Mammalian Biology 81 (2016) 595–603

Contents lists available at ScienceDirect

Mammalian Biology

journal homepage: www.elsevier.com/locate/mambio

Original investigation

Genome-wide single nucleotide polymorphism (SNP) identification and characterization in a non-model organism, the African buffalo (*Syncerus caffer*), using next generation sequencing

Nathalie Smitz[a,b,*], Pim Van Hooft[c], Rasmus Heller[d], Daniel Cornélis[e,f], Philippe Chardonnet[g], Robert Kraus[h,i], Ben Greyling[j], Richard Crooijmans[k], Martien Groenen[k], Johan Michaux[a,e]

[a] *Conservation Genetics, University of Liège, Boulevard du Rectorat 26, 4000 Liège, Belgium*
[b] *Joint Experimental Molecular Unit, Royal Museum for Central Africa, Leuvensesteenweg 11-17, 3080 Tervuren, Belgium*
[c] *Resource Ecology Group, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands*
[d] *Bioinformatics, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark*
[e] *Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), UPR AGIRS, Campus International de Baillarguet, F-34398 Montpellier, France*
[f] *Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD)-RP-PCP, University of Zimbabwe, Harare, Zimbabwe*
[g] *International Foundation for the Conservation of Wildlife (IGF), Rue de Téhéran 15, 75008 Paris, France*
[h] *Department of Biology, University of Konstanz, 78457 Konstanz, Germany*
[i] *Department of Migration and Immuno-Ecology, Max Planck Institute for Ornithology, Am Obstberg 1, 78315 Radolfzell, Germany*
[j] *Agricultural Research Council, Old Olifantsfontein Road, Irene Centurion 0062, South Africa*
[k] *Animal Breeding and Genomics Centre, Wageningen University, Droevendaalsesteeg 1, Wageningen, 6708 PB, The Netherlands*

ARTICLE INFO

Article history:
Received 17 March 2015
Accepted 28 July 2016
Handled by Paul Grobler
Available online 30 July 2016

Keywords:
Population genomics
Conservation
Disease ecology
Molecular markers

ABSTRACT

This study aimed to develop a set of SNP markers with high resolution and accuracy within the African buffalo. Such a set can be used, among others, to depict subtle population genetic structure for a better understanding of buffalo population dynamics. In total, 18.5 million DNA sequences of 76 bp were generated by next generation sequencing on an Illumina Genome Analyzer II from a reduced representation library using DNA from a panel of 13 African buffalo representative of the four subspecies. We identified 2534 SNPs with high confidence within the panel by aligning the short sequences to the cattle genome (*Bos taurus*). The average sequencing depth of the complete aligned set of reads was estimated at 5x, and at 13x when only considering the final set of putative SNPs that passed the filtering criterion. Our set of SNPs was validated by PCR amplification and Sanger sequencing of 15 SNPs. Of these 15 SNPs, 14 amplified successfully and 13 were shown to be polymorphic (success rate: 87%). The fidelity of the identified set of SNPs and potential future applications are finally discussed.

Introduction

The African buffalo (*Syncerus caffer*) has suffered of major population losses during the last century, impacting all but unevenly subspecies. Habitat loss, climatic changes, poaching and diseases are the main challenges currently threatening the species survival, contributing to local buffalo populations decimation. Direct competition for space and resources gradually appeared with the expansion of livestock farming and agriculture. Currently around 75% of the global African buffalo population is located in protected areas (East, 1999). The resulting disruption of natural wildlife population admixture is likely responsible for genetic erosion (Young and Clarke, 2000; Frankham et al., 2002). Isolated populations are likely to have lower reproductive fitness and lose their adaptive genetic variation, while presenting a higher risk of extinction (Frankham et al., 2002). Conservation genetics help in identifying and promoting appropriate management methods to reduce the risks of species extinction through the study of the spatial distribution of mutations between and among populations. Recent technological advances have revolutionized the generation of these genetic resources, allowing DNA-library construction, large-scale

* Corresponding author. Present address: MRAC, Leuvensesteenweg 11-17, 3080, Tervuren, Belgium.
E-mail address: nathalie.smitz@africamuseum.be (N. Smitz).

http://dx.doi.org/10.1016/j.mambio.2016.07.047
1616-5047/

sequencing and identification of single nucleotide polymorphism (SNP) genetic markers (Seeb et al., 2011). SNPs were shown to constitute highly informative markers (Morin et al., 2009) and lead to a better inference of population structure than microsatellites (Liu et al., 2005; Santure et al., 2010). Attention has begun to shift toward SNPs as preferred genetic markers due to their increased power of resolution and accuracy for studying fine scale population structure (Schlötterer, 2004). This is based on their high abundance throughout the genome, simple mutation characteristics, low mutation rates, usability on non-invasive samples and historical DNA, and standardization possibilities between laboratories (Kraus et al., 2014; Morin et al., 2007a,b, 2004; Luikart et al., 2003). SNPs have become an established marker in molecular ecology, evolutionary genetics, and animal breeding (Davey et al., 2011; Kraus et al., 2014, 2012; Morin et al., 2004; Santure et al., 2010).

Despite their attractiveness, some difficulties have been experienced in developing SNP in non-model organisms due to the limited or no genomic resources available, leading to complex laboratory screening of segments of the genome from multiple individuals to yield only a small number of independent SNPs. Next-generation sequencing (NGS) allows to overcome this issue by providing large-scale genome variation studies based on deep sequencing of relatively large genome fractions (>1%) or even the complete genome (Seeb et al., 2011). However, not so long ago, within non-model organisms, the predominant technique has been the targeted gene approach, using regular Sanger sequencing (Sanger et al., 1977), since it does not require species-specific pre-existing DNA data and is applicable to many taxa. A few hundred SNPs were identified using this approach for numerous species (e.g., 158 SNPs, *Sceloporus undulates*; 112 SNPs, *Salmo salar*; 768 SNPs, *Pusa hispida hispida*; 168 SNPs, *Thryothorus pleurostictus*) (Andreassen et al., 2010; Cramer et al., 2008; Olsen et al., 2011; Rosenblum et al., 2006). Only a few SNPs per species (<100) have been developed using the targeted gene approach for animals of conservation concern such as the marmoset (*Saguinus oedipus*), the dhole (*Cuon alpinus*) and the elephant (*Loxodonta africana*) (Aitken et al., 2004). The targeted gene approach, although still widely used, is laborious, time consuming, costly and yields only a fairly limited amount of candidate SNPs in contrast to NGS.

The Reduced Representation Library (RRL) approach is a NGS method that involves a digestion step of multiple genomic DNA samples with restriction enzyme(s), a selection of the resulting restriction fragments and a sequencing step. RRL approaches have been used to generate tens of thousands to millions of candidate SNPs with a genome-wide coverage for example in cattle (Tassell et al., 2008), turkey (Kerstens et al., 2009) and great tit (Van Bers et al., 2010). Alternatively, SNP resources from one species could be used in a closely related species. An Illumina BovineSNP50 BeadChip has been developed for cattle (*Bos taurus*), a close relative to the African buffalo (Matukumalli et al., 2009). This BeadChip scores 54,001 informative SNPs that are uniformly distributed along the entire cattle genome. It has a high cross amplification success rate across cattle breeds (Matukumalli et al., 2009). However, when used on other bovid species, the number of polymorphic sites decreases substantially. Only a few percent of all SNPs on the chip were still polymorphic (Miller et al., 2010) when tested on other species such as the water buffalo (*Bubalus bubalis* −1159 SNPs), the Yak (*Bos grunniens* −949), the North American Bison (*Bison bison* −1604), and the Banteng (*Bos javanicus* −1429) (Michelizzi et al., 2011). Similar results were obtained when testing the OvineSNP50 BeadChip, developed for domestic sheep, on two related ovid species (Miller et al., 2010). Cross-species amplification of SNP assays usually does not work well compared to cross-species amplification of microsatellites (Kraus et al., 2012). Even if genotyping is successful, many polymorphisms in one species are fixed in the other. Moreover, cross-species SNPs may harbor extreme biases in allele frequencies, since they may predominantly be found in regions of the genome under natural selection favoring polymorphism (e.g., balancing selection).

Since cross-species genotyping of SNPs often seems problematic or biased, this study aims to characterize a genome-wide set of SNPs specifically for the African buffalo over its whole distribution area (sub-Saharan Africa). A previous study conducted by Le Roex et al. (2012) already aimed at identifying SNPs in the African buffalo, however their sampling was limited to the Cape buffalo subspecies (*Syncerus caffer caffer*) and to the Hluhluwe-iMfolozi National Park (NP). The buffalo population within this National Park is known to be affected by strong non-equilibrium conditions linked to a founder event (Smitz et al., 2014; Du Toit, 1954; Kappmeier et al., 1998). In the present study, Next Generation Sequencing of reduced representation libraries for SNP discovery was used. The genome of another Bovid species, *Bos taurus*, which diverged from African buffalo approximately 12 million years ago, was used as a reference for mapping the reads (Hassanin and Ropiquet, 2004; Pitra et al., 2002; Robinson and Ropiquet, 2011; TimeTree software- Hedges et al., 2006; Kumar and Hedges, 2011). The present study allowed the identification of 2534 SNPs with high confidence by aligning short sequences of the African buffalo (*Syncerus caffer*) to the cattle genome (*Bos taurus*).

Material and methods

Sample collection and library preparation

A geographically large and diverse panel of African buffalo was sampled: 6 from East and Southern Africa [South Africa(2), Uganda(1), Kenya(1), Ethiopia(1), Namibia(1)] belonging to the *Syncerus caffer caffer* subspecies, and 7 from West and Central Africa [Central African Republic(1), Niger(3), Chad(2), Burkina Faso(1)] belonging to the *S.c. nanus*, *S.c. brachyceros* and *S.c. aequinoctialis* subspecies respectively (Fig. 1). These subspecies were grouped together because phylogenetic studies showed that they form one clade with only minor to moderate F_{ST} differentiation between subspecies, ranging between 0.02 and 0.12 (Smitz et al., 2013; Van Hooft et al., 2002). Sample extraction, selection and RRL library preparation procedures are available as Supplementary information (Supplementary file 1).

Sequence filtering

Prior to the sequence alignment steps, different filters were applied to the raw Illumina sequence data according to several criteria. First, sequences were expected to start with a CT dinucleotide because of the *Alu*I restriction site (between AT and CT). All sequences not bearing this pattern were discarded as potential contamination. Secondly, average quality scores were calculated for each read by taking the mean of all individual scores at each of the 76 positions. Reads presenting low overall phred quality scores were removed (Ewing and Green, 1998). Moreover, end of reads displaying two successive read positions with average phred quality scores lower than 20 were trimmed from the first read position with a phred <20.

Sequence mapping and SNP discovery

Quality filtered and trimmed sequence reads were aligned to the bovine reference genome (*Bos taurus*; UCSC Genome Bioinformatics; http://genome.ucsc.edu/http://genome.ucsc.edu/ (21/01/2015)) since an African buffalo genome sequence is not available. The MosaikAssembler software (Mosaik 1.0.1388- Stromberg, 2010) was used with default settings, specifying a median fragment length of 50 bp (i.e., inner mate distance) with

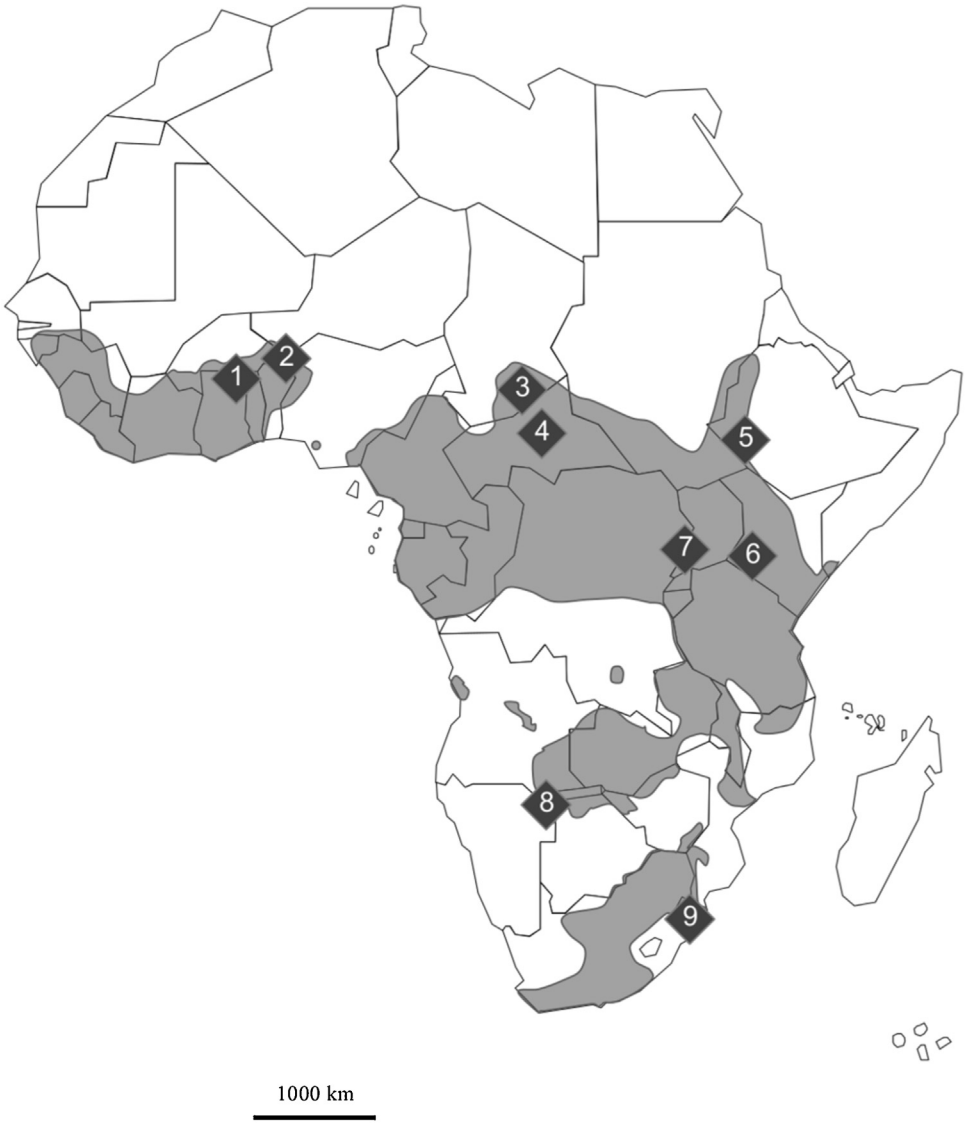

Fig. 1. Sampling localities of African buffalo included in this study. West-central clade (*S.c. nanus*, *S.c brachyceros*, *S.c. aequinoctialis*) and south-eastern lineage (*S.c. caffer*) were sampeled; 1, Pama Reserve; 2, W NP; 3, Aouk NP and Zakouma NP; 4, N'gotto-Kindo; 5, Omo NP; 6, Nakuru NP; 7, Queen Elizabeth NP; 8, Waterberg NP; 9, Hluhluwe Umfolozi NP.

a search radius of 50 bp to search for missing mate and for an alignment that conforms to the paired-end orientation. The ALL alignment mode was used with a hash size of 15 (all hashing strategy), a maximum percentage of mismatches allowed of 15, and a minimum cluster size of 35. Consensus candidate SNPs were extracted from the dataset using the SAMTools software with the pileup function (SAMTools 0.1.7 – Sequence Alignment/Map- (Li et al., 2009a,b)). Candidate SNPs were then filtered for having a phred quality score above 20 (quality of base calling >99%), for having a mapping quality score above 30, and a minor allele occurrence at the polymorphic position of at least three times. Positions that were monomorphic within the SNP discovery pool but showed a fixed difference with the reference genome (*Bos taurus*) were discarded. Finally, SNPs with a four times higher read depth than the average read depth of the RRL (average total number of reads aligned with the reference genome to a unique position) were also discarded (Kerstens et al., 2009), as these are likely to be false SNPs that are the result of alignment of parologous sequences.

For the validation of the filtered SNP set, primers had to be designed in the direct flanking region around the SNPs. Therefore, a buffalo consensus sequence was generated from the RRL sequences and flanking sequences around SNP positions were extracted. Where possible, flanking regions were generated based on the specific African buffalo consensus sequence. If a SNP was situated close to the beginning or end of the reads, flanking regions for each SNP were generated using part of the *Bos taurus* genome and concatenated with African buffalo consensus information ('chimeric flanking sequences' (Jonker et al., 2012)).

Validation

Our set of SNPs was validated by randomly selecting 15 SNPs scattered on the whole genome and by amplifying them by standard PCR. Only SNPs that reached at least an Illumina design score of 0.6 in Illumina's Assay Design Tool available at http://www.illumina.com/support/array/array_software (21/01/2015) were selected. Primers were designed using Primer3 (Rozen and Skaletsky, 2000; http://simgene.com/Primer3 (21/01/2015)), entering our chimeric flanking sequences (Table 1). As the cattle genome was used to align our merged sequences, we wanted to specifically test the sequence conservation between *Bos taurus* and *Syncerus caffer*. If sequence conservation is high, amplifi-

Table 1
Primer sets used for SNP validation step (n/a: no amplification).

Chr nr. bp position	Primer set	SNP	Observed polymorphism
Chr1.75025101	TTTGGATCAGGAGGAACCAG CCCCTTTGGTGGAACATTTA	A/G	Yes
Chr5.14429154	AAAGGATTTCTGTTGGTGGAGA GATTTGCCTTCTCAAACTGGA	A/G	Yes
Chr6.119157451	TGAAATCTAACTGCCTGGGACT CAGGTGTGCTGGTTTACAGG	C/T	Yes
Chr9.106373371	AGTCTGCCTAAAAAGCCCATT CCTCCCACGCACAGACTC	A/C	Yes
Chr10.4010071	TCACCTGAATCCCACCCTTA CTCGAGAAGGGCTTTGTGAC	A/C	Yes
Chr11.72584688	AACACCCCACCTTAATGCAG GTCAGGAGAGGGCTGTCAAG	C/T	Yes
Chr11.70625560	GCCATAAGGGTGGTGTCATC CCATGGACATCCTTTTCCTG	C/T	No
Chr12.20644129	TCCATGCCCATCTGAGATTT CCTGGCCTGACTCTGAGGTA	A/C	Yes
Chr14.80545216	GAGATCCCACTCGGCTGTTA AACCGTGAGCGAAGTGAGAG	A/T	Yes
Chr15.23524504	GATGGACTTGGTGGCAATTT GCCTCAGGACCATTTTCAGA	A/G	Yes
Chr15.81001941	GCTTGTTCAGATGGCACAGA GCCAGTACTCCCCCTAGCTC	A/C	n/a
Chr16.62257511	GCGTTCCTTCAACAACCAAG GCCATCTTGATTTCCTTCCA	C/T	Yes
Chr17.4151052	TCCCAGAGCAGACAGTCTCA CGGTGATCATCTGCTAATGC	A/C	Yes
Chr17.74836302	CCCTCCACTAGCTTCTCAGC AGTGGAGCTGAGGTCTTGGA	C/T	Yes
Chr19.7345443	CATAATCCCAGCCAGTCTCC GAGAGCACCCCTGAGTTGAA	C/T	Yes

cation of SNPs will be successful when primers are designed within the adjacent region of the bovine genome surrounding the read containing the SNP candidate. Therefore, three situations (five SNPs per situation) were tested: (1) both forward and reverse primers designed on the bovine genome where no RRL reads aligned, (2) one forward primer designed on the reference genome and one reverse primer designed on the RRL reads (*Syncerus caffer*), and (3) both primers designed solely from the RRL reads. Thus, validation required SNPs firstly to amplify successfully and secondly to be polymorphic. Each of these 15 SNPs was sequenced in all 13 individuals used in the original discovery pool. The PCR reaction took place in a total volume of 12 μl, consisting of 3 μl DNA (10 ng/μl), 0.5 μl of primers (0.03 μg/μl), 5.2 μl Mastermix (ThermoScientific) and 0.3 μl AmpliTaq® DNA Polymerase. Cycling conditions consisted of 36 cycles for 30 s at 95 °C, 30 s at 50 °C and 30 s at 72 °C. An initial denaturation step preceded the process (5 min at 95 °C), and a final extension step followed the process (10 min at 72 °C). Sequencing was performed on an ABI 3730XL capillary sequencer. The resulting sequences were aligned using CLUSTAL_X (Thompson et al., 1997) as implemented in BIOEDIT v.7.09 (Hall, 1999) and SNPs were validated visually.

Results

Sequencing of the RRL and read filtering

The *Alu*I restriction enzyme was chosen for the construction of the African buffalo RRL since it maximized the quantity of fragments situated in the targeted size range of 100–200 bp, evaluated performing an *in silico* digestion of the *Bos taurus* genome. Correspondences between *in silico* and *in vitro* observed restriction enzyme cleavage patterns were previously demonstrated within other mammal species (Abdurashitov et al., 2006, 2007). In total, 18.5 million paired-end sequences of 76 bp length were generated by the Illumina Genome Analyzer II on two lanes. The genome coverage (Mosaik 1.0.1388, Stromberg, 2010) was estimated at about 5% of coding and non-coding regions (Fig. 2). The average phred quality score per read position dropped below 20 after position 55 for about 5 million reads. To maintain sufficient quality for SNP detection, those reads were trimmed after position 55. The average sequencing depth of the whole aligned set of reads was estimated at 5x and at 13x after filtering steps.

Read alignment to the reference genome

The cattle genome consists of 29 autosomes and the sex chromosomes, with a total estimated genome size of 2.87Gb (Liu et al., 2009). In total, about 60% of the reads were not retained because they did not pass the filtering criteria of the alignment: they were too short, were not unique (i.e., align to more than one location) or contained too many nucleotide differences. Eventually, 6.9 million reads remained for the SNP discovery and could be successfully aligned to the *Bos taurus* reference genome, corresponding to 836.5 million bp. From these reads, 22% were orphans (i.e., only one of the paired read aligned to the reference, while the other did not), while 14% had one paired read that was non-unique. The physical distribution of the identified SNPs across the buffalo genome was estimated using the cattle genome as the reference.

SNP detection

A total of 318,091 putative SNPs were detected. Fixed differences between the discovery panel of African buffalo and the cattle reference genome were discarded (i.e., 22,472). The few SNPs (1.7%) with more than two alleles were also discarded as tri- or tetra-allelic SNPs are uncommon and are more likely to be the result of sequencing errors than real polymorphism (Brookes, 1999). Furthermore, most genotyping assay designs do not allow for more than two alleles. After filtering these SNPs for minor allele count (minor allele occurring at least three times), for minimum phred quality score of 20 and for minimum mapping quality score of 30, 2534 SNPs in which we place high confidence remained distributed across the entire genome (Table 2). The sequencing depth had an average of 13 reads after filtering steps. A total number of 1837 SNPs had an Illumina design score $\geq$0.6.

SNP quality assessment

The ratio of transitions (TS; i.e., C/T or A/G) versus transversions (TV; i.e., A/T, G/C, A/C or G/T) was estimated as a measure for the quality of the SNP discovery. The TS:TV ratio observed within our dataset was 2.38:1 (1784 transitions versus 750 transversions), with a nearly equal number of A/G and T/C mutations (889 A/G and 895 T/C), and the four TV changes occurring at similar frequencies. This is the expected empirical ratio, while ratios substantially lower than 2 can be indicative of random genotyping error (Kraus et al., 2012). The TS:TV ratio remained similar when plotted per read position (Fig. 3), which is a good indication that there was no read position bias, such as false SNPs due to low sequencing quality towards the ends of reads (Kraus et al., 2012). Moreover, SNP predictions were tested by determining whether particular regions of sequence reads presented more SNP candidates than others. Previous studies have shown that tails of reads present excessively more sequencing errors, leading to false SNPs identification (Dohm et al., 2008; Van Bers et al., 2010). Within the SNP set passing all filtering

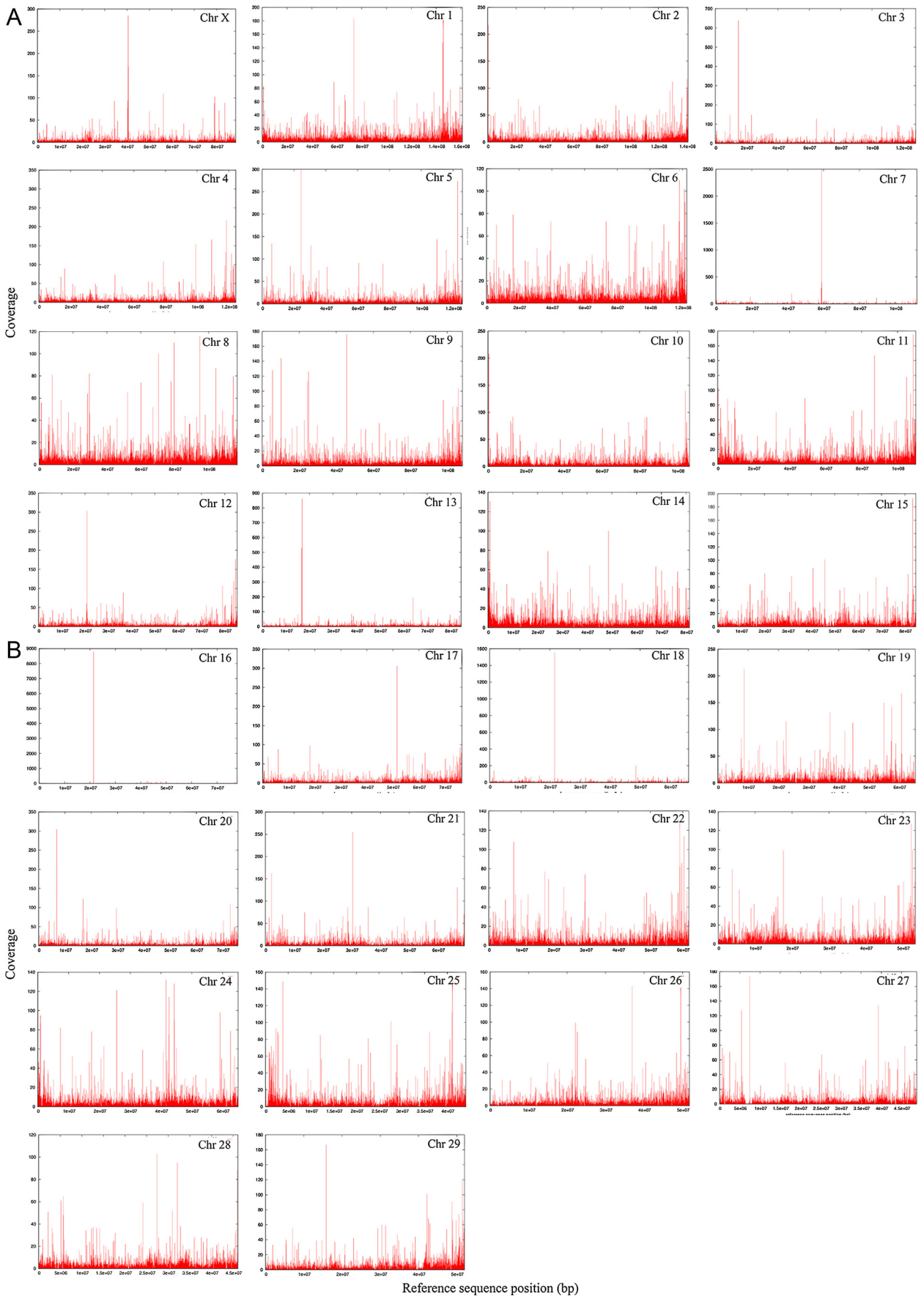

Fig. 2. Reference sequence coverage of the 30 bovine chromosomes (*Bos taurus*).

Table 2
Number of high confidence SNPs on each of the 30 bovine chromosomes (*Bos taurus*).

Chr nr	SNP nr
1	89
2	98
3	135
4	92
5	105
6	51
7	103
8	70
9	54
10	73
11	114
12	58
13	97
14	69
15	80
16	68
17	72
18	90
19	112
20	50
21	102
22	106
23	89
24	89
25	119
26	84
27	73
28	61
29	92
X	39

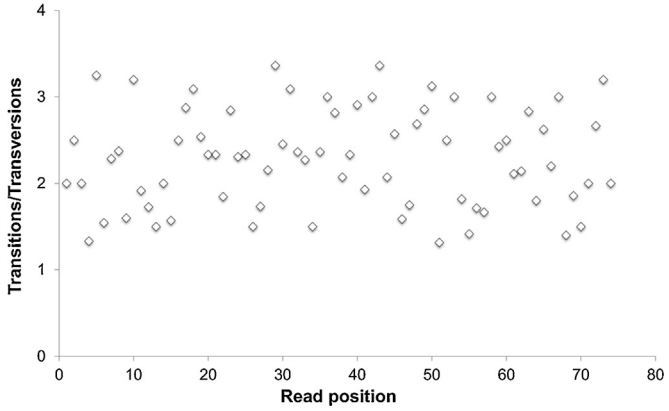

Fig. 3. Average TS:TV ratio at each read position.

criteria, no overrepresentation of SNP candidates in the ends of the reads was observed.

SNP validation

Only one primer pair designed within the RRL reads failed to yield an amplification product. Within the 14 remaining amplification products, one did not contain the expected polymorphism, which means that 87% of the expected SNPs were confirmed by Sanger sequencing of the individuals in the discovery panel. This high percentage of successful PCR amplification is similar to that observed in geese (93% of 384 SNPs tested) using the same chimeric technique (Jonker et al., 2012). The ten primer set entirely or partly designed within the cattle genome gave PCR products with expected SNPs observed. This corroborates the high genome conservation between cattle and buffalo.

Discussion

Model species reference genome and SNP validation

The present study enabled the identification of 2534 SNPs with high confidence in a non-model organism. 1837 SNPs had an Illumina design score ≥0.6, reflecting a high likelihood that assay design will be successful on a modern high throughput SNP genotyping platform. About 30% of the sequence reads could be aligned to the bovine genome, a closely-related species. The study of Jonker et al. (2012) used the same technique to identify SNPs in the Barnacle Goose (*Branta leucopsis*) by aligning 1.77 million reads to the Mallard (*Anas platyrhynchos*) genome (divergence time 30 million years) (Huang et al., 2013). In that study, 16.1% of the reads successfully aligned, subsequently allowing the identification of 2188 high confidence SNPs. In the African buffalo, Le Roex et al. (2012) mapped 19–23% of their short reads (50 bp reads) to the domestic cow genome (*Bos taurus*). Our study confirms that using a genome of a closely related species as a reference standard provides a sufficient number of high confidence SNPs and offers a good alternative to characterize SNPs in non-model species without carrying out tedious steps of deep assembly of redundant contigs (Kerstens et al., 2009; Van Bers et al., 2010).

The chimeric flanking sequences obtained from the cattle genome were also used to generate primers for validation steps. High PCR amplification success with chimeric primers indicated sufficient conservation between the genomes of the two species to use them for genotyping assay design. The high PCR amplification success could likely be attributed to the correspondence of the aligned filtered reads to highly conserved sequences. Van Hooft et al. (1999) previously demonstrated high genome conservation when using microsatellites primers developed for cattle on African buffalo, with 83% successful amplification.

Ascertainment bias

Ascertainment bias can result from the selection of loci from an unrepresentative sample of individuals. To limit this kind of bias, a relatively large pool of samples covering the whole distribution area of the targeted species was selected, comprising all four currently taxonomically recognized subspecies of African buffalos. However, ascertainment bias can also be introduced by limited read depth. By stipulating a minor allele count of three in our protocol, sequencing depth should at least comprise six reads for a SNP to be retained. Our average sequencing depth of the whole putative SNP database was estimated at 5x, which increased to 13x when estimated on the SNP set that passed the filtering criterion. This sequencing depth remains low compared to other studies (e.g., 25x (Van Bers et al., 2010), 58x (Kerstens et al., 2009)). The studies most similar to ours, to our knowledge, are that of Jonker et al. (2012) (*Branta leucopsis*), which yielded an average sequence depth of 9.9x, and that of Le Roex et al. (2012) (*S.c. caffer*), which yield an average sequence depth of 2.7x. Our low sequence depth may be explained by an over-representation of size-fractionated fragments ranging between 100 and 200 bp sliced from the polyacrylamide gel. Consequently, many true rare variants may certainly have been rejected. Moreover, the low depth of coverage also implies that only SNPs present in multiple samples of our DNA pool had a chance to be identified. Overrepresentation of common SNPs over rare SNPs is thus expected to introduce bias into our SNP set. This needs to be taken into account when interpreting genotypic data in future projects. However, by our geographically broad sampling design we avoid the substantial geographic ascertainment bias that is likely present in Le Roex et al. (2012), because their SNP discovery panel was limited in geographic extent.

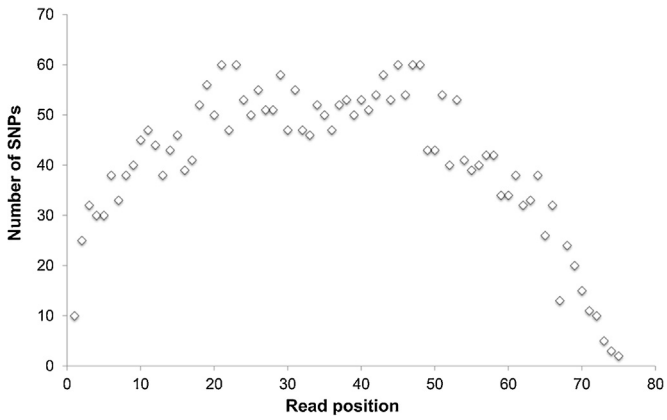

Fig. 4. Cumulative number of SNP occurrence at each read position.

Sequencing error

Different estimators were used to evaluate the risk of false positives in the SNP discovery analysis. A high TS:TV ratio is considered a good measure of SNP validity. A TS:TV ratio of 1:2 would be detected if mutations were random and is therefore an indication for sequencing errors. Higher rates of C/T mutations due to the deamination of methylcytosines in CpG dinucleotides are responsible for a higher TS:TV ratio in real data (Cooper et al., 2010; Scarano et al., 1967; Vignal et al., 2002). Usually, the ratio of 2.1:1 is observed in mammals (DePristo et al., 2011). A ratio significantly lower than this last one can therefore be an indicator of poor quality sequencing data. Our TS:TV ratio of 2.4:1 is similar to the results obtained for example in the study of Kraus et al. (2011) (2.3:1) and Jonker et al. (2012) (2.7:1). It thus indicates that most detected SNP calls were not random, which reflects that our SNPs likely represent true nucleotide polymorphism.

Misidentification of SNPs due to sequencing errors is avoided by excluding variation with a low phred score. Tails of reads often display increasingly more sequencing errors using Illumina's technology. Even though we trimmed our reads during quality checks, a decrease in the number of predicted SNPs in the tail of the reads was observed and may be explained by an associated decrease of the associated phred score (Fig. 4). This decrease in predicted SNPs per position in Illumina sequencing reads was also reported in earlier studies (Kerstens et al., 2009; Ramos et al., 2009; Van Bers et al., 2010), on which the current SNP detection pipeline is based.

The high validation success in our study can also illustrate the quality of the predicted SNPs. From the set of 15, only one did not amplify whose primers were designed based on the generated reads (Table 1). Therefore, a closely related species genome can be used both for mapping and SNP discovery, but also to design the primers for the genotyping step. Another SNP appeared to be monomorphic, yielding a conversion rate of 87% of polymorphic SNPs. This validation success rate is higher than that of Le Roex et al. (2012) working on the Cape buffalo (*S.c. caffer*). There, within the set of 173 SNPs used for the validation, 143 amplified successfully and only 75 were polymorphic. The false positives in SNP discovery in the study of Le Roex et al. (2012) seems to be linked to the low coverage (mean 2.7x), and to the fact that a SNP was inferred if the nucleotide variant was supported by a minimum of two reads (*vs.* 6 in our study). Using such a low cut-off may significantly increase the risk of identifying false positives. Applying a minor allele count of at least three minimizes false positives in the SNP discovery analysis. This approach, however, also dramatically reduces the identification of true nucleotide polymorphisms that could, in principle, be detected even if it would also increase chances of identifying a sequencing error as an SNP.

Utility of SNPs in African buffalo

SNP markers can provide major insights into animal dispersal patterns. This is especially relevant in the light of recent conservation initiatives aiming to restore genetic diversity of wildlife stocks by re-establishing demographic connectivity between wildlife populations of different NP (e.g., Great Limpopo Transfrontier Conservation Area). Dispersal beyond traditional conservancy boundaries, and also national borders, may pose a risk as far as the spread of pathogens are concerned (Cross et al., 2004, 2005). Among wildlife species, buffalos are known to be one of the main wildlife reservoirs for diseases (Rodwell et al., 2001). Since buffalo are closely related to cattle, and may transmit disease directly or indirectly, buffalo also represent an important threat for the African livestock industry, from a conservation, sanitary and economic point of view (Garine-Wichatitsky et al., 2010; Jolles et al., 2005; Michel et al., 2006). For fine scale inference, a larger number of SNPs may be required, as the information content of one SNP is less than that of one microsatellite (i.e., bi-allelic *vs.* multi-allelic markers). Previous studies revealed that four to twelve times more SNPs are needed for population structure inference to match the statistical power of one microsatellite (Liu et al., 2005). For highly dispersive organisms it has been shown that the detection of low levels of differentiation is possible with a minimum of 80 SNPs (Morin et al., 2009; Ryman et al., 2006). Our large set of hundreds of SNPs should thus allow to scale the genetic marker system to the needs of future studies of the interaction between landscape features and microevolutionary processes (Manel et al., 2003).

This SNP database may also be of benefit in the context of selective breeding. Indeed, selective breeding of specific phenotypes of the Cape buffalo subspecies (*S.c. caffer*- South-Eastern Africa) has become an intricate business within private game farming. Females are being selected for horn length, milk production and regular calving intervals, while males are being selected for horn size, body mass and shape, which are desirable to trophy hunters. However, such approaches may lead to distortion of evolutionary natural processes and may reduce the species genetic variability thereby weakening the species resilience in the wild. These practices are not believed to benefit the conservation of global biodiversity, and may even become problematic if genetic dilution occurs through escapes of selected individuals into neighboring wild populations of buffalo. Future development of guidelines in collaboration with the game-farming breeders should allow finding compromises for the long-term conservation of the wildlife species.

Conclusions

Within a highly mobile species such as the African buffalo, the SNP set developed in this study should provide highly valuable and reliable tools for gaining insight into the migratory pattern of this species, known to be a disease reservoir. Our approach yielded higher quality SNPs (as judged by assay conversion rate) and less geographically biased SNPs than a previous study (Le Roex et al., 2012). Furthermore, the construction of chimeric flanking sequence was shown to increase the number of usable SNPs by providing sufficiently large regions for the genotyping assay.

Conflict of interest

The authors declare to have no conflicts of interest. They developed all aspects of this study. The sponsors of the issue had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

The present research study was designed in collaboration with Pim Van Hooft, Rasmus Heller, Johan Michaux, Richard Crooijmans, Martien Groenen and Ben Greyling. Partial samples were provided by Daniel Cornélis and Philippe Chardonnet. Statistical analysis and interpretation was performed by Nathalie Smitz, with assistance of Robert Kraus, Richard Crooijmans and Martien Groenen. All co-authors participated to the paper writing.

Acknowledgements

We like to thank the support of the Research Platform "Production and Conservation in Partnership" (RP-PCP). We would also like to thank F. Jori, B. Van Vuuren, K.L. Kanapeckas and all collectors for providing us the samples used for the discovery of the SNP set. Technical assistance in the laboratory was provided by Bert Dibbits, and Rudy Jonker helped with constructing chimeric flanking sequences. Proof reading assistance was provided by Virginie Winant. This project was supported by the network "Bibliothèque du Vivant" funded by the CNRS, the "Musée National d'Histoire Naturelle", the INRA and the "Centre National de Séquençage". This study is supported by research grants from the FRS-FNRS of Belgium provided to J.R. Michaux and N.M.R. Smitz.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.mambio.2016.07.047.

References

Abdurashitov, M.A., Tomilov, V.N., Chernukhin, V.A., Gonchar, D.A., Degtyarev, S. Kh., 2006. Mammalian chromosomal DNA digestion with restriction endonucleases in silico. Ovchinnikov Bull. Biotechnol. Phys. Chem. Biol. 2, 29–38.

Abdurashitov, M.A., Tomilov, V.N., Chernukhin, V.A., Gonchar, D.A., Degtyarev, S. Kh., 2007. Comparative analysis of human chromosomal DNA digestion with restriction endonucleases in vitro and in silico. Med. Genet. 6, 29–36.

Aitken, N., Smith, S., Schwarz, C., Morin, P.A., 2004. Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. Mol. Ecol. 13, 1423–1431, http://dx.doi.org/10.1111/j.1365-294X.2004.02159.x.

Andreassen, R., Lunner, S., Høyheim, B., 2010. Targeted SNP discovery in Atlantic salmon (Salmo salar) genes using a 3′UTR-primed SNP detection approach. BMC Genomics 11, 706, http://dx.doi.org/10.1186/1471-2164-11-706.

Brookes, A.J., 1999. The essence of SNPs. Gene 234, 177–186.

Cooper, D.N., Mort, M., Stenson, P.D., Ball, E.V., Chuzhanova, N.A., 2010. Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. Hum. Genomics 4, 406–410.

Cramer, E.R.A., Stenzler, L., Talaba, A.L., Makarewich, C.A., Vehrencamp, S.L., Lovette, I.J., 2008. Isolation and characterization of SNP variation at 90 anonymous loci in the banded wren (Thryothorus pleurostictus). Conserv. Genet. 9, 1657–1660, http://dx.doi.org/10.1007/s10592-008-9511-7.

Cross, P., Lloyd-Smith, J., Bowers, J., Hay, C., Hofmeyr, M., Getz, W., 2004. Integrating association data and disease dynamics: an illustration using African buffalo in Kruger National Park. Ann. Zool. Fennici 41, 879–892.

Cross, P.C., Lloyd-Smith, J.O., Getz, W.M., 2005. Disentangling association patterns in fission-fusion societies using African buffalo as an example. Anim. Behav. 69, 499–506, http://dx.doi.org/10.1016/j.anbehav.2004.08.006.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12, 499–510, http://dx.doi.org/10.1038/nrg3012.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K., Maguire, V., Hartl, J.R., Philippakis, C., Angel, A.A., del, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498, http://dx.doi.org/10.1038/ng.806.

Dohm, J.C., Lottaz, C., Borodina, T., Himmelbauer, H., 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res. 36, e105, http://dx.doi.org/10.1093/nar/gkn425.

Du Toit, R., 1954. Trypanosomiasis in Zululand and the control of tsetse flies by chemical means. Onderstepoort J. Vet. Res. 26, 317–387.

East, R., 1999. African Antelope Database 1999. Gland: IUCN, Switzerland and Cambridge.

Ewing, B., Green, P., 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8, 186–194.

Frankham, R., Ballou, J.D., Briscoe, D.A., 2002. Introduction to Conservation Genetics. Cambridge University Press, Cambridge.

Garine-Wichatitsky, M., de Caron, A., Gomo, C., Foggin, C., Dutlow, K., Pfukenyi, D., Lane, E., Bel, S., Le Hofmeyr, M., Hlokwe, T., Michel, A., 2010. Bovine tuberculosis in buffaloes, Southern Africa. Emerg. Infect. Dis. 16, 884–885, http://dx.doi.org/10.1890/02-5266.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Hassanin, A., Ropiquet, A., 2004. Molecular phylogeny of the tribe Bovini (Bovidae, Bovinae) and the taxonomic status of the Kouprey, Bos sauveli Urbain 1937. Mol. Phylogenet. Evol. 33, 896–907, http://dx.doi.org/10.1016/j.ympev.2004.08.009.

Hedges, S.B., Dudley, J., Kumar, S., 2006. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22, 2971–2972, http://dx.doi.org/10.1093/bioinformatics/btl505.

Huang, Y., Li, Y., Burt, D.W., Chen, H., Zhang, Y., et al., 2013. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. Nat. Genet. 45, 776–784, http://dx.doi.org/10.1038/ng.2657.

Jolles, A.E., Cooper, D.V., Levin, S.A., 2005. Hidden effects of chronic tuberculosis in African buffalo. Ecology 86, 2358–2364, http://dx.doi.org/10.1890/05-0038.

Jonker, R.M., Zhang, Q., Van Hooft, P., Loonen, M.J.J.E., Van der Jeugd, H.P., Crooijmans, R.P.M.A., Groenen, M.A.M., Prins, H.H.T., Kraus, R.H.S., 2012. The development of a genome wide SNP set for the Barnacle goose Branta leucopsis. PLoS One 7, e38412, http://dx.doi.org/10.1371/journal.pone.0038412.

Kappmeier, K., Nevill, E.M., Bagnall, R.J., 1998. Review of tsetse flies and trypanosomosis in South Africa. Onderstepoort J. Vet. Res. 65, 195–203.

Kerstens, H.H.D., Crooijmans, R.P.M., Veenendaal, A., Dibbits, B.W., Chin-A-Woeng, T.F.C., den Dunnen, J.T., Groenen, M.A.M., 2009. Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: applied to turkey. BMC Genomics 10, 479, http://dx.doi.org/10.1186/1471-2164-10-479.

Kraus, R.H.S., Kerstens, H.H.D., Van Hooft, P., Crooijmans, R.P.M.A., Van Der Poel, J.J., Elmberg, J., Vignal, A., Huang, Y., Li, N., Prins, H.H.T., Groenen, M.A.M., 2011. Genome wide SNP discovery, analysis and evaluation in mallard (Anas platyrhynchos). BMC Genomics 12, 150, http://dx.doi.org/10.1186/1471-2164-12-150.

Kraus, R.H.S., Kerstens, H.H.D., van Hooft, P., Megens, H.-J., Elmberg, J., Tsvey, A., Sartakov, D., Soloviev, S.A., Crooijmans, R.P.M.A., Groenen, M.A.M., Ydenberg, R.C., Prins, H.H.T., 2012. Widespread horizontal genomic exchange does not erode species barriers among sympatric ducks. BMC Evol. Biol. 12, 45, http://dx.doi.org/10.1186/1471-2148-12-45.

Kraus, R.H.S., Vonholdt, B., Cocchiararo, B., Harms, V., Bayerl, H., Uhn, R.K., Orster, D.W.F., Roos, C., 2014. A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. Mol. Ecol. Resour., http://dx.doi.org/10.1111/1755-0998.12307.

Kumar, S., Hedges, S.B., 2011. TimeTree2: species divergence times on the iPhone. Bioinformatics 27, 2023–2024, http://dx.doi.org/10.1093/bioinformatics/btr315.

Le Roex, N., Noyes, H., Brass, A., Bradley, D.G., Kemp, S.J., Kay, S., van Helden, P.D., Hoal, E.G., 2012. Novel SNP discovery in African buffalo syncerus caffer, using high-throughput sequencing. PLoS One 7, e48792, http://dx.doi.org/10.1371/journal.pone.0048792.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009a. The sequence alignment/map (SAM) format and SAM tools. Bioinformatics 25, 2078–2079.

Li, S., Wan, H., Ji, H., Zhou, K., Yang, G., 2009b. SNP discovery based on CATS and genotyping in the finless porpoise (Neophocaena phocaenoides). Conserv. Genet. 10, 2013–2019.

Liu, N., Chen, L., Wang, S., Oh, C., Zhao, H., 2005. Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. BMC Genet. 6 (Suppl. 1), S26, http://dx.doi.org/10.1186/1471-2156-6-s1-s26.

Liu, Y., Qin, X., Song, X.-Z.H., Jiang, H., Shen, Y., Durbin, K.J., Lien, S., Kent, M.P., Sodeland, M., Ren, Y., Zhang, L., Sodergren, E., Havlak, P., Worley, K.C., Weinstock, G.M., Gibbs, R.A., 2009. Bos taurus genome assembly. BMC Genomics 10, 180, http://dx.doi.org/10.1186/1471-2164-10-180.

Luikart, G., England, P.R., Tallmon, D., Jordan, S., Taberlet, P., 2003. The power and promise of population genomics: from genotyping to genome typing. Nat. Rev. Genet. 4, 981–994, http://dx.doi.org/10.1038/nrg1226.

Manel, S., Schwartz, M.K., Luikart, G., Taberlet, P., 2003. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol. Evol. 18, 189–197, http://dx.doi.org/10.1016/S0169-5347(03)00008-9.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connell, J., Moore, S.S., Smith, T.P.L., Sonstegard, T.S., Van Tassell, C.P., 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4, e5350, http://dx.doi.org/10.1371/journal.pone.0005350.

Michel, A.L., Bengis, R.G., Keet, D.F., Hofmeyr, M., DE Klerk, L.M., Cross, P.C., Jolles, A.E., Cooper, D., Whyte, I.J., Buss, P., Godfroid, J., 2006. Wildlife tuberculosis in South African conservation areas: implications and challenges. Vet. Microbiol. 112, 91–100, http://dx.doi.org/10.1016/j.vetmic.2005.11.035.

Michelizzi, V.N., Wu, X., Dodson, M.V., Michal, J.J., Zambrano-varon, J., Mclean, D.J., Jiang, Z., 2011. A global view of 54, 001 single nucleotide polymorphisms (SNPs) on the illumina BovineSNP50 BeadChip and their transferability to water buffalo. Int. J. Biol. Sci. 7, 18–27.

Miller, J.M., Poissant, J., Kijas, J.W., Coltman, D.W., 2010. A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. Mol. Ecol. Resour. 11, 314–322, http://dx.doi.org/10.1111/j.1755-0998.2010.02918.x.

Morin, P., Luikart, G., Wayne, R.K., 2004. SNPs in ecology, evolution and conservation. Trends Ecol. Evol. 19, 208–216, http://dx.doi.org/10.1016/j.tree.2004.01.009.

Morin, P., Mccarthy, M., 2007a. Highly accurate SNP genotyping from historical and low-quality samples. Mol. Ecol. Notes 7, 937–946, http://dx.doi.org/10.1111/j.1471-8286.2007.01804.x.

Morin, P.A., Aitken, N.C., Rubio-Cisneros, N., Dizon, A.E., Mesnick, S., 2007b. Characterization of 18 SNP markers for sperm whale (Physeter macrocephalus). Mol. Ecol. Notes 7, 626–630, http://dx.doi.org/10.1111/j.1471-8286.2006.01654.x.

Morin, P., Martien, K.K., Taylor, B.L., 2009. Assessing statistical power of SNPs for population structure and conservation studies. Mol. Ecol. Resour. 9, 66–73, http://dx.doi.org/10.1111/j.1755-0998.2008.02392.x.

Olsen, M.T., Volny, V.H., Bérubé, M., Dietz, R., Lydersen, C., Kovacs, K.M., Dodd, R.S., Palsbøll, P.J., 2011. A simple route to single-nucleotide polymorphisms in a nonmodel species: identification and characterization of SNPs in the Artic ringed seal (Pusa hispida hispida). Mol. Ecol. Resour. 11 (Suppl. 1), 9–19, http://dx.doi.org/10.1111/j.1755-0998.2010.02941.x.

Pitra, C., Hansen, A.J., Lieckfeldt, D., Arctander, P., 2002. An exceptional case of historical outbreeding in African sable antelope populations. Mol. Ecol. 11, 1197–1208.

Ramos, A.M., Crooijmans, R.P.M., Affara, N., Amaral, A.J., Archibald, A.L., Beever, J.E., Bendixen, C., Churcher, C., Clark, R., Dehais, P., Hansen, M.S., Hedegaard, J., Hu, Z.-L., Kerstens, H.H., Law, A.S., Megens, H.-J., Milan, D., Nonneman, D.J., Rohrer, G., Rothschild, M.F., Smith, T.P.L., Schnabel, R.D., Van Tassell, C.P., Taylor, J.F., Wiedmann, R.T., Schook, L.B., Groenen, M.M., 2009. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS One 4, e6524, http://dx.doi.org/10.1371/journal.pone.0006524.

Robinson, T.J., Ropiquet, A., 2011. Examination of hemiplasy, homoplasy and phylogenetic discordance in chromosomal evolution of the Bovidae. Syst. Biol. 60, 439–450, http://dx.doi.org/10.1093/sysbio/syr045.

Rodwell, T.C., Whyte, I.J., Boyce, W.M., 2001. Evaluation of population effects of bovine tuberculosis in free-ranging african Buffalo (Syncerus caffer). J. Mammal. 82, 231–238, http://dx.doi.org/10.1644/1545-1542(2001)082<0231:EOPEOB>2.0.CO;2.

Rosenblum, E.B., Belfiore, N.M., Moritz, C., 2006. Anonymous nuclear markers for the eastern fence lizard Sceloporus undulatus. Mol. Ecol. Notes 7, 113–116, http://dx.doi.org/10.1111/j.1471-8286.2006.01547.x.

Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. 132, 365–386.

Ryman, N., Palm, S., André, C., Carvalho, G.R., Dahlgren, T.G., Jorde, P.E., Laikre, L., Larsson, L.C., Palmé, A., Ruzzante, D.E., 2006. Power for detecting genetic divergence: differences between statistical methods and marker loci. Mol. Ecol. 15, 2031–2045, http://dx.doi.org/10.1111/j.1365-294X.2006.02839.x.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467.

Santure, A.W., Stapley, J., Ball, A.D., Birkhead, T.R., Burke, T., Slate, J., 2010. On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. Mol. Ecol. 19, 1439–1451, http://dx.doi.org/10.1111/j.1365-294X.2010.04554.x.

Scarano, E., Iaccarino, M., Grippo, P., Parisi, E., 1967. The heterogeneity of thymine methyl group origin in DNA pyrimidine isostichs of developing sea urchin embryos. Proc. Natl. Acad. Sci. U. S. A. 57, 1394–1400.

Schlötterer, C., 2004. The evolution of molecular markers—just a matter of fashion? Nat. Rev. Genet. 5, 63–69.

Seeb, J.E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., Seeb, L.W., 2011. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in non model organisms. Mol. Ecol. Resour. 11 (Suppl. 1), 1–8, http://dx.doi.org/10.1111/j.1755-0998.2010.02979.x.

Smitz, N., Berthouly, C., Cornélis, D., Heller, R., Van Hooft, P., Chardonnet, P., Caron, A., Prins, H., van Vuuren, B.J., De Iongh, H., Michaux, J., 2013. Pan-African genetic structure in the African buffalo (Syncerus caffer): investigating intraspecific divergence. PLoS One 8, e56235, http://dx.doi.org/10.1371/journal.pone.0056235.

Smitz, N., Cornélis, D., Chardonnet, P., Caron, A., De Garine-wichatitsky, M., Jori, F., Mouton, A., Latinne, A., Pigneur, L., Melletti, M., Kanapeckas, K.L., Michaux, J., 2014. Genetic structure of fragmented southern populations of African Cape buffalo (Syncerus caffer caffer). BMC Evol. Biol., 1–19, http://dx.doi.org/10.1186/s12862-014-0203-2.

Stromberg, M.P., 2010. Enabling high-throughput sequencing data analysis with MOSAIK, PhD, Boston College. http://hdl.handle.net/2345/1332.

Van Tassell, C.P., Smith, T.P.L., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C., Sonstegard, T.S., 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat. Methods 5, 247–252, http://dx.doi.org/10.1038/NMETH.1185.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Van Bers, N.E.M., van Oers, K., Kerstens, H.H.D., Dibbits, B.W., Crooijmans, R.P.M., Visser, M.E., Groenen, M.A., 2010. Genome-wide SNP detection in the great tit Parus major using high throughput sequencing. Mol. Ecol. 19 (Suppl. 1), 89–99, http://dx.doi.org/10.1111/j.1365-294X.2009.04486.x.

Van Hooft, W.F., Hanotte, O., Wenink, P.W., Groen, A.F., Sugimoto, Y., Prins, H.H.T., Teale, A., 1999. Applicability of bovine microsatellite markers for population genetic studies on African buffalo (Syncerus caffer). Anim. Genet. 30, 214–220.

Van Hooft, W.F., Groen, A.F., Prins, H.H.T., 2002. Phylogeography of the African buffalo based on mitochondrial and Y-chromosomal loci: pleistocene origin and population expansion of the Cape buffalo subspecies. Mol. Ecol. 11, 267–279.

Vignal, A., Milan, D., SanCristobal, M., Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 34, 275–305, http://dx.doi.org/10.1051/gse.

Young, A.G., Clarke, G.M., 2000. Genetics, Demography and Viability of Fragmented Populations. Cambridge University Press, Cambridge.