BTB-positive and low body condition) and healthy (BTB-negative and high body condition) males, these MDTA alleles likely occur genome-wide and at relatively high frequencies in both the KNP and HiP population.

Male-deleterious alleles appear to have high frequencies and be under positive selection in KNP despite the negative phenotypic effects [22]. Positive selection has been attributed to a sex-ratio meiotic gene-drive system that appears also to be health-associated (S1 Table) [20–22, 25]. It has been hypothesized that poor health in males (due to low body condition and BTB infection) suppresses sex-ratio distortion genes in this gene-drive system that, when

active, reduces fertility. As a consequence, any allele that has a negative effect on male health may have a positive effect on the relative fertility of males, thereby being under positive selection if the net fitness effect of health and fertility across both sexes is positive. In contrast to positive selection of male-deleterious alleles in KNP, however, selection of male-deleterious alleles appears to be negative in HiP (situated just 280 km further south), resulting in relatively low male-deleterious allele frequencies compared to KNP [20]. As discussed elsewhere, this negative selection has been attributed to incompleteness of the gene-drive system [20, 21]. The hypothesized sex-ratio meiotic gene-drive system and its relationship with male-deleterious alleles have been clarified in Van Hooft et al. 2019 (Table 1 and Fig 1 and S1 Text herein) [20].

It is unlikely that the male-deleterious alleles are restricted to just KNP and HiP. Positive selection and movement of individuals (both diffusive and migratory) may have spread these alleles, together with the linked (hitchhiking) MDTA alleles, across a large geographic range. The interplay between gene flow and geographic variation in selection pressure may have resulted in allele-frequency clines [26–28]. Such clines have previously been observed in KNP, not only for the MDTA alleles but also for a Y-chromosomal haplotype hypothesized to be linked to a suppressor gene from the gene-drive system [22].

Multilocus selection may have had a negative effect on population-level genetic diversity by hitchhiking of loci linked to the male-deleterious alleles in a process known as selective sweep [29, 30]. Multilocus selection may also have resulted in extended linkage disequilibrium (LD) due to increased frequencies of haplotypes containing multiple male-deleterious alleles relative to frequencies of haplotypes expected under linkage equilibrium [31]. This may be particularly so if sex-specific selection, a potential consequence of the aforementioned sex-ratio meiotic gene-drive system, resulted in admixture LD due to different allele frequencies in male and female gametes [32]. The latter is supported by the observation of significant genetic differentiation between male and female calves in KNP [21].

Confirmation of an expansive distribution of expressed male-deleterious alleles is of concern as it would suggest that many African buffalo populations experience a high genetic load, making them sensitive to infectious diseases and other environmental stressors. In this study, we address this concern through analysis of previously published microsatellite data from 1676 buffalo from 34 localities across the African continent (Fig 1). Implicit in our analyses is the assumption that microsatellite alleles occurring at a higher frequency among BTB-positive males of low body condition than among BTB-negative males of high body condition tend to be associated with a male-deleterious allele (i.e., no significance to be expected if this assumption is wrong). With this assumption, we addressed the following three questions:

1. Did selection result in male-deleterious allele-frequency clines across the geographic area under study?
2. Did selection, via genetic hitchhiking, result in decreased population-level genetic diversity?
3. Did selection result in increased LD between male-deleterious alleles?

Materials and methods

Localities and microsatellites

We analysed 1676 animals from 34 localities throughout the buffalo's range (Fig 1 and Table 1). These localities span approximately 37 degrees of latitude (\pm 4500 km from Bénoué NP in Cameroon to HiP in South Africa) and 27 degrees of longitude (\pm 3000 km from Lopé NP in Gabon to Tsavo NP in Kenya).

Table 1. Overview of African buffalo populations included in this study.

Population no.	Population	Lat.	Lon.	usat set	Sample size
1	Bénoué NP	8.33	13.83	B	3
2	Lopé NP	-0.50	11.50	B	2
3	Kidepo Valley NP	3.85	33.75	C	19
4	Murchison Falls NP	2.15	31.81	C	15
5	Laikipia NP	0.40	37.16	C	17
6	Queen Elizabeth NP Mw ^a	-0.06	30.00	C	37
7	Queen Elizabeth NP Is ^a	-0.36	30.00	C	17
8	Lake Nakuru NP	-0.40	36.10	B, C	35, 18
9	Lake Mburo NP	-0.58	30.99	C	15
10	Nairobi NP	-1.36	36.85	C	19
11	Masai Mara GR	-1.47	35.10	B, C	10, 33
12	Serengeti NP ^b	-2.20	34.90	B, E	35, 49
13	Amboseli NP	-2.60	37.20	B, C	20, 19
14	Tsavo NP	-2.70	38.40	B	9
15	Niassa Reserve	-12.25	37.50	D	20
16	Mana Pools NP	-15.90	29.40	D	10
17	Nyakasanga	-16.30	29.10	D	2
18	Victoria Falls NP	-17.97	25.85	D	15
19	Caprivi Strip	-18.10	23.17	F	134
20	Chobe NP	-18.45	24.50	D	22
21	Gorongosa NP	-18.80	34.50	D	7
22	Marromeu GR	-18.85	36.00	D	21
23	Hwange NP	-19.20	27.10	D	6
24	Okavango Delta	-19.30	23.05	D	20
25	Save Valley Conservancy ^c	-19.43	27.54	B	10
26	Malilangwe Wildlife Res.	-21.00	31.90	D	20
27	Gonarezhou NP	-21.65	31.70	D	42
28	Sengwe Safari Area	-22.30	31.40	D	8
29	Crooks corner	-22.42	31.31	D	13
30	Limpopo NP	-23.25	31.90	D	6
31	Manguana	-24.90	32.80	D	4
32	Northern KNP ^d	-23.15	31.30	A, B, D	138, 22, 26
33	Southern KNP ^d	-24.75	31.70	A, B	321, 16
34	HiP ^e	-28.22	31.95	A, D	401, 20

a: Mweya sector and Ishasha sector,

b: Including Maswa GR,

c: Intermediate coordinates, which refer to the origin of the buffalos, weighted by stock size. Buffalo were restocked in 1993 from Gonarezhou NP (38 animals) and Hwange NP (360 animals) [33],

d: Kruger National Park, north and south of Olifants River,

e: Hluhluwe-iMfolozi Park

<https://doi.org/10.1371/journal.pone.0259685.t001>

We used previously published genotype data of the following seventeen microsatellites, originally developed for cattle (*Bos taurus*), which were randomly chosen with respect to genomic location: *BM0719*, *BM1824*, *BM3517*, *BM3205*, *BM4028*, *CSSM019*, *DIK020*, *ETH010*, *ETH225*, *ILSTS026*, *INRA006*, *INRA128*, *SPS115*, *TGLA057*, *TGLA0159*, *TGLA227* and *TGLA263* [20–22, 33–38]. Data for all seventeen microsatellites were only available for KNP

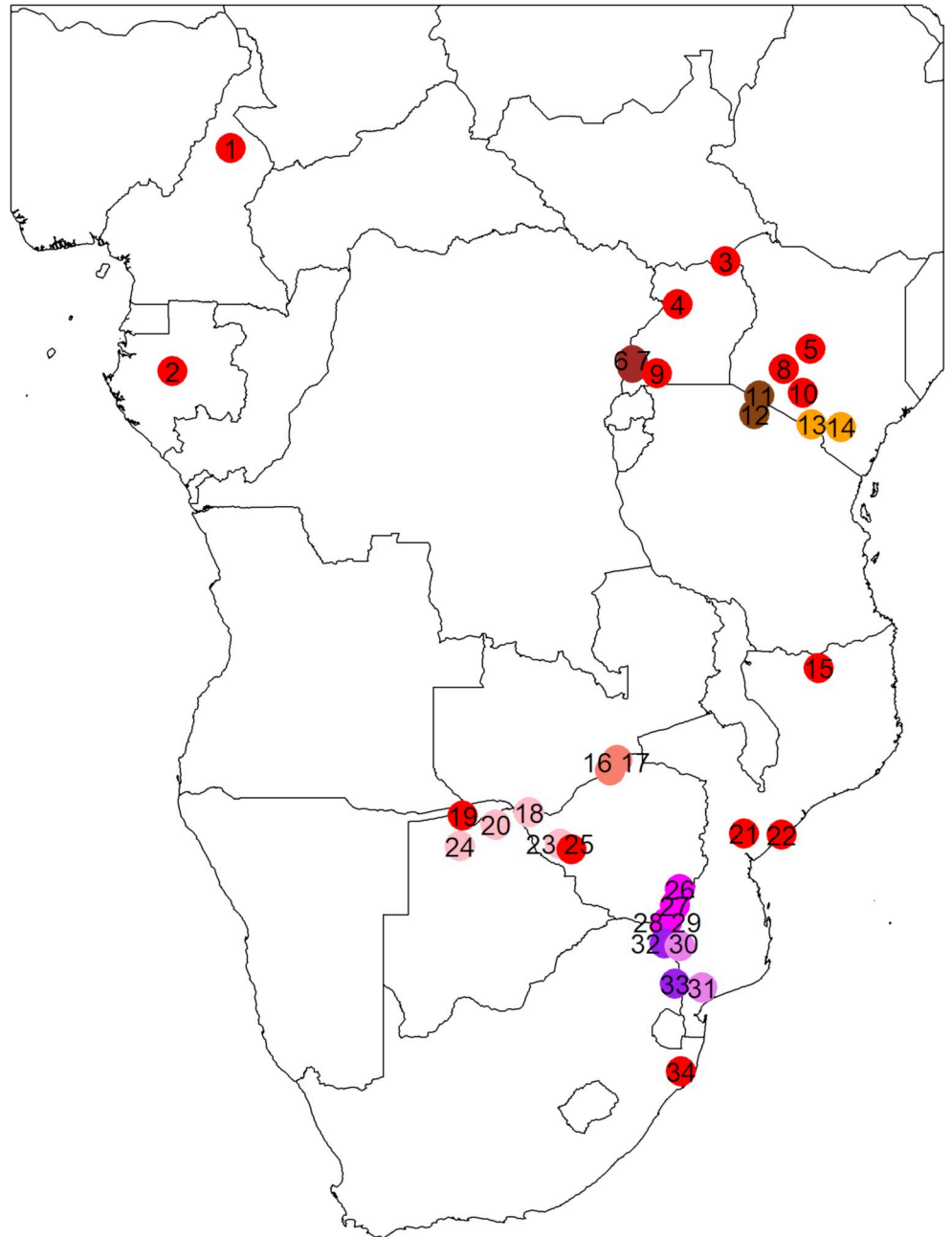


Fig 1. Map of Africa representing the 34 sampling localities.

<https://doi.org/10.1371/journal.pone.0259685.g001>

and HiP (microsatellite set A in Table 2). Five smaller subsets, consisting of 2–11 microsatellites, were available for the other populations (microsatellite sets B–F in Table 2). Raw microsatellite data from microsatellite sets D and F are provided in earlier publications [34, 37], while those from the remaining sets are available from the Dryad Digital Repository [39].

Allele size standardization

Genotype data from different studies obtained with the aforementioned six microsatellite sets were combined after standardizing allele sizes using the Genalex add-in for Excel (version

Table 2. Overview of microsatellite sets used.

Set	μsats	Reference
A	1–17	[20, 22, 38]
B	1, 2, 5, 6, 8–12, 14–16	[33]
C	1, 2, 6, 9–12, 14–16	[35]
D	2–10, 13, 15, 17	[34]
E	3, 13, 17	[36]
F	1, 11	[37]

1: *BM3517*, 2: *BM4028*, 3: *ETH010*, 4: *ETH225*, 5: *INRA006*, 6: *INRA128*, 7: *TGLA227*, 8: *TGLA263*, 9: *CSSM019*, 10: *DIK020*, 11: *TGLA057*, 12: *BM0719*, 13: *BM1824*, 14: *BM3205*, 15: *ILSTS026*, 16: *TGLA159*, 17: *SPS115*.

<https://doi.org/10.1371/journal.pone.0259685.t002>

6.503) [40, 41]. Standardization was necessary, because allele size estimates have been observed to differ by up to 17 bp among platforms [42, 43]. In our case they differed by up to 11 bp (microsatellite set B relative to set A: *BM3517*: 8 bp, *INRA006*: 11 bp, in both cases unambiguous standardization despite large size difference; ≤ 6 bp in the remaining cases). For various populations, two microsatellite sets were available (sets A–D in Tables 2 and 3; sample size per set per population: ≥ 38 individuals), which permitted allele alignment by matching (allele size shift) each microsatellite's allele frequencies while preserving size order. Aligning microsatellite alleles based on frequency information is a valid standardization method when data sets are reasonably large and coming from the same population [44–48]. With two microsatellites sets (sets E and F in Tables 2 and 3; sample size per set per population: ≥ 38 individuals) standardization was only possible in comparison with another microsatellite set applied to another population. Because of the higher risk of alignment errors in the latter case, we also conducted the tests for allele-frequency clines and for correlation between heterozygosity clines and allele-frequency clines without these two microsatellite sets (see Results section).

The allele size shift we used was the one that resulted in the highest Pearson correlation in allele frequencies and lowest F_{ST} between two sets. Given large enough sample sizes, their values are expected to be close to 1 and 0, respectively. Generally, the obtained Pearson correlations were > 0.8 and the F_{ST} values < 0.01 . A detailed description of the allele size standardization procedure is given in S1 Text and S2 Table.

Unaccounted for alignment errors probably increased type II errors and were unlikely to have resulted in a systematic bias of allele frequency estimates. Accuracy of the allele size standardization was strongly supported by the observation that frequency clines of p-MDTA alleles correlated with heterozygosity clines, of which the latter were not influenced by possible

Table 3. Linkage disequilibrium in Kruger National Park among locus pairs showing significant LD in earlier studies.

Locus pair	multi-allelic $ r_{LD} ^a$	p-MDTA r_{LD}	p-random r_{LD}^b	Exact probability ^c
<i>BM1824-BM3205</i>	0.126 (0.358)	0.176	0.031	0.28
<i>BM1824-CSMM019</i>	0.157 (0.189)	0.337	0.076	0.17
<i>CSMM019-BM3205</i>	0.158 (0.154)	0.339	0.031	0.033
Average first three pairs	0.147 (0.234)	0.284	0.046	0.010
<i>ILSTS026-INRA006</i>	0.128 (0.141)	-0.041	0.002	0.76
Average all pairs	0.142 (0.211)	0.203	0.035	0.019

a: Between brackets: multi-allelic $|r_{LD}|$ in Hluhluwe-iMfolozi Park,

b: average r_{LD} when alleles were randomly assigned to the focal allele class,

c: exact 2 times 1-sided probability $p\text{-random} \geq p\text{-MDTA}$ (*ILSTS026-INRA006*: $p\text{-random} \leq p\text{-MDTA}$)

<https://doi.org/10.1371/journal.pone.0259685.t003>

alignment errors (see [Results](#) section). Furthermore, it is unlikely that alignment errors generally resulted in positive LD among p-MDTA alleles, as was also observed in KNP with microsatellite set A only (see [Results](#) section). Additional support for the accuracy of the allele size standardization was provided by the high amount of explained variation among populations in East and southern Africa, with R^2 being 0.78, when regressing pairwise multilocus F_{ST} with geographic distance (predicted F_{ST} increasing from 0.014 at 0 km to 0.11 at 3500 km, [S1 Fig](#)).

Male-deleterious-trait associations

Alleles at the seventeen microsatellite loci (dataset A) have previously been shown to be significantly associated with two deleterious traits in male buffalo from southern KNP (i.e., south of the Olifants River): 1) low body condition and 2) BTB infection (see also [S1 Table](#)) [[21](#), [22](#)]. BTB was present only in southern KNP at the time of sampling, except for three BTB-positive individuals in the north. Here we considered all microsatellite alleles from southern KNP, after excluding rare alleles, as being associated with male-deleterious traits when they had a higher frequency among low-body-condition, BTB-positive males than among high-body-condition, BTB-negative males ([S3 Table](#)). The sample sizes of these two male groups were similar, $N = 35$ and $N = 39$ respectively, with alleles considered rare when observed ≤ 8 times (frequency ≤ 0.054) in total in these two groups. This was the highest possible cut-off point where both male-deleterious-trait-associated (MDTA) alleles and wild-type associated alleles (relatively low frequency among low-body-condition, BTB-positive males) could be identified at each locus (32 MDTA and 36 wild-type associated alleles in total, [S3 Table](#)). As argued in an earlier study, comparison of allele frequencies between the two male groups probably constitutes a reliable indicator of association strength with male-deleterious traits [[20](#)]. In all tests, we pooled the frequencies of MDTA alleles at each microsatellite locus ([S4 Table](#)). Thus, as discussed elsewhere [[20](#)], we assumed two alleles per coding gene: a male-deleterious allele associated with the pooled MDTA (p-MDTA) alleles and a wild-type allele associated with the pooled remaining (p-wildtype) microsatellite alleles.

Allele-frequency clines

We tested whether allele-frequency clines could be attributed to selection by comparing frequency clines of p-MDTA alleles with frequency clines of randomly pooled (p-random) microsatellite alleles. Microsatellite alleles were randomly pooled at each locus with the same number of constituent alleles per locus as the MDTA alleles (without changing the original genotypes). The frequency clines were based on the frequency differences per locus between p-MDTA and p-wildtype alleles (p-MDTA minus p-wildtype) for each of 31 localities. We excluded HiP (location 34 in [Fig 1](#) and [Table 1](#)), because here the male-deleterious alleles had earlier been shown to be under strong negative selection [[20](#)]. Samples from Bénoué NP and Lopé NP (separated by ± 1000 km; locations 1–2) in central Africa and from Mana Pools NP and Nyakasanga (separated by ± 55 km; locations 16–17) in southern Africa were pooled because of small sample size. Intermediate coordinates, weighted by sample size, were used for these pooled samples. We estimated significance as the probability that random pooling of alleles resulted in an average Pearson correlation (r) between frequency difference and latitude at each locus at least strong as with the p-MDTA alleles. To minimize bias in average r we first transformed each individual r to a Fisher's z , after which we back-transformed the average z [[49](#)]. This transformation did not affect probability estimates. Further, to reduce heteroskedasticity, the r values for each locus were weighted by the square-root of the number of individuals genotyped at the locus at each location. We weighted by the square-root because without this transformation the r values would have been strongly biased by the relatively large sample sizes of the two KNP subpopulations.

We also determined whether per-locus allele-frequency differences (p-MDTA minus p-wildtype) were significantly positive in both central-East and southern Africa. If so, this positivity would be indicative of selection across the entire latitudinal range. Again, allele-frequency differences were weighted by the square-root of the number of genotypes per location. We estimated significance as the probability that random pooling of alleles resulted in an average allele-frequency difference per locus at least as high as with the p-MDTA alleles.

In both tests, we applied 100,000 randomizations implemented using Excel 2016. The *P*-values are reported as 2 times 1-sided values. We plotted average per-locus frequency difference against latitude, with the linear regression line obtained by weighing the population at each location by the square-root of the number of genotyped individuals multiplied by the average number of genotyped loci per individual.

Heterozygosity clines

We conducted two statistical tests for an effect of selection on expected heterozygosity (H_e), with H_e estimates at each locus based on the original non-pooled microsatellite alleles. Unbiased H_e estimates [50] were obtained with the Excel add-in Microsatellite Toolkit [51]. To make H_e estimates independent from assumptions about allele size standardization we only used microsatellite set A for KNP and used the weighted average H_e of microsatellite sets B and C (i.e., H_e was separately estimated for each set) for Lake Nakuru NP, Masai Mara GR and Amboseli NP. Only one microsatellite set was available for each of the remaining localities. We would like to stress here that the H_e estimates were unaffected by any assumptions about male-deleterious alleles.

In the first heterozygosity test, we determined whether selection had an influence on population-level H_e by correlating heterozygosity clines with the aforementioned allele-frequency clines. A negative correlation would indicate a negative influence. As before, we excluded HiP (location 34 in Fig 1 and Table 1) because of earlier observed negative selection [20]. We also excluded Bénoué NP (location 1), Lopé NP (location 2) and Nyakasanga (location 17) because of small sample size and the admixed population from Save Valley Conservancy (location 25; buffalo were restocked in 1993 with 38 animals from Gonarezhou NP and 360 animals from Hwange NP [33]) because admixing may positively bias H_e estimates. We estimated significance as the probability that random pooling of alleles resulted in a Pearson correlation between the H_e cline (correlation between H_e and latitude) and the allele-frequency cline (correlation between allele frequency and latitude) at each locus at least as strong as the correlation between the H_e cline and the observed p-MDTA allele-frequency cline. We applied 100,000 randomizations implemented using Excel 2016. The *P*-values are reported as 2 times 1-sided values.

We plotted multilocus H_e against latitude, including only localities for which the same microsatellites were analysed and with the linear regression line obtained by weighing the population at each location by the square-root of the total number of genotypes. One group of seventeen localities between 24.9°S and 2.2°S and a second group of twelve localities between 2.7°S and 3.9°N were selected for analysis using ten microsatellites for each group, with five microsatellites shared between the two groups. Multilocus H_e estimates in the southern group were based on microsatellites *BM1824*, *BM4028*, *ETH010*, *INRA006*, *INRA128*, *CSSM019*, *DIK020*, *ILSTS026*, *SPS115* and *TGLA263*; and those in the eastern group on microsatellites *BM3517*, *BM4028*, *INRA128*, *BM0719*, *BM3205*, *CSSM019*, *DIK020*, *ILSTS026*, *TGLA057* and *TGLA159*. Multilocus H_e of the northern localities was multiplied with 0.966, which was the ratio between the two microsatellite sets in Serengeti NP; the only locality in East Africa analysed with both sets.

In the second heterozygosity test we determined whether recent (since < 40 years before sampling) and historical (since > 100 years before sampling) selection pressures were correlated. In particular, we conducted a Spearman correlation between the per-locus southern/northern KNP H_e ratio (H_e southern KNP/ H_e northern KNP), indicative of a short-term hitchhiking effect caused by the introduction of BTB around 1960 [21, 22], and the weighted Pearson correlation (by square-root of number of genotypes) between H_e and latitude across Africa (24.9°S—3.9°N, excluding southern KNP for independence with aforementioned H_e ratio), indicative of a long-term hitchhiking effect. A significant relationship cannot be attributed to random genetic drift considering that in such case the fractional decrease in per-locus H_e would be determined by effective population size and gene flow, which are not locus specific. Despite a small genetic distance between northern and southern KNP ($F_{ST} = 0.0034$, 95% CI = 0.0023, 0.0045; estimated with FSTAT 2.9.3 [52, 53]), the latter had a significantly lower multilocus H_e (multilocus H_e northern KNP = 0.66, multilocus H_e southern KNP = 0.64; Student's paired t-test: $P = 0.014$, $N = 17$).

Linkage disequilibrium

Four microsatellite pairs showed highly significant LD in both KNP and HiP [20, 22]. Three of these occur close on chromosome 1 in cattle and its homologous chromosome in African buffalo (whole genome sequence data of cattle and contig sequence data of buffalo at NCBI website): *BM1824-CSMM019* (0.5 Mb in buffalo; microsatellite sets A and D in Table 2), *CSSM019-BM3205* (5.4 Mb; sets A-C), *BM1824-BM3205* (5.9 Mb; set A) [54]. The fourth pair consists of microsatellites *ILSTS026* and *INRA006* (sets A, B and D), which are probably ~19 Mb apart on the same chromosome in buffalo. This chromosome corresponds to a fusing of chromosomes 2 and 3 in cattle, with the former harbouring *ILSTS026* and the latter *INRA006*, in both cases ~9.5 Mb from the end of the left-side of their p arms [54–56]. Although 19 Mb is still a relatively large distance, recombination rate may well be reduced in the fused region [57]. Two additional microsatellite pairs are at close chromosomal distance, although they did not show significant LD in KNP or HiP: *INRA006-TGLA263* (27.5 Mb in buffalo, chromosome 3 in cattle; microsatellite sets A, B and D in Table 2) and *TGLA057-INRA128* (28.3 Mb, chromosome 1 in cattle; sets A-C) [54]. The chromosomal distances at the aforementioned locus pairs (range: 0.5–28.3 Mb) correspond to recombination rates of 0.6–31%/generation, based on a recombination rate of 1.1 cM/Mb at chromosomes 1 to 3 in cattle.

We conducted three statistical tests for LD, with LD estimated per population per locus pair, assuming two alleles per locus: the p-MDTA and the p-wildtype allele. Here, the latter also included the rare original microsatellite alleles (observed ≤ 8 times among low-body-condition, BTB-positive males and high-body-condition, BTB-negative males from southern KNP) because in LD analyses frequencies must add up to 1 (in contrast to the frequency-cline tests where rare alleles were excluded).

We measured LD as the Pearson correlation coefficient, r_{LD} , for a 2x2 table representing biallelic locus pairs. When the linkage phase of all genotypes is known, r_{LD} can be derived from the observed and expected haplotype frequencies (χ^2 value, four possible haplotypes) and the total number of gametes or chromosomes analysed (k), using the following equation [28]:

$$r_{LD} = \sqrt{\frac{\chi^2}{k}} \quad (1)$$

A negative sign is added to r_{LD} when pairs of focal alleles occur less often than expected (i.e. when they tend to be in repulsion). However, when the linkage phase of double heterozygotes is unknown, such as in our case, r_{LD} can be derived from maximum-likelihood estimates of

the haplotype frequencies (ML r_{LD}) [58, 59]. We estimated ML r_{LD} using Genalex, following Weir's method and assuming Hardy-Weinberg equilibrium [59]. In our case, positive ML r_{LD} values indicate that pairs of male-deleterious alleles tend to be in coupling.

Except for KNP, unbiased estimates of r_{LD} were not possible, because for this a minimum sample size of 50 non-double heterozygotes is required, particularly with interlocus distances > 10 Mb [60, 61]. Therefore, we determined only the sign of r_{LD} in the first two LD tests and estimated significance by randomizing genotypes and alleles. We excluded populations with less than five non-double heterozygotes (i.e., < 10 known haplotypes). Any bias or inaccuracy due to low sample size equally applied to the observed and the randomized data and therefore was unlikely to affect probability estimates. We pooled the samples from the neighbouring populations in Mana Pools NP-Nyakasanga (~55 km distance; locations 16–17 in Fig 1 and Table 1) and in Limpopo NP-Manguana (~200 km distance; locations 30–31) to increase sample size. Intermediate coordinates, weighted by sample size, were used for these pooled samples. Further, we excluded HiP (location 34) because of earlier observed negative selection [20] and the admixed population from Save Valley Conservancy (location 25) because admixing may bias LD estimates.

The sign of the ML r_{LD} was always identical to that of r_{LD} for the non-double heterozygotes alone, except when the latter was close to zero and one haplotype from each of the two possible double-heterozygous haplotype pairs (one in coupling and one in repulsion) was rare. Such cases, as well as cases where r_{LD} of the non-double heterozygotes was exactly zero, were excluded from our analyses (different signs: *CSSM019-BM3205* in Kidepo Valley NP and *TGLA057-INRA128* in Masai Mara GR; $r_{LD} = 0$: *CSSM019-BM3205* in Murchison Falls NP and *ILSTS026-INRA006* in Limpopo NP). Further, Weir's method can only estimate the sign of ML r_{LD} when all four possible haplotypes are observed among the non-double heterozygotes. With only three observed haplotypes, therefore, sign of r_{LD} was only included if linkage phase of the double heterozygotes did not affect it (i.e., all double heterozygotes in repulsion resulted in the same sign as all double heterozygotes in coupling). The occurrence of only two haplotypes was always due to monomorphism at a locus, in which case LD cannot be estimated in principle.

LD estimates of the three closest linked locus pairs, *BM1824-BM3205*, *BM1824-CSMM019* and *CSMM019-BM3205*, for populations that are part of a larger metapopulation may not be independent of each other (indeed, sign was always identical when part of the same metapopulation; sign test: $P = 0.0078$, $N = 8$). We excluded *BM1824-BM3205* because this locus pair was only analysed in KNP. Further, to err on the conservative side, we used only one estimate per metapopulation for the other two locus pairs: the fraction of populations with positive r_{LD} . With respect to *BM1824-CSMM019* and *CSMM019-BM3205*, the following six population location lists were considered to each constitute a separate metapopulation because of their proximity and landscape-corridor connectedness (Fig 1). Metapopulation 1: northern KNP, southern KNP (locations 32–33 in Fig 1 and Table 1); Metapopulation 2: Malilangwe Wildlife Res., Gonarezhou NP, Sengwe Safari Area, Crooks Corner (locations 26–29); Metapopulation 3: Victoria Falls NP, Chobe NP, Hwange NP, Okavango Delta (locations 18, 20, 23–24); Metapopulation 4: Amboseli NP, Tsavo NP (locations 13–14); Metapopulation 5: Masai Mara GR, Serengeti NP (locations 11–12), Metapopulation 6: Queen Elizabeth NP Mw, Queen Elizabeth NP Is (locations 6–7). We assumed independence of r_{LD} estimates for the other three physically linked locus pairs, even when part of the same metapopulation, because of their relatively large interlocus distances. For these locus pairs, the expected recombination rate (~21–31% per generation) likely exceeds the migration rate between nearby populations. LD estimates of the neighbouring locus pairs *BM1824-CSMM019* and *CSMM019-BM3205* in KNP, the only population in our analyses that included both these pairs, were considered independent from

each other in the absence of selection, which was supported by the results of the third LD test (no meaningful correlation in r_{LD} between the two locus pairs when using random allele assignments).

In the first two LD tests, we hypothesized that, in comparison with random allele associations between locus pairs, pairs of p-MDTA alleles were coupled more often than expected and that these coupled pairs occurred on average more southerly than uncoupled pairs. The latter observation suggests an LD cline, although the statistical power to detect this was quite low. Even when the fraction of coupled p-MDTA allele pairs at linked microsatellites in the buffalo genome linearly decreases from a value as high as 0.8 at 24.9°S to 0.5 (expected fraction in the absence of selection) at 4.8°N, the probability of observing a significant ($\alpha = 0.05$) or near-significant ($\alpha = 0.1$) difference in average latitude with a Student's t -test between coupled and repulsed allele pairs was only 34% and 46%, respectively (60 r_{LD} estimates from 24 locations using 50 randomizations).

In both LD tests significance was estimated using two randomization schemes. In the first scheme, the genotypes at one locus per locus pair were randomized for each of the 27 locations (non-randomized data: 60 r_{LD} estimates in total from five locus pairs and 24 locations). In the second scheme, the non-rare microsatellite alleles (observed ≥ 9 times among low-body-condition, BTB-positive males, as well as high-body-condition, BTB-negative males from southern KNP) were randomly assigned to the focal-allele class (i.e., alleles coded as 1 vs. non-assigned alleles coded as 0; without changing the original genotypes), using the same number of constituent alleles per focal allele as the MDTA alleles. While the first randomization scheme merely tested whether LD was significant or occurred significantly more often at southern localities, the second scheme tested whether this could be attributed to selection. We estimated significance as the probability that randomized data resulted in: for LD test 1, a total fraction of locus pairs (across five locus pairs and 27 locations) with positive r_{LD} at least as large as the observed data with the p-MDTA alleles as focal alleles; or, LD test 2, an average latitude of locus pairs with positive r_{LD} at least as low as the observed data with the p-MDTA alleles as focal alleles. All populations were included in the randomizations, including those whose r_{LD} sign could not be determined, as a result of which the subset of datapoints differed between randomizations (i.e., < 27 locations in most randomizations). We applied 100,000 randomizations implemented using Excel 2016. P -values are reported as 2 times the 1-sided value.

As a control, we estimated the total fraction of locus pairs with positive r_{LD} across all populations for: 1) the five closely linked locus pairs (excluding *BM1824-BM320*, which was only analysed in KNP) with the genotypes at one locus randomized per location, and 2) locus pairs with interlocus distance > 45 Mb (seven locus pairs) or that occur on different chromosomes (123 locus pairs). Both fractions are expected to be close to 0.5 in the absence of selection (with free recombination) [62]. In the first control we conducted 100,000 randomizations of the total dataset, resulting in 2,422,396 randomized r_{LD} values.

In the third LD test, we determined whether earlier observed LD in KNP (treated as a single population) at locus pairs *BM1824-BM3205*, *BM1824-CSMM019*, *CSMM019-BM3205* and *ILSTS026-INRA006* based on the original non-pooled microsatellite alleles [20, 22], could be attributed not only to physical proximity but also to selection. Similar to the first two LD tests, we randomly assigned microsatellite alleles to the focal-allele class according to the second randomization scheme, and we estimated significance as twice the exact probability that random assignments resulted in a r_{LD} value at least as high as with the p-MDTA alleles as focal alleles. We would like to stress here that this LD test was not meaningfully influenced by possible alignment errors (similar results when only microsatellite set A was included, which was used for 88% of the KNP samples). Additionally, we tested whether r_{LD} values based on random allele assignments were correlated between nearby locus pairs *BM1824-BM3205*, *BM1824-*

CSMM019 and *CSMM019-BM3205*. Weak correlations would indicate that in the absence of selection r_{LD} values at these locus pairs can be treated as (nearly) independent observations.

For comparison, we estimated multi-allelic $|r_{LD}|$ (sign not relevant with > 2 alleles) in KNP and HiP based on the non-pooled microsatellite alleles at the four locus pairs with earlier observed LD, according to Equation 3 in Zhao et al. 2005 and excluding alleles with frequency < 0.075 [63]. For each allele pair we estimated ML $|r_{LD}|$ with Genalex, following Weir's method and assuming Hardy-Weinberg equilibrium [59]. Comparison of r_{LD} values based on p-MDTA alleles between KNP and HiP was not possible because of low p-MDTA frequencies in HiP, resulting in inaccurate r_{LD} estimates (p-MDTA frequencies: *BM1824*: 0.0059, *BM3205*: 0.065, *INRA006*: 0.028).

Results

Allele-frequency clines

We observed allele-frequency clines between 24.9°S (treated as negative latitude values—see Fig 2) and 4.8°N with average per-locus Pearson $r = -0.22$ (frequency difference between p-MDTA and p-wildtype alleles; p-MDTA alleles: $r = -0.53$, p-wildtype alleles: $r = -0.27$), which could be attributed to selection ($P = 0.025$, 12 out of 17 loci showed a negative correlation; when excluding microsatellite sets E and F: $r = -0.18$, $P = 0.042$; S5 Table). Plotting average frequency increase of p-MDTA alleles across loci relative to p-wildtype alleles explained 74% of the variation (Fig 2). Predicted average frequency increase varied between 0.30 (95% CI: 0.27, 0.33) at 24.9°S to 0.03 (95% CI: -0.01, 0.07) at 4.8°N (a decrease of 0.0090 per degree latitude) and could be attributed to selection across the whole latitudinal range (average frequency increase; southern Africa: 0.24, $P = 0.00095$; central-East Africa: 0.08, $P = 0.083$; when excluding microsatellite sets E and F: southern Africa: 0.24, $P = 0.00066$; central-East Africa: 0.07, $P = 0.14$). The HiP population deviated from the overall trend, showing a clear frequency decrease of p-MDTA alleles relative to p-wildtype alleles (0.10 decrease) in sharp contrast to the other southern African populations (range: 0.13–0.37 increase). This frequency decrease has earlier been attributed to negative selection [20].

Heterozygosity clines

There was a significant negative correlation between H_e clines and allele-frequency clines, which could be attributed to selection (Pearson $r = -0.62$, $P = 0.0079$, 17 loci; attribution to selection: $P = 0.029$; when excluding microsatellite sets E and F: $r = -0.55$, attribution to selection: $P = 0.070$; S5 Table). Regressing multilocus H_e against latitude explained 84% of the variation (Fig 3 and S6 Table; 85% when also including *ABS010* and *AGLA293* which were not part of microsatellite set A, but which were included in the original publications of sets B, C and D, S2 Fig; 74% when including only the seven microsatellites analysed in both East and southern Africa with sets A, B and D, plus *ABS010* and *AGLA293*, S3 Fig). Predicted multilocus H_e decreased by 16% (95% CI: 13%, 19%) at 24.9°S compared to 3.9°N. Again, the HiP population deviated from the overall trend, showing by far the lowest multilocus H_e of all populations (0.53 vs. range: 0.61–0.76). Low genetic diversity, which was also observed with mtDNA and at the Y-chromosome, has earlier been attributed to a historical bottleneck (≤ 75 individuals between 1895 and 1930) [20].

There was a significant Spearman correlation between the per-locus southern/northern KNP H_e ratio on the one hand and the continent-wide Pearson correlation between H_e and latitude on the other (Spearman $\rho = -0.55$, $P = 0.021$, N loci = 17, Fig 4; Spearman $\rho = -0.51$, $P = 0.025$, N loci = 19, when also including *ABS010* and *AGLA293*, S4 Fig). This indicates that selective sweeps in southern KNP caused by the introduction of BTB around 1960 [21, 22] were correlated with historical selective sweeps (since > 100 years before sampling) across Africa.

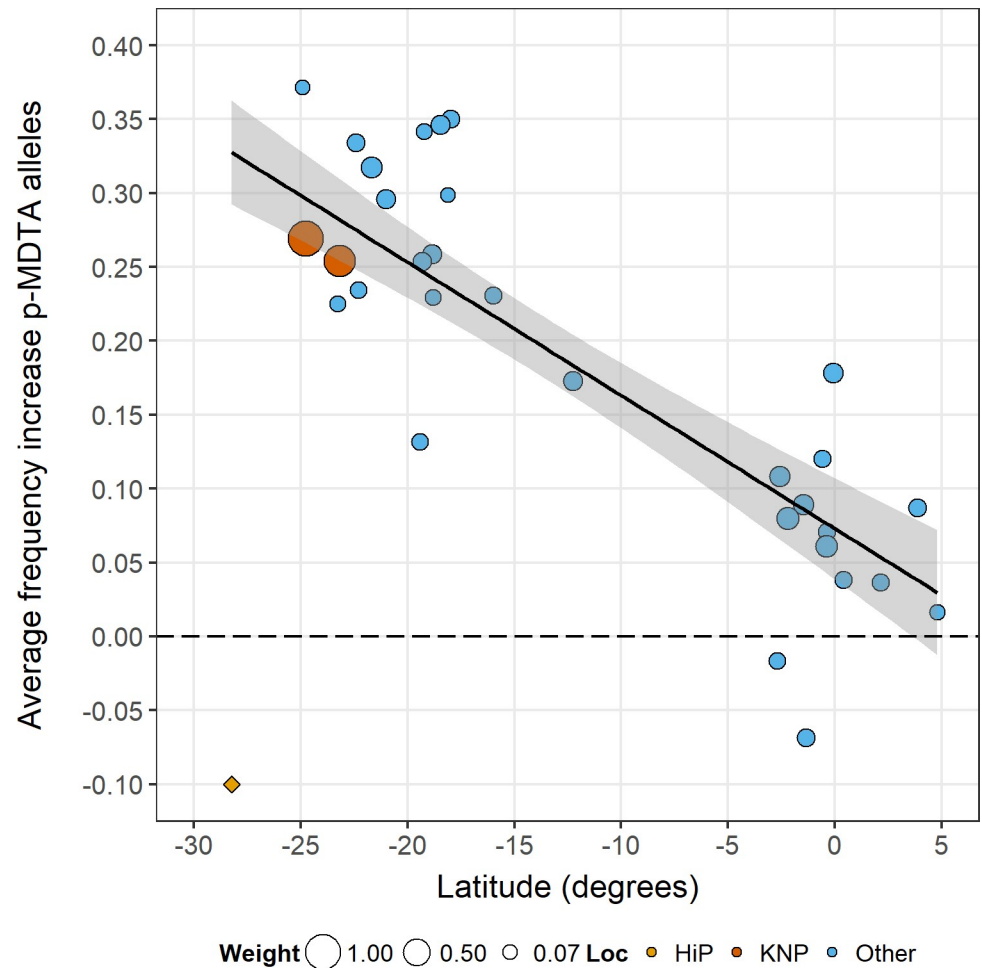


Fig 2. Multilocus-allele-frequency cline.

<https://doi.org/10.1371/journal.pone.0259685.g002>

Linkage disequilibrium

The total fraction of locus pairs with positive r_{LD} across populations and linked loci was 0.66, which was higher than expected under random association (LD Test 1; Wilson score 95% CI: 0.53–0.77, $N = 53$, randomizing genotypes: $P = 0.015$, randomizing allele assignments: $P = 0.0038$; [S7 Table](#)), although these locus pairs were not observed at significantly low latitudes (LD Test 2; $P > 0.35$; however, this test had low statistical power—see [Methods](#) section). The fraction with positive r_{LD} was > 0.5 at all five linked locus pairs (Sign test: $P = 0.063$) but was particularly high for interlocus distances ≥ 19 Mb (three locus pairs: 0.68, $N = 40$; randomizing genotypes: $P = 0.028$, randomizing allele assignments: $P = 0.0052$). Surprisingly, the fraction with positive r_{LD} was also relatively high in East Africa (four locus pairs: 0.74, $N = 19$; randomizing genotypes: $P = 0.031$, randomizing allele assignments: $P = 0.0081$). Both among randomized genotypes at the former five locus pairs and among observed genotypes at the remaining unlinked locus pairs, the total fraction with positive r_{LD} was very close to the expected value of 0.5 (randomized genotypes: 0.500, $N = 2,422,396$; remaining loci: 0.501, $N = 1095$).

On average, the four locus pairs that earlier showed significant LD in KNP based on the non-pooled microsatellite alleles also showed significantly positive r_{LD} values with the p-

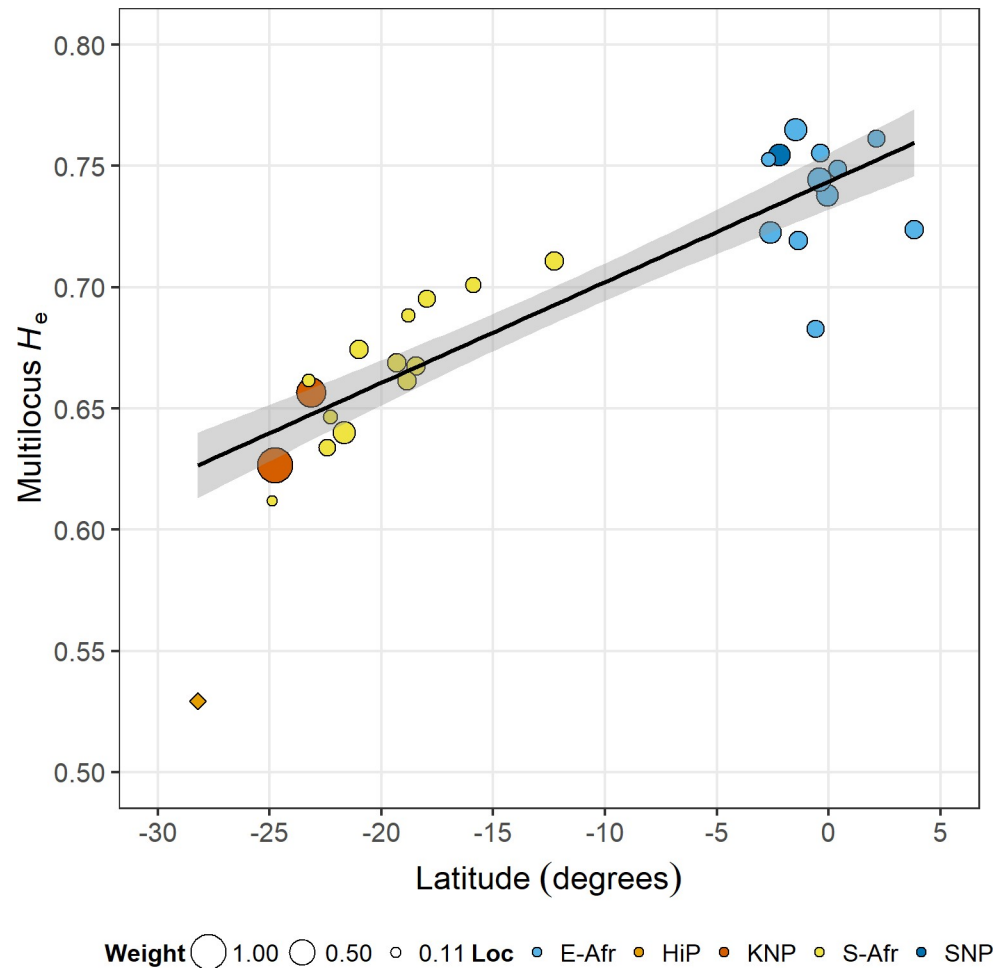


Fig 3. Multilocus- H_e cline.

<https://doi.org/10.1371/journal.pone.0259685.g003>

MDTA alleles, particularly the three neighbouring pairs *BM1824-BM3205*, *BM1824-CSMM019* and *CSMM019-BM3205* (LD Test 3; all pairs: average $r_{LD} = 0.203$, exact $P = 0.019$; neighbouring pairs: average $r_{LD} = 0.284$, exact $P = 0.010$; Table 3). This indicates that LD was not only caused by physical linkage but also by selection. The r_{LD} values based on random allele assignments were close to zero on average (average $r_{LD} = 0.035$, Table 3), while no meaningful correlation in r_{LD} was observed between the neighbouring locus pairs (Pearson $r \leq 0.12$). The latter two observations indicate that, in the absence of selection, strong LD based on non-pooled alleles does not necessarily result in strong LD based on pooled alleles, and that r_{LD} values at neighbouring locus pairs can be statistically independent. The three LD tests based on random allele assignments combined indicate highly significant LD due to selection (Test 1: $Z = 2.48$, after excluding KNP for independence with Test 3; Test 2: $Z = -0.31$; Test 3: $Z = 2.35$; Z-transform test: combined $Z = 2.61$, combined $P = 0.0090$) [64].

Discussion

We observed continent-wide frequency clines of p-MDTA alleles with decreasing frequencies in a northerly direction, which could be attributed to selection ($P = 0.025$). Further, the multi-locus- H_e cline ($P = 0.029$) and the correlation between short- and long-term decrease in per-

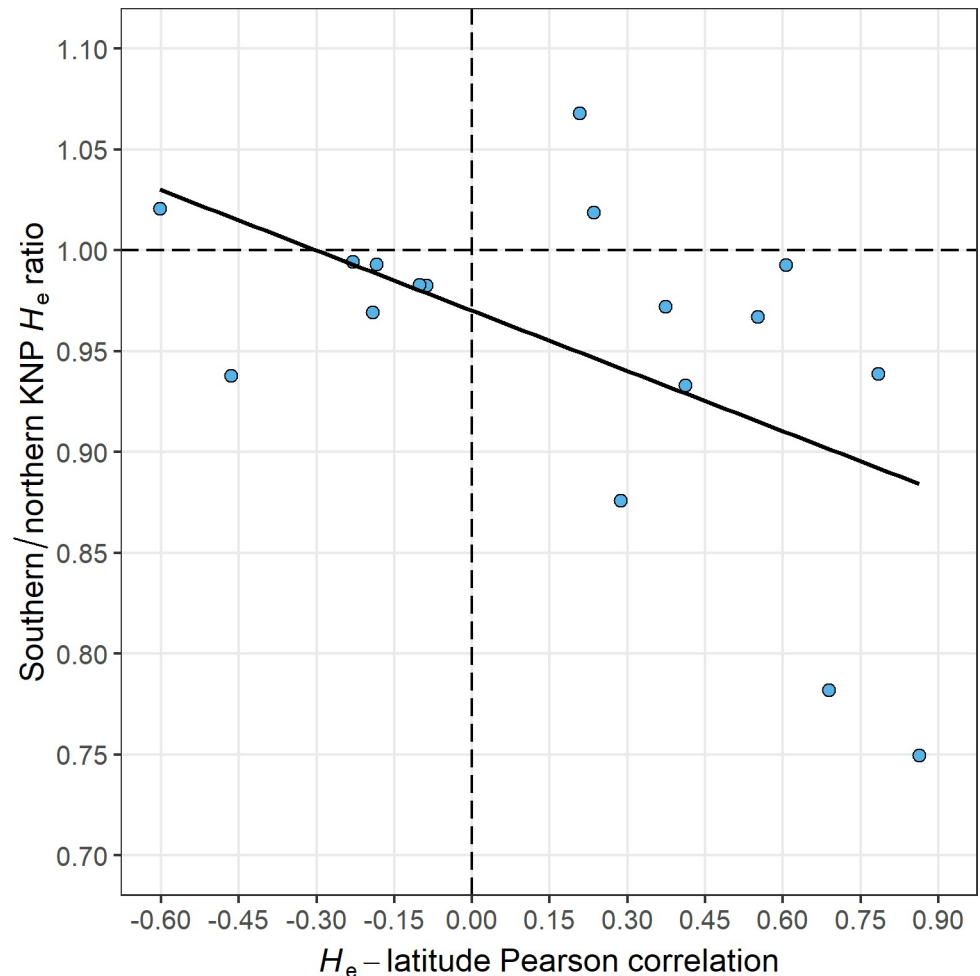


Fig 4. Correlation between per-locus southern/northern KNP H_e ratio and H_e -latitude Pearson correlation.

<https://doi.org/10.1371/journal.pone.0259685.g004>

locus H_e ($P = 0.023$) suggests that via genetic hitchhiking genome-wide selection has had a negative effect on population-level genetic diversity. Although no significant LD cline was observed ($P > 0.35$), both the high fraction of locus pairs with positive r_{LD} and the high average r_{LD} in KNP could be explained by selection ($P \leq 0.015$ and $P = 0.019$, respectively). Positive LD is indicative of genome-wide selection resulting in increased frequencies of haplotypes consisting of multiple deleterious alleles (relative to linkage equilibrium). Relatively high p-MDTA allele frequencies ($P = 0.083$) and significant positive LD ($P \leq 0.031$) in East Africa indicate that the male-deleterious alleles are active across the entire latitudinal range.

Our results may be explained by continent-wide and genome-wide selection for male-deleterious traits, indicative of a high genetic load, with selection strength decreasing from south to north. The only exception is the population in HiP, where male-deleterious alleles appear to be under negative selection (relatively low p-MDTA allele frequencies). The selection pressures in East and southern Africa must be high to prevent destruction of the allele-frequency clines and haplotypes by LD decay. Three statistically independent results, i.e., the allele-frequency clines, the short vs. long-term H_e -decrease correlation and LD, two of which did not rely on interlaboratory allele alignment (short vs. long-term H_e -decrease and LD in KNP), provide strong support for earlier reported observations of genome-wide selection (S1 Table) [20–22].

Allele-frequency clines

Common explanations for allele-frequency clines due to selection assume the advance of a favourable mutation (paradoxically in our case male-deleterious alleles), an environmental gradient, or a combination of the two [28, 65]. The advance of a favourable mutation assumes South Africa as the place of origin of the male-deleterious alleles. However, we consider it unlikely that all male-deleterious alleles originated in a relatively small area. The presence of a linear environmental gradient seems an unlikely explanation considering the wide range of habitats and climates in the sampled region that overall are non-linearly distributed. To our knowledge, there is no linear disease gradient in this region either. Further, the advance of a favourable mutation, an environmental gradient or a disease gradient would not explain the high frequencies of the male-deleterious alleles, which normally are under negative selection [66]. It is also implausible that the allele-frequency clines are a result of the selective agent being present only in the most southern part of the range (e.g., South Africa) considering the occurrence of positive LD in East Africa. LD decay is expected to be very fast in the absence of selection (6–31% per generation at the locus pairs analysed in East Africa), considering the chromosomal distances involved.

Instead of the aforementioned explanations, we posit that the continent-wide allele-frequency clines are caused by the same sex-ratio meiotic gene-drive system that earlier has been hypothesized to be the ultimate selective agent in KNP and HiP. We do so in light of our earlier work showing that both male-deleterious alleles and gene-drive system interact with diseases, such as BTB, and other environmental factors, such as droughts [20–22]. We hypothesize that there is a similar continent-wide frequency cline of gene-drive alleles, also with decreasing frequencies in a northerly direction, considering that in KNP MDTA allele-frequency clines co-occurred with an allele-frequency cline of a Y-chromosomal haplotype linked to a Y-suppressor gene (Y haplotype 557) [22]. Our explanation for the allele-frequency clines implies that frequencies of male-deleterious alleles on the autosomes are maintained at specific equilibrium values in each population depending on the local frequencies of the gene-drive alleles on the sex chromosomes. This would require some form of balancing selection (which does not preclude periods of positive or negative selection such as observed in KNP and HiP, respectively [20–22]).

Genetic hitchhiking

The increase in frequency of male-deleterious alleles to equilibrium can be described as a “partial selective sweep” [30], with each sweep causing a hitchhiking effect and decreased H_e at linked loci, such as observed for the microsatellites in this study. Although selective sweeps are generally observed with directional selection, they can also occur during the early phase of balancing selection [67]. In a recent study in KNP and HiP, decreased H_e at neighbouring loci *BMS1617* and *IFNG* relative to other nearby loci has been attributed to a selective sweep, providing additional support for this scenario [68]. The relatively low microsatellite H_e in southern Africa is unlikely to be caused by low effective population size considering the correlation between short-term hitchhiking effects in southern KNP and long-term hitchhiking effects across Africa, and considering that mtDNA nucleotide diversity (π) is actually considerably higher in southern Africa than in East Africa (D-loop: 0.047 vs. 0.036; complete mitogenome: 0.0048 vs. 0.0035) [69, 70]. The observed correlation between short-term and long-term hitchhiking effects indicates that environmental stressors such as drought and diseases have been consistently acting as selective agents for long periods of time.

Interestingly, pairwise sequentially Markovian coalescent (PSMC) models based on whole-genome sequences indicate relatively high instantaneous coalescence rates since ~20,000 years ago in South Africa, but not in Kenya [71, 72]. According to population genetic theory, high

coalescence rates are associated with low heterozygosity because heterozygosity is proportional to the coalescent time [73]. Although the authors attributed the high coalescence rates in South Africa to a low effective population size, they may also be a signal of selection and hitchhiking effects since the introduction of the aforementioned sex-ratio meiotic gene-drive system [74, 75]. If so, then we predict low heterozygosity to be genome-region specific (i.e., near male-deleterious alleles) and not genome wide. It is relevant to note here that multilocus heterozygosity at microsatellites is a reliable indicator of genome-wide heterozygosity, considering that relative differences in multilocus heterozygosity among South African populations were similar between genome-wide sequence data and microsatellites (observed heterozygosity genome-wide sequence data: KNP = 0.00324, HiP = 0.00261, Addo NP = 0.00222; multilocus H_e microsatellites: KNP = 0.686 [38 loci], HiP = 0.581 [38 loci], Addo NP = 0.447 [18 loci]; S8 Table) [20, 22, 68, 71, 76, 77].

Linkage disequilibrium

Positive LD at linked loci indicates increased frequencies of deleterious-allele haplotypes relative to linkage equilibrium. We detected long-distance LD at genomic distances ≥ 5.9 Mb at four locus pairs; these pairs occur on chromosomes 1 and 3 in cattle. Long-distance LD appears to occur genome-wide considering that significant LD has earlier been observed in KNP at genomic distances of 7.7–17.9 Mb at three locus pairs (six loci) in a 34 Mb region, occurring on chromosome 5 in cattle ($D' = 0.28$ – 0.43 ; i.e., LD is at 28–43% of its maximum possible value) [54, 68]. The observation of long-distance LD in KNP at chromosomal distances ≥ 5.9 Mb is surprising, especially considering that high haploid and autosomal diversity indicate a large historical effective population size for the KNP buffalo (34 mitochondrial D-loop haplotypes with gene diversity $H = 0.94$; 15 Y-chromosomal haplotypes with $H = 0.74$; whole-genome sequencing: high genome-wide observed heterozygosity compared to other mammals: $H = 0.00324$, low median inbreeding coefficient: $f = 0.00060$) [21, 24, 71, 78]. According to population genetic theory, a large effective population size limits long-distance LD because of increased LD decay with chromosomal distance [62].

LD in buffalo populations extends across much larger chromosomal distances than in other natural mammal populations that we are aware of for which LD decay has been estimated. In chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) LD extends to a distance of ~ 0.15 Mb, in Arizona wild mice (*Mus musculus domesticus*) to a distance of ~ 0.2 Mb and in Iberian wild boar (*Sus scrofa*) to a distance of ~ 0.5 Mb. In gray wolf (*Canis lupus*) and coyote (*Canis latrans*) LD extends to a distance of ~ 5 Mb, even in small or bottlenecked populations apart from the wolf population of Isle Royale, which consisted of just 10–30 individuals [79–84]. Further, the half-length of LD (the distance at which LD is 50% of its maximal value) in two isolated Canadian populations of bighorn sheep (*Ovis canadensis*) is only 25%–40% of that in KNP buffalo, despite their small size of less than 200 individuals each (4.6–7.5 Mb vs. ~ 20 Mb at chromosome 5 in buffalo) [68, 85, 86].

LD between distant loci in large outbreeding buffalo populations, despite fast LD decay due to recombination (~ 21 – 31% per generation for three out of six locus pairs), indicates strong selection pressures. Simulation studies, however, indicate that multilocus selection may lower the minimum selection pressure necessary for a given level of LD and may slow down LD decay in a multilocus cline (although even then selection is probably still relatively strong) [31, 87]. Further, allele frequency differences between male and female gametes due to sex-specific selection, as hypothesized in KNP based on genetic data from male and female calves [21], may have resulted in admixture LD between distant loci, especially when selection is strong [32].

Genetic load

A substantial number of high-frequency deleterious alleles of seemingly large effect [21, 22], with many co-occurring in haplotypes (relative to what one would expect if no linkage existed among these alleles), indicates a high genetic load. Most populations of African buffalo, however, seem to have been large in the evolutionary past (long-term N_e of *S. c. caffer*: ~ 48,000) [71] and to be stable after their recovery from the rinderpest pandemic of 1889–1895 that decimated buffalo populations across the whole of Africa [33, 88]. At face value, this seems to support the view, advocated by some population geneticists, that genetic load and genetic diversity in general play a smaller role in ecology than one might expect [6, 89].

To bolster the significance of ecological versus genetic factors in explaining the stability of buffalo populations, we suggest the following five non-mutually exclusive explanations:

1. Male-deleterious alleles are epigenetically suppressed in a large fraction of animals [20, 21].
2. Net-deleterious effects on female health are diminished or even prevented by male-deleterious alleles or haplotypes with negative phenotypic effects in males but positive phenotypic effects in females [22].
3. Negative phenotypic effects only become evident in stressful periods, such as those caused by droughts and disease outbreaks. For example, in 2001 average body condition was lower in southern KNP, a region characterized by relatively high frequencies of MDTA alleles, than in northern KNP, but only significantly so at the end of the dry season [90].
4. Selection is mostly soft, where selective death due to male-deleterious alleles would otherwise be replaced by nonselective death [6].
5. Interspecific competition is limited, thereby minimizing the ecological effects of genetic load [6]. A recent review has failed to find conclusive evidence for interspecific competition between African buffalo and other large mammals [91], which may be due to ecological separation [92].

Conclusions

Our analyses and results reveal that selection resulted in a continent-wide and genome-wide distribution of high-frequency male-deleterious alleles in the African buffalo, with many co-occurring in haplotypes (relative to what one would expect if no linkage existed among these alleles), and decreased population-level genetic diversity via genetic hitchhiking. Since most populations appear to be stable, this suggests that natural populations of mammals, depending on their genetic background, can withstand a high genetic load. Nevertheless, we expect that a high genetic load makes many buffalo populations vulnerable to environmental stressors such as droughts and disease outbreaks. Since buffalo play an important role in the maintenance and transmission of a variety of economically important livestock diseases, which may well have been augmented by a high genetic load, particularly in southern Africa, our results have relevance to livestock husbandry in areas where cattle occur in close proximity to buffalo herds.

Supporting information

S1 Text. Allele size standardization.
(DOCX)

S1 Table. Earlier reported associations of microsatellite alleles with low body condition and bovine tuberculosis (BTB) infection risk.

(DOCX)

S2 Table. Allele size standardization.

(DOCX)

S3 Table. Overview of alleles selected for analysis.

(CSV)

S4 Table. p-MDTA allele frequencies per population.

(CSV)

S5 Table. Per-locus Pearson correlation between ‘p-MDTA minus p-wildtype’ allele frequency difference and latitude, and between H_e and latitude.

(CSV)

S6 Table. H_e per locus per population.

(CSV)

S7 Table. Linkage disequilibrium per population.

(CSV)

S8 Table. H_e per locus in KNP, HiP and Addo NP.

(CSV)

S1 Fig. Increase of pairwise F_{ST} with geographic distance.

(DOCX)

S2 Fig. Multilocus- H_e cline based on twelve microsatellites per locality.

(DOCX)

S3 Fig. Multilocus- H_e cline based on seven microsatellites analysed in both East and southern Africa (microsatellite sets A, B and D) plus *ABS010* and *AGLA293*.

(DOCX)

S4 Fig. Correlation between per-locus southern/northern KNP H_e ratio and H_e -latitude Pearson correlation, based on 19 microsatellites.

(DOCX)

Acknowledgments

Genotype data from microsatellite sets C and E were kindly provided by R. Heller and K. H. Røed, respectively.

Author Contributions

Conceptualization: Pim van Hooft.

Formal analysis: Pim van Hooft.

Investigation: Pim van Hooft.

Visualization: Pim van Hooft.

Writing – original draft: Pim van Hooft.

Writing – review & editing: Pim van Hooft, Wayne M. Getz, Barend J. Greyling, Bas Zwaan, Armanda D. S. Bastos.

References

1. Gravel S. When is selection effective? *Genetics*. 2016; 203(1):451–62. <https://doi.org/10.1534/genetics.115.184630> PMID: 27010021
2. Frankham R, Bradshaw CJA, Brook BW. Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biol Conserv*. 2014; 170:56–63. <https://doi.org/10.1016/j.biocon.2013.12.036> PubMed PMID: WOS:000333774500008.
3. O'Grady JJ, Brook BW, Reed DH, Ballou JD, Tonkyn DW, Frankham R. Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biol Conserv*. 2006; 133(1):42–51. <https://doi.org/10.1016/j.biocon.2006.05.016> PubMed PMID: WOS:000241823900004.
4. Charlesworth D, Willis JH. Fundamental concepts in genetics: The genetics of inbreeding depression. *Nat Rev Genet*. 2009; 10(11):783–96. PubMed PMID: ISI:000270915400014. <https://doi.org/10.1038/nrg2664> PMID: 19834483
5. Nur N. Population growth rate and the measurement of fitness: a critical reflection. *Oikos*. 1987; 48(3):338–41. <https://doi.org/10.2307/3565523>
6. Agrawal AF, Whitlock MC. Mutation Load: The Fitness of Individuals in Populations Where Deleterious Alleles Are Abundant. *Annual Review of Ecology, Evolution, and Systematics*, Vol 43. 2012;43:115–+. <https://doi.org/10.1146/annurev-ecolsys-110411-160257> PubMed PMID: WOS:000311573500006.
7. Bouzat JL. Conservation genetics of population bottlenecks: the role of chance, selection, and history. *Conserv Genet*. 2010; 11(2):463–78. <https://doi.org/10.1007/s10592-010-0049-0> PubMed PMID: WOS:000275455700011.
8. Whiteman NK, Matson KD, Bollmer JL, Parker PG. Disease ecology in the Galapagos Hawk (*Buteo galapagoensis*): host genetic diversity, parasite load and natural antibodies. *Proc R Soc B-Biol Sci*. 2006; 273(1588):797–804. <https://doi.org/10.1098/rspb.2005.3396> PubMed PMID: WOS:000236279800005. PMID: 16618672
9. Reid JM, Arcese P, Keller LF. Inbreeding depresses immune response in song sparrows (*Melospiza melodia*): direct and inter-generational effects. *Proc R Soc B-Biol Sci*. 2003; 270(1529):2151–7. <https://doi.org/10.1098/rspb.2003.2480> PubMed PMID: WOS:000186067000010. PMID: 14561279
10. Ortego J, Aparicio JM, Calabuig G, Cordero PJ. Risk of ectoparasitism and genetic diversity in a wild lesser kestrel population. *Mol Ecol*. 2007; 16(17):3712–20. <https://doi.org/10.1111/j.1365-294X.2007.03406.x> PubMed PMID: WOS:000249242700021. PMID: 17845443
11. Trinkel M, Cooper D, Packer C, Slotow R. Inbreeding depression increases susceptibility to bovine tuberculosis in lions: an experimental test using an inbred-outbred contrast through translocation. *J Wildl Dis*. 2011; 47(3):494–500. <https://doi.org/10.7589/0090-3558-47.3.494> PubMed PMID: WOS:000292798500002. PMID: 21719814
12. Benton CH, Delahay RJ, Smith FAP, Robertson A, McDonald RA, Young AJ, et al. Inbreeding intensifies sex- and age-dependent disease in a wild mammal. *J Anim Ecol*. 2018; 87(6):1500–11. <https://doi.org/10.1111/1365-2656.12878> PubMed PMID: WOS:000446663700002. PMID: 29938787
13. Townsend AK, Taff CC, Wheeler SS, Weis AM, Hinton MG, Jones ML, et al. Low heterozygosity is associated with vector-borne disease in crows. *Ecosphere*. 2018; 9(10):13. <https://doi.org/10.1002/ecs2.2407> PubMed PMID: WOS:000449566200023.
14. Queiros J, Vicente J, Alves PC, de la Fuente J, Gortazar C. Tuberculosis, genetic diversity and fitness in the red deer, *Cervus elaphus*. *Infect Genet Evol*. 2016; 43:203–12. <https://doi.org/10.1016/j.meegid.2016.05.031> PubMed PMID: WOS:000379701100027. PMID: 27245150
15. Goedbloed DJ, Van Hooft P, Megens HJ, Bosch T, Lutz W, Van Wieren SE, et al. Host genetic heterozygosity and age are important determinants of porcine circovirus type 2 disease prevalence in European wild boar. *Eur J Wildl Res*. 2014; 60:803–10. <https://doi.org/10.1007/s10344-014-0850-z>
16. Coltman DW, Pilkington JG, Smith JA, Pemberton JM. Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution*. 1999; 53(4):1259–67. PubMed PMID: ISI:000082389200024. <https://doi.org/10.1111/j.1558-5646.1999.tb04538.x> PMID: 28565537
17. Townsend AK, Clark AB, McGowan KJ, Buckles EL, Miller AD, Lovette IJ. Disease-mediated inbreeding depression in a large, open population of cooperative crows. *Proceedings of the Royal Society B: Biological Sciences*. 2009; 276(1664):2057–64. <https://doi.org/10.1098/rspb.2008.1852> PMID: 19324784
18. Acevedo-Whitehouse K, Gulland F, Greig D, Amos W. Disease susceptibility in California sea lions. *Nature*. 2003; 422(6927):35. <https://doi.org/10.1038/422035a> PMID: 12621424

19. Acevedo-Whitehouse K, Spraker TR, Lyons E, Melin SR, Gulland F, DeLong RL, et al. Contrasting effects of heterozygosity on survival and hookworm resistance in California sea lion pups. *Mol Ecol*. 2006; 15(7):1973–82. <https://doi.org/10.1111/j.1365-294X.2006.02903.x> PMID: 16689912
20. van Hooft P, Getz WM, Greyling BJ, Bastos ADS. A natural gene drive system influences bovine tuberculosis susceptibility in African buffalo: possible implications for disease management. *PLoS ONE*. 2019; 14(9):e0221168. <https://doi.org/10.1371/journal.pone.0221168> PMID: 31483802
21. van Hooft P, Dougherty ER, Getz WM, Greyling BJ, Zwaan BJ, Bastos ADS. Genetic responsiveness of African buffalo to environmental stressors: a role for epigenetics in balancing autosomal and sex chromosome interactions? *PLoS ONE*. 2018; 13(2):e0191481. <https://doi.org/10.1371/journal.pone.0191481> PMID: 29415077
22. van Hooft P, Greyling BJ, Getz WM, van Helden PD, Zwaan BJ, Bastos ADS. Positive selection of deleterious alleles through interaction with a sex-ratio suppressor gene in African buffalo: a plausible new mechanism for a high frequency anomaly. *PLoS ONE*. 2014; 9(11):e111778. <https://doi.org/10.1371/journal.pone.0111778> PMID: 25372610
23. Lindholm AK, Dyer KA, Firman RC, Fishman L, Forstmeier W, Holman L, et al. The ecology and evolutionary dynamics of meiotic drive. *Trends Ecol Evol*. 2016; 31(4):315–26. <https://doi.org/10.1016/j.tree.2016.02.001> PubMed PMID: WOS:269204739900009.
24. van Hooft P, Greyling BJ, Prins HHT, Getz WM, Jolles AE, Bastos ADS. Selection at the Y chromosome of the African buffalo driven by rainfall. *PLoS ONE*. 2007; 2(10): <https://doi.org/10.1371/journal.pone.0001086>
25. van Hooft P, Prins HHT, Getz WM, Jolles AE, van Wieren SE, Greyling BJ, et al. Rainfall-driven sex-ratio genes in African buffalo suggested by correlations between Y-chromosomal haplotype frequencies and foetal sex ratio. *BMC Evol Biol*. 2010; 10:10.1186/471-2148-10-106. <https://doi.org/10.1186/1471-2148-10-106> PubMed PMID: MEDLINE:20416038. PMID: 20416038
26. May RM, Endler JA, McMurtrie RE. Gene frequency clines in presence of selection opposed by gene flow. *Am Nat*. 1975; 109(970):659–76. <https://doi.org/10.1086/283036> PubMed PMID: WOS: A1975BC40400004. PMID: 29513550
27. Slatkin M. Gene flow and selection in a cline. *Genetics*. 1973; 75(4):733–56. PubMed PMID: WOS: A1973S097400015. <https://doi.org/10.1093/genetics/75.4.733> PMID: 4778791
28. Charlesworth B, Charlesworth D. *Elements of evolutionary genetics*. Greenwood Village: Roberts and Company Publishers; 2010. 734 p. <https://doi.org/10.1038/ng0910-734> PMID: 20802475
29. Connallon T, Clark AG. Antagonistic versus nonantagonistic models of balancing selection: Characterizing the relative timescales and hitchhiking effects of partial selective sweeps. *Evolution*. 2013; 67(3):908–17. <https://doi.org/10.1111/j.1558-5646.2012.01800.x> PMID: 23461340
30. Coop G, Ralph P. Patterns of neutral diversity under general models of selective sweeps. *Genetics*. 2012; 192(1):205–24. <https://doi.org/10.1534/genetics.112.141861> PMID: 22714413
31. Hastings A. Linkage disequilibrium, selection and recombination at 3 loci. *Genetics*. 1984; 106(1):153–64. PubMed PMID: WOS:A1984RW57300011. <https://doi.org/10.1093/genetics/106.1.153> PMID: 17246189
32. Úbeda F, Haig D, Patten MM. Stable linkage disequilibrium owing to sexual antagonism. *Proc R Soc B Biol Sci*. 2011; 278(1707):855–62. <https://doi.org/10.1098/rspb.2010.1201> PMID: 20861051
33. van Hooft WF, Groen AF, Prins HHT. Microsatellite analysis of genetic diversity in African buffalo (*Syncerus caffer*) populations throughout Africa. *Mol Ecol*. 2000; 9(12):2017–25. PubMed PMID: ISI:000166112700008. <https://doi.org/10.1046/j.1365-294x.2000.01101.x> PMID: 11123614
34. Smitz N, Cornelis D, Chardonnet P, Caron A, de Garine-Wichatitsky M, Jori F, et al. Genetic structure of fragmented southern populations of African Cape buffalo (*Syncerus caffer caffer*). *BMC Evol Biol*. 2014;14. doi: 203 <https://doi.org/10.1186/1471-2148-14-14> PubMed PMID: WOS:000344956900001. PMID: 24450606
35. Heller R, Lorenzen ED, Okello JBA, Masembe C, Siegmund HR. Mid-Holocene decline in African buffaloes inferred from Bayesian coalescent-based analyses of microsatellites and mitochondrial DNA. *Mol Ecol*. 2008; 17(22):4845–58. <https://doi.org/10.1111/j.1365-294X.2008.03961.x> PubMed PMID: WOS:000260735300007. PMID: 19140976
36. Ernest EM, Haanes H, Bitanyi S, Fyumagwa RD, Msoffe PL, Bjornstad G, et al. Influence of habitat fragmentation on the genetic structure of large mammals: evidence for increased structuring of African buffalo (*Syncerus caffer*) within the Serengeti ecosystem. *Conserv Genet*. 2012; 13(2):381–91. <https://doi.org/10.1007/s10592-011-0291-0> PubMed PMID: WOS:000303476600009.
37. Epps CW, Castillo JA, Schmidt-Kuentzel A, du Preez P, Stuart-Hill G, Jago M, et al. Contrasting Historical and Recent Gene Flow among African Buffalo Herds in the Caprivi Strip of Namibia. *J Hered*. 2013; 104(2):172–81. <https://doi.org/10.1093/jhered/ess142> PubMed PMID: WOS:000315155900002. PMID: 23341534

38. Greyling BJ, Kryger P, du Plessis S, van Hooft WF, van Helden P, Getz WM, et al. Development of a high-throughput microsatellite typing approach for forensic and population genetic analysis of wild and domestic African Bovini. *Afr J Biotechnol*. 2008; 7(5):655–60. PubMed PMID: ISI:000254900200026.
39. van Hooft P, Greyling BJ, Getz WM, Heller R, Røed KH, Bastos ADS. Microsatellite data from various African buffalo (*Syncerus caffer*) populations throughout Africa. 2021. <https://doi.org/10.5061/dryad.s7h44j14c>
40. Peakall R, Smouse PE. GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*. 2012; 28(19):2537–9. <https://doi.org/10.1093/bioinformatics/bts460> PubMed PMID: WOS:000309687500024. PMID: 22820204
41. Peakall R, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes*. 2006; 6(1):288–95. <https://doi.org/10.1111/j.1471-8286.2005.01155.x> PubMed PMID: WOS:000235725600089.
42. LaHood ES, Moran P, Olsen J, Grant WS, Park LK. Microsatellite allele ladders in two species of Pacific salmon: preparation and field-test results. *Mol Ecol Notes*. 2002; 2(2):187–90. <https://doi.org/10.1046/j.1471-8286.2002.00174.x> PubMed PMID: WOS:000176371900036.
43. Ellis JS, Gilbey J, Armstrong A, Balstad T, Cauwelier E, Cherbonnel C, et al. Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (*Salmo salar* L.). *Genetica*. 2011; 139(3):353–67. <https://doi.org/10.1007/s10709-011-9554-4> PubMed PMID: WOS:000288514100007. PMID: 21279823
44. Presson AP, Sobel E, Lange K, Papp JC. Merging microsatellite data. *J Comput Biol*. 2006; 13(6):1131–47. <https://doi.org/10.1089/cmb.2006.13.1131> PubMed PMID: WOS:000240091600001. PMID: 16901233
45. Paiva SR, Mariante AD, Blackburn HD. Combining US and Brazilian Microsatellite Data for a Meta-Analysis of Sheep (*Ovis aries*) Breed Diversity: Facilitating the FAO Global Plan of Action for Conserving Animal Genetic Resources. *J Hered*. 2011; 102(6):697–704. <https://doi.org/10.1093/jhered/esr101> PubMed PMID: WOS:000296100900006. PMID: 22013019
46. Presson AP, Sobel EM, Pajukanta P, Plaisier C, Weeks DE, Åberg K, et al. Merging microsatellite data: Enhanced methodology and software to combine genotype data for linkage and association analysis. *BMC Bioinformatics*. 2008;9. <https://doi.org/10.1186/1471-2105-9-9> PMID: 18182098
47. Smolyanitsky AG, Khromov-Borisov NN, Perepechina IO, Ivanov PL, Kornienko IV, Zamaraev VS, et al. Towards Russian reference population data on STR loci. *International Congress Series 2004*. p. 242–4.
48. Zhang Y, Vankan D, Zhang Y, Barker JSF. Genetic differentiation of water buffalo (*Bubalus bubalis*) populations in China, Nepal and south-east Asia: Inferences on the region of domestication of the swamp buffalo. *Anim Genet*. 2011; 42(4):366–77. <https://doi.org/10.1111/j.1365-2052.2010.02166.x> PMID: 21749419
49. Alexander RA. A note on averaging correlations. *Bulletin of the Psychonomic Society*. 1990; 28(4):335–6. <https://doi.org/10.3758/BF03334037>
50. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*. 1978; 89:583–90. <https://doi.org/10.1093/genetics/89.3.583> PMID: 17248844
51. Park SDE. Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection [PhD]. Dublin: University of Dublin; 2001.
52. Goudet J. FSTAT (Version 1.2): A computer program to calculate F-statistics. *J Hered*. 1995; 86(6):485–6. PubMed PMID: ISI:A1995TL74700013.
53. Goudet J. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Updated from Goudet (1995). 2001.
54. Ihara N, Takasuga A, Mizoshita K, Takeda H, Sugimoto M, Mizoguchi Y, et al. A comprehensive genetic map of the cattle genome based on 3802 microsatellites. *Genome Res*. 2004; 14(10A):1987–98. <https://doi.org/10.1101/gr.2741704> PubMed PMID: WOS:000224405900018. PMID: 15466297
55. Gallagher DS, Womack JE. Chromosome conservation in the bovidae. *J Hered*. 1992; 83(4):287–98. PubMed PMID: WOS:A1992JJ19600008. <https://doi.org/10.1093/oxfordjournals.jhered.a111215> PMID: 1401875
56. Weng ZQ, Saatchi M, Schnabel RD, Taylor JF, Garrick DJ. Recombination locations and rates in beef cattle assessed from parent-offspring pairs. *Genet Sel Evol*. 2014; 46:14. <https://doi.org/10.1186/1297-9686-46-14> PubMed PMID: WOS:000338459100002. PMID: 24552175
57. Guerrero RF, Kirkpatrick M. Local adaptation and the evolution of chromosome fusions. *Evolution*. 2014; 68(10):2747–56. <https://doi.org/10.1111/evo.12481> PMID: 24964074
58. Schaid DJ. Linkage disequilibrium testing when linkage phase is unknown. *Genetics*. 2004; 166(1):505–12. <https://doi.org/10.1534/genetics.166.1.505> PubMed PMID: WOS:000220021800040. PMID: 15020439

59. Weir BS. Genetic Data Analysis. Sunderland: Sinauer Associates, Inc.; 1990.
60. Xu L, Zhu B, Wang Z, Xu L, Liu Y, Chen Y, et al. Evaluation of linkage disequilibrium, effective population size and haplotype block structure in Chinese cattle. *Animals*. 2019; 9(3). <https://doi.org/10.3390/ani9030083> PMID: 30845681
61. Khatkar MS, Nicholas FW, Collins AR, Zenger KR, Cavanagh JAL, Barris W, et al. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics*. 2008;9. <https://doi.org/10.1186/1471-2164-9-9> PMID: 18186939
62. Slatkin M. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*. 2008; 9(6):477–85. <https://doi.org/10.1038/nrg2361> PubMed PMID: WOS:000255953500015. PMID: 18427557
63. Zhao H, Nettleton D, Soller M, Dekkers JCM. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genet Res*. 2005; 86(1):77–87. <https://doi.org/10.1017/S001667230500769X> PMID: 16181525
64. Whitlock MC. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J Evol Biol*. 2005; 18(5):1368–73. <https://doi.org/10.1111/j.1420-9101.2005.00917.x> PubMed PMID: WOS:000231508900031. PMID: 16135132
65. de Jong MA, Collins S, Beldade P, Brakefield PM, Zwaan BJ. Footprints of selection in wild populations of *Bicyclus anynana* along a latitudinal cline. *Mol Ecol*. 2013; 22(2):341–53. <https://doi.org/10.1111/mec.12114> PubMed PMID: WOS:000312996600007. PMID: 23171126
66. Henn BM, Botigue LR, Bustamante CD, Clark AG, Gravel S. Estimating the mutation load in human genomes. *Nat Rev Genet*. 2015; 16(6):333–43. <https://doi.org/10.1038/nrg3931> PubMed PMID: WOS:000354655800009. PMID: 25963372
67. Weigand H, Leese F. Detecting signatures of positive selection in non-model species using genomic data. *Zoological Journal of the Linnean Society*. 2018; 184(2):528–83. <https://doi.org/10.1093/zoolinnean/zly007>
68. Lane-deGraaf KE, Amish SJ, Gardipee F, Jolles A, Luikart G, Ezenwa VO. Signatures of natural and unnatural selection: evidence from an immune system gene in African buffalo. *Conserv Genet*. 2015; 16(2):289–300. <https://doi.org/10.1007/s10592-014-0658-0> PubMed PMID: WOS:000351287400004.
69. Smitz N, Berthouly C, Cornélis D, Heller R, Van Hooft P, Chardonnet P, et al. Pan-African Genetic Structure in the African Buffalo (*Syncerus caffer*): Investigating Intraspecific Divergence. *PLoS ONE*. 2013; 8(2):e56235. <https://doi.org/10.1371/journal.pone.0056235> PMID: 23437100
70. Heller R, Brueniche-Olsen A, Siegismund HR. Cape buffalo mitogenomics reveals a Holocene shift in the African human-megafauna dynamics. *Mol Ecol*. 2012; 21(16):3947–59. <https://doi.org/10.1111/j.1365-294X.2012.05671.x> PubMed PMID: WOS:000306897500006. PMID: 22725969
71. de Jager D, Glanzmann B, Möller M, Hoal E, van Helden P, Harper C, et al. High diversity, inbreeding and a dynamic Pleistocene demographic history revealed by African buffalo genomes. *Sci Rep*. 2021; 11(1). <https://doi.org/10.1038/s41598-021-83823-8> PMID: 33633171
72. Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, et al. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science*. 2019; 364(6446). <https://doi.org/10.1126/science.aav6202> PMID: 31221828
73. Campbell RB. Coalescent size versus coalescent time with strong selection. *Bulletin of Mathematical Biology*. 2007; 69(7):2249–59. <https://doi.org/10.1007/s11538-007-9218-9> PMID: 17546476
74. Palamara PF, Terhorst J, Song YS, Price AL. High-throughput inference of pairwise coalescence times identifies signals of selection and enriched disease heritability. *Nature Genet*. 2018; 50(9):1311–7. <https://doi.org/10.1038/s41588-018-0177-x> PMID: 30104759
75. Johri P, Riall K, Becher H, Excoffier L, Charlesworth B, Jensen JD. The impact of purifying and background selection on the inference of population history: problems and prospects. *Mol Biol Evol*. 2021. <https://doi.org/doi.org/10.1093/molbev/msab050>
76. de Jager D, Harper CK, Bloomer P. Genetic diversity, relatedness and inbreeding of ranched and fragmented Cape buffalo populations in southern Africa. *PLoS ONE*. 2020; 15(8 August). <https://doi.org/10.1371/journal.pone.0236717> PMID: 32797056
77. O'Ryan C, Harley EH, Bruford MW, Beaumont M, Wayne RK, Cherry MI. Microsatellite analysis of genetic diversity in fragmented South African buffalo populations. *Animal Conservation*. 1998; 1(2):85–94. <https://doi.org/10.1111/j.1469-1795.1998.tb00015.x>
78. Greyling BJ. Genetic variation, structure and dispersal among Cape buffalo populations from the Hluhluwe-Imfolozi and Kruger National Parks of South Africa [PhD thesis]. Pretoria: University of Pretoria; 2007.
79. Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, et al. Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics*. 2009; 181(4):1493–505. <https://doi.org/10.1534/genetics.108.098830> PubMed Central PMCID: PMC19189949. PMID: 19189949

80. Laurie CC, Nickerson DA, Anderson AD, Weir BS, Livingston RJ, Dean MD, et al. Linkage disequilibrium in wild mice. *PLoS Genet.* 2007; 3(8):1487–95. <https://doi.org/10.1371/journal.pgen.0030144> PubMed Central PMCID: PMC17722986. PMID: 17722986
81. Munoz M, Bozzi R, Garcia-Casco J, Nunez Y, Ribani A, Franci O, et al. Genomic diversity, linkage disequilibrium and selection signatures in European local pig breeds assessed with a high density SNP chip. *Sci Rep.* 2019; 9:14. <https://doi.org/10.1038/s41598-018-36846-7> PubMed PMID: WOS:000486567800022. PMID: 30626888
82. De Manuel M, Kuhlwlilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, et al. Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science.* 2016; 354(6311):477–81. <https://doi.org/10.1126/science.aag2602> PubMed Central PMCID: PMC27789843. PMID: 27789843
83. Herrero-Medrano JM, Megens HJ, Groenen MAM, Ramis G, Bosse M, Pérez-Enciso M, et al. Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. *BMC Genet.* 2013;14. <https://doi.org/10.1186/1471-2156-14-14> PMID: 23448225
84. Pilot M, Malewski T, Moura AE, Grzybowski T, Oleński K, Ruść A, et al. On the origin of mongrels: Evolutionary history of free-breeding dogs in Eurasia. *Proceedings of the Royal Society B: Biological Sciences.* 2015; 282(1820). <https://doi.org/10.1098/rspb.2015.2189> PMID: 26631564
85. Miller JM, Poissant J, Malenfant RM, Hogg JT, Coltman DW. Temporal dynamics of linkage disequilibrium in two populations of bighorn sheep. *Ecol Evol.* 2015; 5(16):3401–12. <https://doi.org/10.1002/ece3.1612> PMID: 26380673
86. Miller JM, Poissant J, Kijas JW, Coltman DW, Int Sheep Genomics C. A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Mol Ecol Resour.* 2011; 11(2):314–22. <https://doi.org/10.1111/j.1755-0998.2010.02918.x> PubMed PMID: WOS:000287200200010. PMID: 21429138
87. Baird SJE. A simulation study of multilocus clines. *Evolution.* 1995; 49(6):1038–45. <https://doi.org/10.1111/j.1558-5646.1995.tb04431.x> PubMed PMID: WOS:A1995TR69500002. PMID: 28568511
88. Bengis RG, Kock RA, Fischer J. Infectious animal diseases: the wildlife/livestock interface. *Rev Sci Tech Off Int Epizoot.* 2002; 21(1):53–65. PubMed PMID: ISI:000174945900004. <https://doi.org/10.20506/rst.21.1.1322> PMID: 11974630
89. Teixeira JC, Huber CD. The inflated significance of neutral genetic diversity in conservation genetics. *Proc Natl Acad Sci U S A.* 2021; 118(10). <https://doi.org/10.1073/pnas.2015096118> PMID: 33608481
90. Caron A, Cross PC, Du Toit JT. Ecological implications of bovine tuberculosis in African buffalo herds. *Ecol Appl.* 2003; 13(5):1338–45. PubMed PMID: ISI:000186599400013.
91. Prins HHT. Interspecific resource competition in antelopes: search for evidence. In: Bro-Jørgensen J, Mallon DP, editors. *Antelope conservation in the 21st century: from diagnosis to action.* Conservation Science and Practice Series. Oxford: Wiley-Blackwell; 2016.
92. Traill LW. Seasonal utilization of habitat by large grazing herbivores in semi-arid Zimbabwe. *South Afr J Wildl Res.* 2004; 34(1):13–24. PubMed PMID: WOS:000223168300002.