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.

© 2016 Deutsche Gesellschaft für Säugetierkunde. Published by Elsevier GmbH. All rights reserved.

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/© 2016 Deutsche Gesellschaft für Säugetierkunde. Published by Elsevier GmbH. All rights reserved.

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

Download English Version:

https://daneshyari.com/en/article/2193279

Download Persian Version:

https://daneshyari.com/article/2193279

Daneshyari.com