# A continent-wide high genetic load in African buffalo revealed by a cline in the frequency of deleterious alleles and linkage disequilibrium

Pim van Hooft[1], Wayne Getz[2,3], Ben Greyling[4], Bas Zwaan[1], and Armanda Bastos[5]

[1]Wageningen University & Research
[2]University of California, Berkeley
[3]University of California Berkeley
[4]Agricultural Research Council
[5]University of Pretoria

May 8, 2020

## Abstract

Recent microsatellite studies of two African buffalo (Syncerus caffer) populations in South Africa indicated a high genetic load due to genome-wide high-frequency occurrence of deleterious alleles. These alleles have a negative effect on two male traits: body condition and bovine tuberculosis resistance. Here, we study the occurrence of these male-deleterious alleles throughout most of the buffalo's range using previously published microsatellite data (2-17 microsatellite loci genotyped on 1676 animals from 34 localities). We uncovered a continent-wide frequency cline of microsatellite alleles associated with male-deleterious traits, with frequencies decreasing from 25 °S to 5 °N (adjusted R2 = 0.87). Further, there was a continent-wide presence of linkage disequilibrium (LD) between male-deleterious-trait-associated alleles at five linked locus pairs (interlocus distance: 0.5-28 Megabase). The fraction of positive interlocus associations among these locus pairs was 0.65 (95% CI: 0.52-0.77; expected fraction with free recombination: 0.5), indicating that many male-deleterious alleles co-occur in haplotypes. We argue that the allele-frequency cline and LD likely result from genome-wide balancing selection of male-deleterious alleles with equilibrium allele and haplotype frequencies determined by the activity of a sex-chromosomal gene-drive system, the latter which was indicated in earlier studies and seemingly originated in southern Africa. The selection pressures involved must be high to prevent destruction of the allele-frequency cline and haplotypes by LD decay. Since most buffalo populations are stable, our results indicate that natural mammal populations can withstand a high genetic load. Nevertheless, we expect that a high genetic load makes many buffalo populations relatively sensitive to environmental stresses.

## Keywords

allele frequency cline, deleterious allele, genetic load, linkage disequilibrium, microsatellite, *Syncerus caffer*

## Introduction

Several microsatellite studies on African buffalo (*Syncerus caffe* r) from Kruger National Park (KNP) and Hluhluwe-iMfolozi Park (HiP), South Africa, have identified deleterious alleles with a negative effect on male body condition and resistance to bovine tuberculosis (BTB) (van Hooft et al., 2018; van Hooft, Getz, Greyling, & Bastos, 2019; van Hooft et al., 2014). Two types of microsatellites were observed: one type containing alleles associated with negative phenotypic effects in both sexes (deleterious-effect—DE—associated loci and alleles), and one type containing alleles associated with negative phenotypic effects in males but positive phenotypic effects in females (sexually-antagonistic-effect—SAE—associated loci and alleles). These microsatellite alleles are probably linked to male-deleterious alleles (from hereon, DE, SAE and male-

1

deleterious-trait-associated alleles will refer to microsatellite alleles and male-deleterious alleles to alleles at protein-coding genes). The male-deleterious alleles probably occur genome-wide at high frequencies in both populations and seem to be of large effect considering the notable frequency differences between year-cohorts and between unhealthy (BTB-positive and low body condition) and healthy (BTB-negative and high body condition) males.

Male-deleterious alleles seemed to have high frequencies and be under positive selection in KNP despite the negative phenotypic effects (van Hooft et al., 2014). Positive selection has been attributed to a sex-ratio meiotic gene-drive system (van Hooft et al., 2018; van Hooft et al., 2019; van Hooft et al., 2014; van Hooft et al., 2010). It was hypothesized that poor health (due to low body condition and BTB infection) in males suppresses sex-ratio distortion genes in this gene-drive system, which when active results in reduced fertility. As a consequence, any allele that has a negative effect on male health may have a positive effect on male relative fertility and thereby be under positive selection if the net fitness effect on health and fertility across both sexes is positive. However, in contrast to positive selection of male-deleterious alleles in KNP, selection of male-deleterious alleles appears to be negative in HiP, which is situated just 280 km further south, resulting in relatively low male-deleterious allele frequencies compared to KNP (van Hooft et al., 2019). This negative selection has been attributed to incompleteness of the gene-drive system, as discussed elsewhere (van Hooft et al., 2018; van Hooft et al., 2019).

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) DE and SAE microsatellite alleles, across a large part of southern Africa (van Hooft et al., 2018). Their range possibly extends as far as East Africa, considering the high DE allele frequencies in this region (average frequency Kruger: 0.69, average frequency East Africa: 0.47) (van Hooft et al., 2014). The interplay between migration (gene flow) and selection may have resulted in allele-frequency clines, which would be a strong indicator of selection acting across a wide geographic range (Charlesworth & Charlesworth, 2010; May, Endler, & McMurtrie, 1975; Slatkin, 1973). Such clines have previously been observed in KNP, not only for the DE and SAE alleles but also for a Y-chromosomal haplotype hypothesized to be linked to a suppressor gene from the gene-drive system (van Hooft et al., 2014). Further, multilocus selection of male-deleterious alleles at short chromosomal distances may have resulted in linkage disequilibrium (LD) due to increased frequencies of haplotypes consisting of multiple male-deleterious alleles (relative to haplotype frequencies expected under linkage equilibrium) (Hastings, 1984). This may be particularly so if sex-specific selection results in admixture LD due to different allele frequencies in male and female gametes (Úbeda, Haig, & Patten, 2011). Such differences have earlier been hypothesized in KNP for both DE and SAE alleles based on genetic data from male and female calves (van Hooft et al., 2018). In combination with a selection gradient, multilocus sex-specific selection may have resulted in an LD cline with highest LD where selection is strongest. A wide distribution of male-deleterious alleles may have substantial management implications, because it suggests that many African buffalo populations experience a high genetic load (reduction in relative fitness due to genetic factors) and thereby may be relatively sensitive to environmental stresses.

In this study, we analysed previously published microsatellite data from 1676 African buffalo from 34 localities across the African continent (Figure 1). We addressed the following four questions. 1) Do DE and SAE alleles occur throughout the range of African buffalo? 2) Do the spatial distributions of alleles constitute allele-frequency clines? 3) Can allele-frequency clines, if present, be attributed to selection? 4) Did selection result in an additional LD cline?

**Materials and methods**

*Localities and microsatellites*

We analysed 1676 animals from 34 localities throughout the buffalo's range (Figure 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). Samples from Bénoué NP and Lopé NP (separated by $\pm$ 1000 km) in central Africa and from Mana Pools

NP and Nyakasanga (separated by $\pm$ 55 km) in southern Africa were pooled because of small sample size. Intermediate coordinates, weighted by sample size, were used for these pooled samples.

We used previously published genotype data of the following seventeen microsatellites: *BM0719* , *BM1824* , *BM3517* ,*BM3205* , *BM4028* , *CSSM019* , *DIK020* ,*ETH010* , *ETH225* , *ILSTS026* , *INRA006*,*INRA128* , *SPS115* , *TGLA057* , *TGLA0159* ,*TGLA227* and *TGLA263* (Epps et al., 2013; Ernest et al., 2012; Greyling et al., 2008; Heller, Lorenzen, Okello, Masembe, & Siegismund, 2008; Smitz et al., 2014; van Hooft et al., 2018; van Hooft et al., 2019; van Hooft et al., 2014; van Hooft, Groen, & Prins, 2000). Data for all seventeen microsatellites were only available for KNP and HiP (microsatellite set A in Table 2). Five smaller subsets, consisting of 2-11 microsatellites, were available for the other populations (microsatellite set B-F in Table 2). Raw microsatellite data from microsatellite sets D and F are provided in earlier publications (Epps et al., 2013; Smitz et al., 2014), while those from the remaining sets are available from the Dryad Digital Repository (van Hooft et al., 2020).

*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 6.503) (Peakall & Smouse, 2006; Peakall & Smouse, 2012). Standardization was necessary, because allele size estimates have been observed to differ up to 7 bp among platforms (LaHood, Moran, Olsen, Grant, & Park, 2002). In our case they differed up to 11 bp (microsatellite set B relative to set A:*BM3517* : 8 bp, *INRA006* : 11 bp, [?]5 bp in the remaining cases). For various populations, two microsatellite sets were available (sets A-D in Tables 1 and 2), which permitted standardization by shifting allele sizes by a constant amount in one of the two sets. With two microsatellites sets (sets E and F in Tables 1 and 2) standardization was only possible in comparison with another microsatellite set applied to another population. The 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. Using the standardized data to regress pairwise multilocus $F_{ST}$ values on geographic distance explained 78% of the genetic variation among populations in East and southern Africa (Figure S1), thereby providing further support for the accuracy of the allele size standardization. A detailed description of the allele size standardization is given in Text S1.

*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 (van Hooft et al., 2018; van Hooft et al., 2014). 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 (observed <15 times), as being potentially 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. The sample sizes of these two male groups were similar, $N = 35$ and $N = 39$ respectively. 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 (van Hooft et al., 2019). In contrast to that study, we did not make a further distinction among alleles according to their association with female-beneficial traits. Thus, as discussed elsewhere (van Hooft et al., 2019), we assumed two alleles per coding gene with the aforementioned sexually-antagonistic trait associations due to linkage with a second coding gene.

As in earlier studies, we made a distinction between two groups of microsatellites based on whether their alleles showed highest frequency in low or high body-condition individuals from southern KNP (van Hooft et al., 2018; van Hooft et al., 2019; van Hooft et al., 2014). From our first group of eight microsatellites (*BM3517* , *BM4028* ,*ETH010* , *ETH225* , *INRA006* , *INRA128* ,*TGLA227* and *TGLA263* ) we identified eight high-frequency alleles, one at each locus (frequency > 0.63 in all eight cases), that were associated with low body condition in both sexes and thus likely linked to sex-independent deleterious alleles (van Hooft

et al., 2014). As in earlier studies, we refer to this group as the deleterious-effect-associated loci (DE loci) (van Hooft et al., 2018; van Hooft et al., 2019; van Hooft et al., 2014). From our second group of nine microsatellites (*BM0719* , *BM1824* , *BM3205* ,*CSSM019* , *DIK020* , *ILSTS026* , *SPS115* ,*TGLA057* and *TGLA159* ) we identified a number of alleles (36 out of 53, excluding rare alleles with frequency <0.05) that showed sexually antagonistic associations. They were either associated with low body condition in males and high body condition in females (19 alleles) or *vice versa* (17 alleles). Alleles of this second group are assumed to be linked to haplotypes consisting of closely linked male-specific deleterious and female-specific beneficial alleles (van Hooft et al., 2019; van Hooft et al., 2014). As in earlier studies, we refer to the second group as the sexually-antagonistic-effect-associated loci (SAE loci) (van Hooft et al., 2018; van Hooft et al., 2019; van Hooft et al., 2014). The two microsatellite groups are largely independent because only two out of 72 DE-SAE locus pairs (*INRA006* -*ILSTS026* and *TGLA057* -*INRA128* ) are characterized by an interlocus distance <29 Megabase (Mb).

*Analysis of allele-frequency clines*

We conducted three statistical tests for the occurrence of allele-frequency clines, which are explained in detail in the following paragraphs. First, we estimated the significance of regression of average allele frequency across loci against latitude. Second, we tested whether we could attribute the regressions (clines) to selection. Finally, we tested whether selection resulted in increased allele frequencies across the entire latitudinal range. In these tests, we excluded HiP, because here the male-deleterious alleles have earlier been shown to be under strong negative selection (van Hooft et al., 2019). In all tests, we pooled the frequencies of male-deleterious-trait-associated alleles at each locus (Table S1). We refer to the pooled alleles as DE and SAE alleles. Allele-frequency clines were based on plotting the DE and SAE allele frequencies as a function of latitude. The results of the three tests were not meaningfully influenced by sample-size weights, potential errors in allele frequency correction (see next paragraph), or by allele size standardization (Text S2).

In the first test (regression of average allele frequency against latitude), we corrected the average allele frequencies across microsatellites to account for the fact that populations were analysed with different microsatellite subsets using KNP frequencies as a standard for this correction. Specifically, we first calculated the correction factor for the $i$ [th] population in terms of the relative frequencies of the alleles found in population $i$ compared with the full microsatellite set of 17 alleles found in both the northern (NK) and southern (SK) Kruger populations. If we use the notation $f^{NK}$ and $f^{SK}$ to represent the average frequencies of the 17 alleles found in northern and southern KNP and $f_i^{NK}$ and $f_i^{SK}$ to represent the averages in northern and southern KNP over only those alleles found in population $i$ (some subset of the 17 alleles found in KNP), then our corrected average frequency $f_i^{cor}$ for population $i$ , in terms of the observed average frequency $f_i^{obs}$ of the subset of alleles in population $i$ is given by:

$$f_i^{cor} = f_i^{obs} (f^{NK} + f^{SK})/(f_i^{NK} + f_i^{SK}) (1)$$

This correction assumes a similar frequency ratio in all populations. The ratios were estimated for all loci combined, and for DE and SAE loci separately. The frequency correction had only a minor effect as the ratios varied by no more than a factor of 1.1 (range: 0.92-1.09), except for the Caprivi Strip population for which data from only two loci were available.

The large variation in number of genotyped individuals and number of genotyped loci between populations resulted in significant heteroskedasticity in our regression models (modified Breusch-Pagan test, Text S3) (Wooldridge, 2013). Correction for heteroskedasticity is possible by weighing each population by the number of genotyped individuals multiplied by the average number of genotyped loci per individual, because the standard deviations of the error terms are expected to scale linearly with the square-root of 'within-group sample size'. However, we took a slightly different approach by weighing the regressions (of average allele frequency across loci against latitude) by the sum of the square-roots of the number of genotyped individuals per population (i.e., by $\sum_{l=1}^{n} \sqrt{g_l}$, where $g_l$ is the number of genotyped individuals at locus $l$ , with $l = 1,\ldots,17$). This is more appropriate here instead of the square-root of the sums because it gives more weight

to the number of genotyped loci rather than to the number of genotyped individuals. In this way, relatively low weight was given to the Caprivi Strip population despite its large sample size (two genotyped loci, 134 genotyped individuals). Also, we weighted by square-root per locus, because the adjusted $R^2$ values were strongly positively biased by the relatively large sample sizes of the two KNP subpopulations. Besides significance in the Breusch-Pagan test, the effect of heteroskedasticity was also evident from the increased adjusted $R^2$ values relative to the unweighted regressions. The regressions were conducted with SPSS 23.

As a control, we also estimated the allele-frequency clines of the remaining alleles from southern KNP observed [?]15 times that were not associated with male-deleterious traits, again after pooling frequencies of individual alleles. We based these control clines on the alleles that were closest in size (number of short tandem repeats) to the non-pooled male-deleterious-trait associated alleles. Further, where possible, we selected the same number of non-pooled alleles per locus as with the male-deleterious-trait-associated alleles (Table S1). Only at three loci was the number of selected remaining control alleles smaller than the comparison alleles (*BM3517* and *TGLA227* : one instead of two alleles; *BM1824* : two instead of four alleles).

In constructing the allele-frequency clines, we used an updated definition of DE and SAE alleles according to a recent genetic study in HiP (i.e., pooling all alleles with a relatively high frequency in BTB-positive males with low body condition) (van Hooft et al., 2019), rather than the one applied for the allele-frequency clines in KNP (van Hooft et al., 2014). Also with the new definition, a significant cline was observed in KNP for DE and SAE alleles combined and for DE alleles specifically, but not for SAE alleles (DE loci: adjusted$R^2 = 0.23$, $P = 0.0042$; SAE loci: adjusted $R^2 = 0.04$, $P = 0.15$; all loci: adjusted $R^2 = 0.23$, $P = 0.0040$; frequency against latitude, weighted by square-root of number of sampled individuals per herd; Figures S2 and S3). However, the Spearman correlation between average SAE allele frequency and latitude, when weighted by square-root of sample size, was close to significance ($\rho$ = -0.36, $P = 0.054$).

In the second test (attribution of allele-frequency clines to selection), we estimated the significance of the average per-locus Pearson correlation between allele frequency and latitude, with the individual per-locus correlations weighted by the square-root of the number of genotypes per population. In the third test (increased allele frequencies across whole latitudinal range), we estimated whether average frequencies per locus were significantly higher for the DE and SAE alleles than for the control alleles, with the per-locus frequencies weighted by the square-root of the number of genotypes per population. In both tests, randomized pooled frequencies were obtained by replacing the non-pooled male-deleterious-trait-associated alleles (as mentioned before, DE and SAE alleles consisted of pools of male-deleterious-trait-associated alleles) with a random selection of non-pooled alleles (73 alleles observed [?]15 times in southern KNP, consisting of 33 male-deleterious-trait-associated alleles and 40 remaining alleles). We estimated significance as the probability that a random selection of non-pooled alleles resulted in a stronger average Pearson correlation per locus (Test 2) or a higher average frequency per locus (Test 3) than the male-deleterious-trait-associated alleles. We applied 100,000 randomizations implemented using Excel 2016.

*Analysis of linkage disequilibrium*

Four microsatellite pairs showed significant LD in both KNP and HiP (van Hooft et al., 2019; van Hooft et al., 2014). 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), *CSSM01* 9-*BM3205* (5.4 Mb), *BM1824* -*BM3205* (5.9 Mb) (Ihara et al., 2004). The fourth pair consists of microsatellites *ILSTS026* and *INRA006* , which are probably ˜18 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 (Text S4) (Gallagher & Womack, 1992; Goudet, 1995, 2001; Ihara et al., 2004; Lane-deGraaf et al., 2015; Liu & Muse, 2005; Weng, Saatchi, Schnabel, Taylor, & Garrick, 2014). 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) and *TGLA057* -*INRA128* (28.3 Mb, chromosome 1) (Ihara et al., 2004). The chromosomal distances at the aforementioned locus pairs

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 (Mouresan et al., 2019; Weng et al., 2014).

In this study, we conducted two statistical tests for LD, with LD estimated per population per locus pair, assuming two alleles per locus: the DE allele or SAE allele on the one hand and the wild-type allele (the pool of all original microsatellite alleles not associated with a male-deleterious trait, including the rare alleles) on the other. In these tests, we did not include the *BM1824 -BM3205* pair, because this specific combination of microsatellites was only genotyped in KNP and HiP and because their LD would not be independent of the *BM1824 -CSMM019* and *CSMM019 -BM3205* locus pairs. Thus, the two LD tests were conducted with the following five locus pairs: *BM1824 -CSMM019* (0.5 Mb), *CSSM01 9-BM3205* (5.4 Mb), *ILSTS026 -INRA006* (~18 Mb), *INRA006 -TGLA263* (27.5 Mb) and *TGLA057 -INRA128* (28.3 Mb).

We measured LD by the Pearson correlation coefficient for a 2x2 table, $r_{LD}$, for bi-allelic locus pairs. When the linkage phase is known, $r_{LD}$ can be derived from the observed and expected haplotype frequencies ($\chi^2$ value, four possible haplotypes) and the total number of gametes or chromosomes analysed ($k$), using Equation 2 below (Charlesworth & Charlesworth, 2010). A negative sign is added to $r_{LD}$ when most pairs of focal alleles are in repulsion. However, when the linkage phase is unknown, as in our case, $r_{LD}$ can be derived from maximum-likelihood estimates of the haplotype frequencies (Schaid, 2004; Weir, 1990). In our case, positive $r_{LD}$ values indicate that most pairs of male-deleterious alleles are in coupling (i.e., DE-DE, SAE-SAE or DE-SAE).

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

We estimated $r_{LD}$ with Genalex, following Weir's method (Weir, 1990). Genalex provides three estimates of $r_{LD}$: 1) $r_{LD}$ based on estimated frequencies of double heterozygotes in coupling assuming Hardy-Weinberg equilibrium, 2) $r_{LD}$ based on estimated frequencies of double heterozygotes in repulsion assuming Hardy-Weinberg equilibrium, 3) $r_{LD}$ based on estimated frequencies of both types of double heterozygotes without assuming Hardy-Weinberg equilibrium. The first two estimates should be identical, but the maximum-likelihood procedure may result in deviations, especially with small sample sizes. We used the $r_{LD}$ estimates that assume HWE (i.e., estimates 1 and 2). We only included cases where estimates 1 and 2 were identical and of the same sign as estimate 3 (67 out of 78 cases), except for two cases where estimates 1 and 2 were of opposite sign but very close to zero ($|r_{LD}| = 0.0002$) while estimate 3 was zero. For these two cases we assumed $r_{LD}=0$.

We pooled the samples from the neighbouring populations in Mana Pools NP-Nyakasanga (~55 km distance) and in Limpopo NP-Manguana (~200 km distance) to increase sample size, after checking that in the latter case this pooling, considering the large geographic distance, did not bias estimates of $r_{LD}$ (paired sample $t$-test on 30 locus pairs, comparing weighted average $r_{LD}$ of the two separate populations with $r_{LD}$ of the pooled population: pooled population: average $r_{LD}$ = -0.010, two separation populations: average $r_{LD}$ = 0.002; $P = 0.52$; Pearson $r = 0.91$). It was not possible to estimate $r_{LD}$ for all populations and locus pairs because of small sample sizes. However, in some cases we were able to get an estimate of $r_{LD}$ by a leave-one-out approach (*CSMM019 -BM3205* in Queen Elizabeth NP (both sectors) and Amboseli NP, *BM1824 -CSMM019* in Limpopo NP-Manguana, *INRA006 -TGLA263* in Mana Pools NP-Nyakasanga). Hereby, we calculated average $r_{LD}$ for all possible population samples with one individual excluded, with multiple estimates per population always being of the same sign. We excluded HiP because of earlier observed negative selection (van Hooft et al., 2019), the two central African populations because of small sample size, and the admixed population from Save Valley Conservancy because admixing may bias LD estimates (Pfaff et al., 2001).

In the first LD test, we hypothesized that pairs of male-deleterious-trait-associated alleles (i.e.; DE-DE, SAE-SAE and DE-SAE allele pairs) are coupled more often than expected with random association, resulting in positive $r_{LD}$ values. We conducted a $\chi^2$ test to determine whether the total fraction of locus pairs (across loci and across populations) with a positive $r_{LD}$ value significantly deviated from 0.5, considering that positive and negative $r_{LD}$ values are equally likely in the absence of selection (Slatkin, 2008). LD estimates

of the two closest linked locus pairs, *BM1824* - *CSMM019* and *CSMM019* - *BM3205* , for populations that are part of a larger metapopulation may not be independent of each other. To err on the conservative side, we used the average $r_{LD}$ value per metapopulation across both locus pairs (KNP was the only metapopulation in which both locus pairs were analysed). With respect to these locus pairs, the following four groups of populations were considered to be part a metapopulation: 1) northern KNP - southern KNP, 2) Malilangwe Wildlife Res. - Gonarezhou NP - Sengwe Safari Area - Crooks Corner, 3) Victoria Falls NP - Chobe NP - Okavango Delta, 4) Amboseli NP - Tsavo NP, 5) Queen Elizabeth NP Mw - Queen Elizabeth NP Is. 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 (~20-31% per generation) likely exceeds the migration rate between nearby populations.

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 with genotypes randomized per locus per population, and 2) locus pairs with interlocus distance >45 Mb (seven locus pairs) or that occur on different chromosomes (116 locus pairs). Both fractions are expected to be close to 0.5 (free recombination). In the first control we conducted 30 randomizations of the total dataset, resulting in 1541 randomized $r_{LD}$ values. The $r_{LD}$ values in the second control were estimated with GDA, instead of Genalex, because it can better handle large data sets with missing data (Labate, 2000). At the five closely linked locus pairs, Genalex and GDA always provided identical estimates.

In the second LD test, we estimated the significance of regression of $r_{LD}$ for each population against the natural logarithm of the variable $l$ =latitude (67 data points) over the range [-24.75°, 3.85°], i.e., we fitted the parameters $a$ and $b$ in the equation:

$$r_{LD} = a \ln(l - l_0) + b \quad (3)$$

using a general linear mixed model (GLMM) approach, specifically the 'lme4' (version 1.1.21), 'lmerTest' (version 3.1.0) and 'performance' (version 0.4.4.1) packages in R (Bates, Machler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2017). We applied log-transformation to account for the obvious non-linearity in the data, using $l_0$ = -27. Locus pair (five levels), population (25 levels, including three singletons) and metapopulation (57 levels, including 52 singletons) were included as random factors. GLMMs were not weighted by sample size as this only had a minor effect on parameter estimates (< 0.5% difference). Goodness-of-fit was estimated with marginal $R^2$; the proportion of variability explained by the fixed effects (Nakagawa & Schielzeth, 2013).

## Results

### *Allele-frequency clines*

We observed a highly significant allele-frequency cline between 24.9 °S and 4.8 °N for both the DE and SAE alleles ($P$ <0.0001; DE loci: adjusted $R^2$ = 0.78, SAE loci: adjusted $R^2$ = 0.77; Figure 2, Table S2). The two slopes were similar, differing by a factor of only 1.2. Predicted DE allele frequencies decreased from 0.75 (95% CI: 0.72-0.77) at 24.9 °S to 0.46 (95% CI: 0.42-0.51) at 4.8 °N (a decrease of 0.0095 per degree latitude). Predicted SAE allele frequencies decreased from 0.40 (95% CI: 0.37-0.43) at 24.9 °S to 0.17 (95% CI: 0.13-0.20) at 4.8 °N (a decrease of 0.0079 per degree latitude). When considered together, the clines were significantly stronger than clines expected under a random selection of alleles (DE loci: $P$ = 0.081, SAE loci: $P$ = 0.043, all loci: $P$ = 0.013), which indicates that they cannot be explained by genetic drift. The cline based on both the DE and SAE alleles explained as much as 87% of the variation (adjusted $R^2$; Figure 3, Table S2). Predicted allele frequencies decreased from 0.57 (95% CI: 0.55-0.59) at 24.9 °S to 0.29 (95% CI: 0.27-0.32) at 4.8 °N (a decrease of 0.0091 per degree latitude). No clines were observed for the remaining alleles, which were not associated with male-deleterious traits (negative adjusted $R^2$ values). Frequencies of DE alleles, but not SAE alleles, in East and central Africa were significantly higher than pooled frequencies based on a random selection of alleles (six DE alleles: multiply the 1-sided $P$ = 0.013 by 2, nine SAE alleles: 1-sided $P$ = 0.70).

The populations from HiP and Nairobi NP deviated from the cline, showing relatively low DE and SAE allele frequencies that were close to the frequencies of the remaining microsatellite alleles. In HiP 18 out of 20 and in Nairobi NP 16 out of 20 of the non-pooled male-deleterious-trait-associated alleles at the microsatellites analysed in both populations showed relatively low frequencies compared to nearby populations (HiP: compared with northern and southern KNP, Nairobi NP: compared with other East African populations; Table S3). Low frequencies were observed for most non-pooled alleles, which is not expected under genetic drift (Wilcoxon signed rank exact test; HiP:$P = 0.0010$; Nairobi NP: $P = 0.0083$, Holm-Šidák correction based on 12 possible focal populations: $P = 0.095$, Table S3). In HiP, low frequencies have earlier been attributed to negative selection (van Hooft et al., 2019).

*LD cline*

The total fraction of locus pairs across populations and linked loci with positive$r_{LD}$ was 0.65 (95% CI: 0.52-0.77), which is higher than expected under random association (LD test 1: $N = 55$, exact $P = 0.030$; $r_{LD}$ values for all locus pairs and populations are presented in Table S4). 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 [?]18 Mb (three locus pairs: 0.69, $N = 42$, exact $P = 0.020$). Surprisingly, the fraction with positive$r_{LD}$ was also relatively high in East Africa (four locus pairs: 0.69, $N = 26$, exact $P = 0.076$). 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: 0.501; 95% CI: 0.476, 0.526; $N = 1541$, $\chi^2 = 0.006$, $P = 0.94$; remaining loci: 0.484; 95% CI: 0.457, 0.512; $N = 1255$, $\chi^2 = 1.21$, $P = 0.27$; in both cases same subset of populations as with the linked loci, sample size [?]6, $|r_{LD}| > 0.001$).

There was a non-significant logarithmic relationship between$r_{LD}$ and latitude (LD test 2: Eq. 3; $a = -0.055$, $b = 0.189$, $P = 0.26$, marginal$R^2 = 0.032$; Figure 4, Table S5). When considered together, the two LD tests were significant ($Z$-transform test: $P = 0.020$). The decreasing trend in$r_{LD}$ with latitude could largely be attributed to the two closest linked locus pairs; *BM1824* -*CSMM019* and*CSMM019* -*BM3205* (Eq. 3; $a = -0.214$, $b = 0.595$, $P = 0.014$, marginal $R^2 = 0.36$; other three locus pairs: $a = -0.0006$, $b = 0.056$, $P = 0.99$; Figure 4, Table S5). Predicted $r_{LD}$ decreased from 0.45 at 25 °S to -0.13 at 4 °N. Similarly to the allele-frequency clines, the LD cline did not extent beyond 25 °S, with a relatively strong drop of $r_{LD}$ in HiP, being non-significant here despite high sample size ($r_{LD} = -0.055$, $N = 418$, $P = 0.26$), relative to KNP ($r_{LD} = 0.32-0.34$, $N = 155-328$, $P < 0.00002$). Significant positive$r_{LD}$ after Bonferroni correction was observed for both locus pairs in the four most southern populations; northern KNP, southern KNP, Limpopo NP and Manguana ($P < 0.00012$; Bonferroni-corrected $\alpha = 0.00073$; Table S4). The two highest$r_{LD}$ values at any locus pair were observed in the same metapopulation; Limpopo NP-Manguana (*BM1824* -*CSMM019* :$r_{LD} = 0.93$, $P = 0.00012$;*INRA006* -*TGLA263* : $r_{LD} = 0.68$,$P = 0.017$; Table S4).

**Discussion**

We observed a continent-wide allele-frequency cline with decreasing frequencies in a northerly direction for both the DE and SAE loci; two groups of microsatellites that are largely independent of one another (interlocus distance <29 Mb at only two out of 72 DE-SAE locus pairs). We could attribute these clines to selection (DE loci: $P = 0.083$, SAE loci: $P = 0.043$, all loci: $P = 0.013$). Further, we observed an LD cline at the two most closely linked locus pairs ($P = 0.014$, other three physically linked locus pairs:$P = 0.99$), with decreasing interlocus associations between male-deleterious-trait-associated alleles in a northerly direction. Most of the interlocus associations at the five physically linked locus pairs were positive ($P = 0.030$). Continent-wide positive LD in combination with an LD cline (combined $P = 0.020$) is indicative of genome-wide selection resulting in increased frequencies of deleterious-allele haplotypes (relative to linkage equilibrium). Relatively high DE allele frequencies and near-significant positive LD in East Africa indicate that the male-deleterious alleles are active across the whole latitudinal range. Apparently, there is continent-wide and genome-wide selection for male-deleterious traits with selection strength decreasing from south to north. The only exceptions are the populations in HiP and Nairobi NP, where male-deleterious alleles appear to be under negative selection. According to Figure 2, these two populations seem to have lost many (if not

8

most in Nairobi NP) of their male-deleterious alleles. The selection pressures in East and southern Africa must be high to prevent destruction of the allele-frequency clines and haplotypes by LD decay. These results provide strong support for earlier reported observations of genome-wide selection in KNP and HiP (van Hooft et al., 2018; van Hooft et al., 2019; van Hooft et al., 2014).

Common explanations for allele-frequency clines due to selection assume the advance of a favourable mutation (in our case male-deleterious alleles), an environmental gradient, or a combination of the two (Charlesworth & Charlesworth, 2010; de Jong, Collins, Beldade, Brakefield, & Zwaan, 2013). 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, particularly those also deleterious to females (linked to the DE alleles), which normally are under negative selection (Henn, Botigue, Bustamante, Clark, & Gravel, 2015). 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 argue 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, although the male-deleterious alleles and gene-drive system have been shown to interact with diseases, such as BTB, and other environmental factors, such as droughts (van Hooft et al., 2018; van Hooft et al., 2019; van Hooft et al., 2014). 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 DE and SAE allele-frequency clines co-occurred with an allele-frequency cline of a Y-chromosomal haplotype linked to a Y-suppressor gene (Y haplotype 557) (van Hooft et al., 2014). According to this hypothesis, the gene-drive system originated in southern Africa and subsequently spread north, thereby forming a selection gradient. Relatively low frequencies of Y haplotype 557 in other southern African populations (except HiP) support this hypothesis (frequency in KNP: 0.24, 95% CI: 0.19, 0.30; average frequency in four other populations: 0.07, 95% CI: 0.02, 0.18 (Smitz et al., 2014; van Hooft et al., 2010)). The male-deleterious alleles (which occur on the autosomes) probably originated at various localities throughout Africa, considering that mutations leading to such alleles can in principle occur in any population, after which they formed an allele-frequency cline in response to the aforementioned selection gradient (formed by the gene-drive system, which occurs on the sex chromosomes).

When the gene-drive system is incomplete, selection may become negative, resulting in relatively low DE and SAE allele frequencies. This has earlier has been argued for HiP and may also apply for Nairobi NP (van Hooft et al., 2019). In both cases, incompleteness of the gene drive system may be due to strong genetic drift in these small and isolated populations (HiP: [?] 75 individuals between 1895 and 1930, Nairobi NP: current census size ~150; these small sizes probably contributed to the relatively large pairwise $F_{ST}$ values with other populations, Figure S1) (Heller, Okello, & Siegismund, 2010; van Hooft et al., 2019).

Our explanation for the allele-frequency clines implies that frequencies of male-deleterious alleles and haplotypes 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 (van Hooft et al., 2018; van Hooft et al., 2019; van Hooft et al., 2014)). We previously hypothesized that epigenetic suppression of male-deleterious alleles results in balancing selection of Y-chromosomal genes in the gene-drive system (van Hooft et al., 2018). Taking cognizance of an association between epigenetic suppression and low pre-birth rainfall, we argued that males with many active

9

male-deleterious alleles should tend to produce offspring with suppressed male-deleterious alleles and *vice versa* (males with few active alleles producing offspring with many active alleles). We further hypothesize here that the same process also results in balancing selection of male-deleterious alleles and haplotypes.

Positive LD at linked loci indicates increased frequencies of deleterious-allele haplotypes relative to linkage equilibrium. We detected long-distance LD at two pairs of loci, both ~28 Mb apart; these pairs occur on chromosomes 1 and 3 in cattle. Long-distance LD appears to occur genome-wide considering that significant LD has also been observed in KNP for the *INFNG - GLYCAM1* locus pair, which has an interlocus distance of 18 Mb and occurs on chromosome 5 in cattle ($D$ ' = 0.28, i.e. LD is at 28% of its maximum possible value) (Ihara et al., 2004; Lane-deGraaf et al., 2015). This is a relatively large distance, especially considering that high haploid diversity indicates a large effective population size for the KNP buffalo (34 mitochondrial D-loop haplotypes with $H$ (gene diversity) = 0.94, 15 Y-chromosomal haplotypes with $H$ = 0.74; census size ~37,000 in 2010) (Greyling, 2007; van Hooft et al., 2018; van Hooft et al., 2007). According to population genetic theory, a large effective population size limits long-distance LD because of increased LD decay with chromosomal distance (Slatkin, 2008). 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, in Iberian wild boar (*Sus scrofa* ) to a distance of ~0.5 Mb, and in gray wolf (*Canus lupus* ) and coyote (*Canus latrans* ) to a distance of ~5 Mb, even in small or bottlenecked populations (except for the wolf population of Isle Royale, which consisted of just 10-30 individuals) (De Manuel et al., 2016; Gray et al., 2009; Herrero-Medrano et al., 2013; Laurie et al., 2007; Munoz et al., 2019). 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 17%-28% of that in KNP buffalo, despite their small size of less than 200 individuals each (4.6-7.5 Mb vs. 26.4 Mb, Text S4) (Miller, Poissant, Kijas, Coltman, & Int Sheep Genomics, 2011; Miller, Poissant, Malenfant, Hogg, & Coltman, 2015).

LD between distant loci in large outbreeding buffalo populations despite fast LD decay due to recombination (~20-31% per generation), indicates strong selection pressures. However, simulation studies 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) (Baird, 1995; Hastings, 1984). 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 (van Hooft et al., 2018), may have resulted in admixture LD between distant loci, especially when selection is strong (Úbeda et al., 2011).

A high number of high-frequency deleterious alleles of seemingly large effect (van Hooft et al., 2018; van Hooft et al., 2014), with many co-occurring in haplotypes (relative to what one would expect if no linkage existed among these alleles), indicates a high genetic load. However, most populations of African buffalo seem to be stable after their recovery from the rinderpest pandemic of 1889-1895, which decimated buffalo populations across the whole of Africa (Bengis, Kock, & Fischer, 2002; van Hooft et al., 2000). At face value, this seems to support the view, advocated by some population geneticists, that genetic load plays a smaller role than one might expect in ecology (Agrawal & Whitlock, 2012).

We suggest the following five reasons, among others that surely exist, as non-mutually exclusive explanations why buffalo populations are still stable.

1. Male-deleterious alleles are epigenetically suppressed in a large fraction of animals (van Hooft et al., 2018; van Hooft et al., 2019).
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 (i.e., associated with SAE microsatellite alleles) (van Hooft et al., 2014).
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 DE and SAE alleles, than in northern KNP, but only significantly so at the end of the dry season (Caron, Cross, & Du Toit, 2003). It may even be that male-deleterious alleles are mainly expressed during times of high environmental stress considering that most genetic samples used in the studies on KNP and HiP were collected during dry seasons (van Hooft et al., 2018; van Hooft et al., 2019; van Hooft et al., 2014), and then mainly in animals born during stressful periods because these cohorts do not experience epigenetic suppression (van Hooft et al., 2018; van Hooft et al., 2019).

4. Selection is mostly soft, where selective death due to male-deleterious alleles would otherwise be replaced by nonselective death due to environmental and ecological conditions, e.g. droughts, diseases and intraspecific resource competition (Agrawal & Whitlock, 2012).

5. Interspecific competition is limited, thereby minimizing the ecological effects of genetic load (Agrawal & Whitlock, 2012). A recent review has failed to find conclusive evidence for interspecific competition between African buffalo and other large mammals (Prins, 2016), which may be due to ecological separation (Traill, 2004).

Our analyses and results reveal 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). Since most populations appear to be stable, this indicates that, under specific circumstances, natural populations of mammals can withstand a high genetic load. Nevertheless, we expect that a high genetic load makes many buffalo populations vulnerable to environmental stresses 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 (Michel & Bengis, 2012), which may well have been augmented by a high genetic load, particularly in southern Africa, our results have relevance to livestock husbandry in areas were cattle graze in close proximity to buffalo herds.

**Acknowledgements**

**References**

Agrawal, A. F., & Whitlock, M. C. (2012). Mutation Load: The Fitness of Individuals in Populations Where Deleterious Alleles Are Abundant. *Annual Review of Ecology, Evolution, and Systematics, Vol 43, 43* , 115-+. doi:10.1146/annurev-ecolsys-110411-160257

Baird, S. J. E. (1995). A simulation study of multilocus clines. *Evolution, 49* (6), 1038-1045. doi:10.2307/2410429

Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67* (1), 1-48.

Bengis, R. G., Kock, R. A., & Fischer, J. (2002). Infectious animal diseases: the wildlife/livestock interface. *Revue Scientifique Et Technique De L Office International Des Epizooties, 21* (1), 53-65.

Caron, A., Cross, P. C., & Du Toit, J. T. (2003). Ecological implications of bovine tuberculosis in African buffalo herds. *Ecological Applications, 13* (5), 1338-1345.

Charlesworth, B., & Charlesworth, D. (2010). *Elements of evolutionary genetics* . Greenwood Village: Roberts and Company Publishers.

de Jong, M. A., Collins, S., Beldade, P., Brakefield, P. M., & Zwaan, B. J. (2013). Footprints of selection in wild populations of *Bicyclus anynana* along a latitudinal cline. *Molecular Ecology, 22* (2), 341-353. doi:10.1111/mec.12114

De Manuel, M., Kuhlwilm, M., Frandsen, P., Sousa, V. C., Desai, T., Prado-Martinez, J., . . . Marques-Bonet,

T. (2016). Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science, 354* (6311), 477-481. doi:10.1126/science.aag2602

Epps, C. W., Castillo, J. A., Schmidt-Kuentzel, A., du Preez, P., Stuart-Hill, G., Jago, M., & Naidoo, R. (2013). Contrasting Historical and Recent Gene Flow among African Buffalo Herds in the Caprivi Strip of Namibia. *Journal of Heredity, 104* (2), 172-181. doi:10.1093/jhered/ess142

Ernest, E. M., Haanes, H., Bitanyi, S., Fyumagwa, R. D., Msoffe, P. L., Bjornstad, G., & Roed, K. H. (2012). Influence of habitat fragmentation on the genetic structure of large mammals: evidence for increased structuring of African buffalo (*Syncerus caffer* ) within the Serengeti ecosystem. *Conservation Genetics, 13* (2), 381-391. doi:10.1007/s10592-011-0291-0

Gallagher, D. S., & Womack, J. E. (1992). Chromosome conservation in the bovidae. *Journal of Heredity, 83* (4), 287-298.

Goudet, J. (1995). FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity, 86* (6), 485-486.

Goudet, J. (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Updated from Goudet (1995).

Gray, M. M., Granka, J. M., Bustamante, C. D., Sutter, N. B., Boyko, A. R., Zhu, L., . . . Wayne, R. K. (2009). Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics, 181* (4), 1493-1505. doi:10.1534/genetics.108.098830

Greyling, B. J. (2007). *Genetic variation, structure and dispersal among Cape buffalo populations from the Hluhluwe-Imfolozi and Kruger National Parks of South Africa.* (PhD PhD thesis), University of Pretoria, Pretoria. (etd-07152008-090238)

Greyling, B. J., Kryger, P., du Plessis, S., van Hooft, W. F., van Helden, P., Getz, W. M., & Bastos, A. D. S. (2008). Development of a high-throughput microsatellite typing approach for forensic and population genetic analysis of wild and domestic African Bovini.*African Journal of Biotechnology, 7* (5), 655-660.

Hastings, A. (1984). Linkage disequilibrium, selection and recombination at 3 loci. *Genetics, 106* (1), 153-164.

Heller, R., Lorenzen, E. D., Okello, J. B. A., Masembe, C., & Siegismund, H. R. (2008). Mid-Holocene decline in African buffalos inferred from Bayesian coalescent-based analyses of microsatellites and mitochondrial DNA. *Molecular Ecology, 17* (22), 4845-4858. doi:10.1111/j.1365-294X.2008.03961.x

Heller, R., Okello, J. B. A., & Siegismund, H. (2010). Can small wildlife conservancies maintain genetically stable populations of large mammals? Evidence for increased genetic drift in geographically restricted populations of Cape buffalo in East Africa. *Molecular Ecology, 19* (7), 1324-1334. doi:10.1111/j.1365-294X.2010.04589.x

Henn, B. M., Botigue, L. R., Bustamante, C. D., Clark, A. G., & Gravel, S. (2015). Estimating the mutation load in human genomes. *Nature Reviews Genetics, 16* (6), 333-343. doi:10.1038/nrg3931

Herrero-Medrano, J. M., Megens, H. J., Groenen, M. A. M., Ramis, G., Bosse, M., Pérez-Enciso, M., & Crooijmans, R. P. M. A. (2013). Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. *Bmc Genetics, 14* . doi:10.1186/1471-2156-14-106

Ihara, N., Takasuga, A., Mizoshita, K., Takeda, H., Sugimoto, M., Mizoguchi, Y., . . . Sugimoto, Y. (2004). A comprehensive genetic map of the cattle genome based on 3802 microsatellites. *Genome Research, 14* (10A), 1987-1998. doi:10.1101/gr.2741704

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82* (13), 1-26.

Labate, J. A. (2000). Software for population genetic analyses of molecular marker data. *Crop Science, 40* (6), 1521-1528. doi:10.2135/cropsci2000.4061521x

LaHood, E. S., Moran, P., Olsen, J., Grant, W. S., & Park, L. K. (2002). Microsatellite allele ladders in two species of Pacific salmon: preparation and field-test results. *Molecular Ecology Notes, 2* (2), 187-190. doi:10.1046/j.1471-8286.2002.00174.x

Lane-deGraaf, K. E., Amish, S. J., Gardipee, F., Jolles, A., Luikart, G., & Ezenwa, V. O. (2015). Signatures of natural and unnatural selection: evidence from an immune system gene in African buffalo. *Conservation Genetics, 16* (2), 289-300. doi:10.1007/s10592-014-0658-0

Laurie, C. C., Nickerson, D. A., Anderson, A. D., Weir, B. S., Livingston, R. J., Dean, M. D., . . . Nachman, M. W. (2007). Linkage disequilibrium in wild mice. *Plos Genetics, 3* (8), 1487-1495. doi:10.1371/journal.pgen.0030144

Liu, K., & Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics, 21* (9), 2128-2129.

May, R. M., Endler, J. A., & McMurtrie, R. E. (1975). Gene frequency clines in presence of selection opposed by gene flow. *American Naturalist, 109* (970), 659-676. doi:10.1086/283036

Michel, A. L., & Bengis, R. G. (2012). The African buffalo: a villain for inter-species spread of infectious diseases in southern Africa. *The Onderstepoort journal of veterinary research, 79* (2), 453-453. doi:10.4102/ojvr.v79i2.453

Miller, J. M., Poissant, J., Kijas, J. W., Coltman, D. W., & Int Sheep Genomics, C. (2011). A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources, 11* (2), 314-322. doi:10.1111/j.1755-0998.2010.02918.x

Miller, J. M., Poissant, J., Malenfant, R. M., Hogg, J. T., & Coltman, D. W. (2015). Temporal dynamics of linkage disequilibrium in two populations of bighorn sheep. *Ecology and Evolution, 5* (16), 3401-3412. doi:10.1002/ece3.1612

Mouresan, E. F., Gonzalez-Rodriguez, A., Canas-Alvarez, J. J., Munilla, S., Altarriba, J., Diaz, C., . . . Varona, L. (2019). Mapping Recombination Rate on the Autosomal Chromosomes Based on the Persistency of Linkage Disequilibrium Phase Among Autochthonous Beef Cattle Populations in Spain. *Frontiers in Genetics, 10* , 12. doi:10.3389/fgene.2019.01170

Munoz, M., Bozzi, R., Garcia-Casco, J., Nunez, Y., Ribani, A., Franci, O., . . . Ovilo, C. (2019). Genomic diversity, linkage disequilibrium and selection signatures in European local pig breeds assessed with a high density SNP chip. *Scientific Reports, 9* , 14. doi:10.1038/s41598-019-49830-6

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4* (2), 133-142. doi:10.1111/j.2041-210x.2012.00261.x

Peakall, R., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes, 6* (1), 288-295. doi:10.1111/j.1471-8286.2005.01155.x

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics, 28* (19), 2537-2539. doi:10.1093/bioinformatics/bts460

Pfaff, C. L., Parra, E. J., Bonilla, C., Hiester, K., McKeigue, P. M., Kamboh, M. I., . . . Shriver, M. D. (2001). Population structure in admixed populations: Effect of admixture dynamics on the pattern of linkage disequilibrium. *American Journal of Human Genetics, 68* (1), 198-207. doi:10.1086/316935

Prins, H. H. T. (2016). Interspecific resource competition in antelopes: search for evidence. In J. Bro-Jørgensen & D. P. Mallon (Eds.), *Antelope conservation in the 21st century: from diagnosis to action.* Oxford: Wiley-Blackwell.

Schaid, D. J. (2004). Linkage disequilibrium testing when linkage phase is unknown. *Genetics, 166* (1), 505-512. doi:10.1534/genetics.166.1.505

Slatkin, M. (1973). Gene flow and selection in a cline. *Genetics, 75* (4), 733-756.

Slatkin, M. (2008). Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics, 9* (6), 477-485. doi:10.1038/nrg2361

Smitz, N., Cornelis, D., Chardonnet, P., Caron, A., de Garine-Wichatitsky, M., Jori, F., . . . Michaux, J. (2014). Genetic structure of fragmented southern populations of African Cape buffalo (*Syncerus caffer caffer* ). *Bmc Evolutionary Biology, 14* . doi:203 10.1186/s12862-014-0203-2

Traill, L. W. (2004). Seasonal utilization of habitat by large grazing herbivores in semi-arid Zimbabwe. *South African Journal of Wildlife Research, 34* (1), 13-24.

Úbeda, F., Haig, D., & Patten, M. M. (2011). Stable linkage disequilibrium owing to sexual antagonism. *Proceedings of the Royal Society B: Biological Sciences, 278* (1707), 855-862. doi:10.1098/rspb.2010.1201

van Hooft, P., Dougherty, E. R., Getz, W. M., Greyling, B. J., Zwaan, B. J., & Bastos, A. D. S. (2018). Genetic responsiveness of African buffalo to environmental stressors: a role for epigenetics in balancing autosomal and sex chromosome interactions? *PLoS ONE, 13* (2), e0191481.

van Hooft, P., Getz, W. M., Greyling, B. J., & Bastos, A. D. S. (2019). A natural gene drive system influences bovine tuberculosis susceptibility in African buffalo: possible implications for disease management. *PLoS ONE, 14* (9), e0221168.

van Hooft, P., Greyling, B. J., Getz, W. M., Heller, R., Røed, K. H., & Bastos, A. D. S. (2020). *Microsatellite data African buffalo populations* [Individual genotypes]. Dryad. Retrieved from: http://doi.org

van Hooft, P., Greyling, B. J., Getz, W. M., van Helden, P. D., Zwaan, B. J., & Bastos, A. D. S. (2014). 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, 9* (11), e111778.

van Hooft, P., Greyling, B. J., Prins, H. H. T., Getz, W. M., Jolles, A. E., & Bastos, A. D. S. (2007). Selection at the Y chromosome of the African buffalo driven by rainfall. *PLoS ONE, 2* (10), 10.1371/journal.pone.0001086. doi:10.1371/journal.pone.0001086

van Hooft, P., Prins, H. H. T., Getz, W. M., Jolles, A. E., van Wieren, S. E., Greyling, B. J., . . . Bastos, A. D. S. (2010). Rainfall-driven sex-ratio genes in African buffalo suggested by correlations between Y-chromosomal haplotype frequencies and foetal sex ratio. *BMC Evol Biol, 10* , 10.1186/1471-2148-1110-1106. doi:10.1186/1471-2148-10-106

van Hooft, W. F., Groen, A. F., & Prins, H. H. T. (2000). Microsatellite analysis of genetic diversity in African buffalo (*Syncerus caffer* ) populations throughout Africa. *Molecular Ecology, 9* (12), 2017-2025.

Weir, B. S. (1990). *Genetic Data Analysis* . Sunderland: Sinauer Associates, Inc.

Weng, Z. Q., Saatchi, M., Schnabel, R. D., Taylor, J. F., & Garrick, D. J. (2014). Recombination locations and rates in beef cattle assessed from parent-offspring pairs. *Genetics Selection Evolution, 46* , 14. doi:10.1186/1297-9686-46-34

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (Fifth ed.): South-Western, Cengage Learning.

**Data Accessibility Statement**

The raw genotype data from microsatellite sets A, B, C and E will be made available to the Dryad Digital Repository on acceptance of the manuscript.

**Author Contributions**

14

P. van Hooft analysed the data and wrote the paper. All authors discussed and commented on the paper.

**Tables**

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

| Population no. | Population | Lat. | Lon. | µsat 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+ | -0.06 | 30.00 | C | 37 |
| 7 | Queen Elizabeth NP Is+ | -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++ | -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§ | -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¶ | -23.15 | 31.30 | A, B, D | 138, 22, 26 |
| 33 | Southern KNP¶ | -24.75 | 31.70 | A, B | 321, 16 |
| 34 | HiP | -28.22 | 31.95 | A, D | 401, 20 |

+: Mweya sector and Ishasha sector, ++§: Including Maswa GR, §: 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) (van Hooft et al., 2000). ¶: north and south Olifants River.

Table 2: Overview of microsatellite sets used

| Set | DE µsats | SAE µsats | Reference |
|---|---|---|---|
| A | 1-8 | 9-17 | (Greyling et al., 2008; van Hooft et al., 2019; van Hooft et al., 2014) |

15

| Set | DE µsats | SAE µsats | Reference |
|-----|----------|-----------|-----------|
| B | 1, 2, 5, 6, 8 | 9-12, 14-16 | (van Hooft et al., 2000) |
| C | 1, 2, 6 | 9-12, 14-16 | (Heller et al., 2008) |
| D | 2-8 | 9, 10, 13, 15, 17 | (Smitz et al., 2014) |
| E | 3 | 13, 17 | (Ernest et al., 2012) |
| F | 1 | 11 | (Epps et al., 2013) |

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*

**Figures legends**

Figure 1: Map of Africa representing the 34 sampling localities

Locality numbers are explained in Table 1.

Figure 2: Frequency clines of DE and SAE alleles

Scatter plot and regression lines with 95% confidence interval. Blue data points: frequency of DE alleles averaged across loci (adjusted$R^2$ = 0.78, excluding HiP), orange data points: frequency of SAE alleles averaged across loci (adjusted$R^2$ =0.77, excluding HiP). Light blue and light orange data points (HiP at 28.2 °S and Nairobi NP at 1.4 °S) indicate negative outliers.
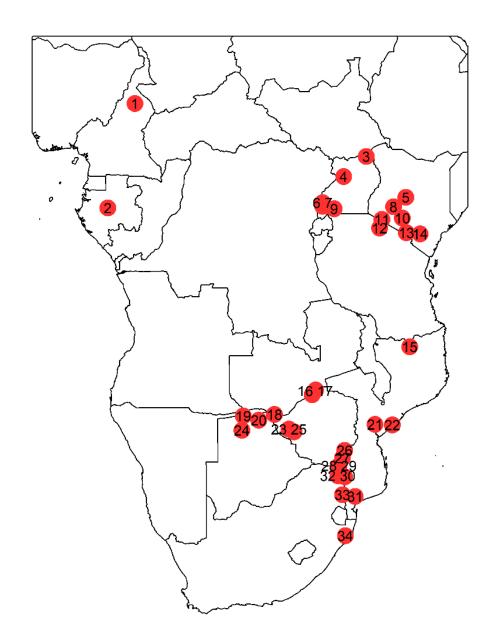
Figure 3: Combined frequency cline of DE and SAE alleles

Scatter plot and regression lines with 95% confidence interval. Orange data points: frequency of DE and SAE alleles averaged across loci (adjusted $R^2$ = 0.87, excluding HiP). Light orange data points (HiP at 28.2 °S and Nairobi NP at 1.4 °S) indicate negative outliers. Grey data points: frequency of remaining microsatellite alleles not associated with male-deleterious traits averaged across loci (adjusted $R^2$ = -0.01, excluding HiP).
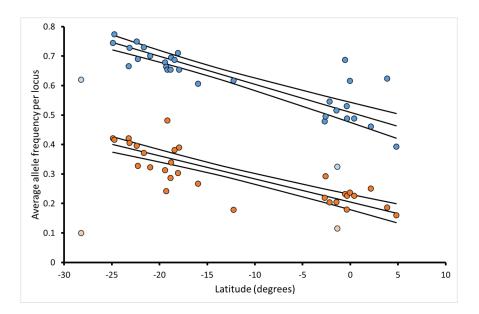
Figure 4: LD cline at the two most closely linked locus pairs

Blue data points: *BM1824* -*CSMM019* , orange data points:*CSMM019* -*BM3205* , light blue data point: HiP at 28.2 °S, grey data points: remaining three locus pairs. The regression line is the best fitting curve through the data, given by$r_{LD} = a \ln(\text{latitude} + 27) + b$ , with$a$ = -0.214 ($P$ = 0.014) and $b$ = 0.595.
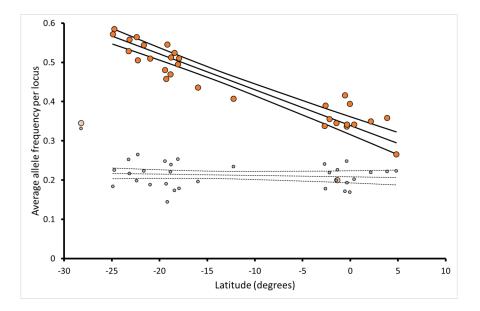
**Figures**

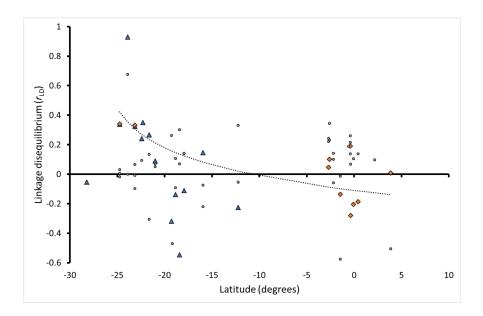16

Figure 1

Figure 2



Figure 3

Figure 4

**Supporting Information**

Text S1: Allele size standardization

Text S2: Dependency of results on sample size weighing, frequency correction and allele size standardization

Text S3: Results of the modified Breusch-Pagan test

Text S4: LD in KNP buffalo as a function of interlocus distance

Figure S1: Increase of pairwise $F_{ST}$ with geographic distance

Figure S2: Frequency clines of DE and SAE alleles in KNP

Figure S3: Combined frequency cline of DE and SAE alleles in KNP

Table S1: Overview of alleles selected for analysis

Table S2: Corrected average DE and SAE allele frequencies per population

Table S3: Allele frequencies in Nairobi NP and HiP relative to averages in nearby populations

Table S4: Linkage disequilibrium in the populations studied

Table S5: General linear mixed models of LD (expressed as $r_{LD}$) against latitude