Phylogenetic analysis and nucleotide diversity of 69 cattle breeds including German Black Pied cattle using WGS

G.B. Neumann[1*], P. Korkuć[1], D. Arends[1], M.J. Wolf[2], K. May[2], S. König[2], G.A. Brockmann[1]

[1] Humboldt University of Berlin, Albrecht Daniel Thaer- Institute for Agricultural and Horticultural Sciences, Invalidenstr. 42, 10115 Berlin, Germany; [2] Justus-Liebig-University of Gießen, Institute of Animal Breeding and Genetics, Ludwigstr. 21, 35390 Gießen, Germany; *guilherme.neumann@hu-berlin.de

Abstract

Local populations such as the endangered German Black Pied cattle (DSN) are an important genetic resource for the future of livestock breeding. To provide more information on the genome composition of DSN, we performed phylogenetic analysis and calculated genome-wide nucleotide diversity using variants from whole-genome sequencing of 302 DSN, and 1,394 individuals from 68 taurine breeds. Phylogenetic analysis detected four main clusters that could be assigned to their respective geographical origin: Northern Europe, Central Europe, Jersey and Guernsey islands, and an area comprising Eastern Europe, Central Italy, and Asia. DSN showed close relationship to Dutch Friesian Red, Dutch Belted, Holstein, Eastern Flanders White Red, and Kholmogory. Nucleotide diversity was medium in DSN (0.001528) while it was highest in breeds from Asia and Eastern Europe (>0.0017). Nucleotide diversity was particularly high on BTA 23 between 19 and 37 Mb where the bovine major histocompatibility complex and additional immune genes are located.

Introduction

Local indigenous cattle breeds are a crucial source of genetic diversity for livestock breeding, which might be important for future local adaptations. According to the Food and Agriculture Organization of the United Nations (FAO), 84% of all local breeds in Europe are considered at risk (FAO, 2021). This is also the case for the dual-purpose German Black Pied cattle population (DSN, "Deutsches Schwarzbuntes Niederungsrind"). Its initial breeding in the North Sea region of Germany and the Netherlands dates back to the 18th century. From there, animals were brought by emigrants to North America where they contributed to the breeding program responsible for the generation of the modern high-yielding dairy breed Holstein (Brade and Brade, 2013). Because of its lower milk yield, DSN has been almost entirely replaced by Holstein. Today, only about 2,500 DSN herd book cows remain. The current breeding goal for DSN aims to preserve DSN-typical and beneficial characteristics. For an effective conservation plan, genetic diversity measurements are needed to guarantee an optimised pool of genetic variants that provides sufficient adaptation capacity to changing environments and prevents inbreeding depression (Kristensen et al., 2015). In this study, we compared the genetic relationship of 69 taurine cattle breeds including DSN using phylogenetic analysis and nucleotide diversity (π). Nucleotide diversity has been shown to be a good start point when evaluating the genetic diversity status of a given population (Kardos et al., 2021). It is a region-wide metric used to quantify the degree of polymorphisms within a population.

Methods

Genomic data. Sequence variants were available from whole-genome sequencing data of 302 DSN animals (Neumann et al., 2021). Further sequence variants of 1,394 animals from 68 other *Bos taurus* breeds and 1 Auroch (*Bos primigenius*) were obtained from the 1000 Bull Genomes Project - Run 9 (Hayes and Daetwyler, 2018). The sequence variants were detected based on the ARS-UCD1.2 reference genome. For high data quality, we required a sequencing coverage of at least 8-fold and at least 5 animals per breed. For breeds with >30 animals, 30 animals were randomly selected. For Holstein, 30 animals per country (Germany, Denmark, Netherlands, U.S.) and 30 labelled as "Holstein Friesian", and not only "Holstein" in the 1000 Bull Genomes Project were randomly selected. The sequence variants from 302 sequenced DSN and from the 1000 Bull Genomes Project were merged using BCFtools v1.9. A total of 74,834,092 biallelic variants (68,486,380 SNPs and 6,347,712 indels) occurring in the tranche 99% (from the Variant Recalibration performed by 1000 Bulls Genome Project) with a call rate ≥90% were considered for the analysis.

Phylogenetic analysis. A genome-wide phylogenetic tree was built for all *Bos taurus* autosomes (BTA). The 74,834,092 sequence variants were first pruned removing all variants in high linkage ($r^2 > 0.6$) with PLINK v1.9 (Purcell et al. 2) using a windows-size of 50 bp and a step-size of 5 bp. Using the resulting 21,426,106 variants, the Manhattan distance between animals was calculated and the UPGMA algorithm implemented in the biotite library in Python was used for clustering. The tree was rooted using Auroch as an outgroup. The phylogenetic tree was visualized using iTOL v6 (Letunic et al. 2019). Branches of the phylogenetic tree represent the position of the majority of animals from each breed.

Genomic diversity. Nucleotide diversity (π) was calculated per windows of 10 kb (π_{window}) for all BTAs using 74,834,092 sequence variants and the package scikit-allel in Python as:

$$\pi_{window} = \frac{\frac{n}{n-1}\sum_{ij} p_i p_j \pi_{ij}}{10\,kb} \quad (1),$$

where p is the frequency of the respective i^{th} and j^{th} haplotypes, π_{ij} is the number of nucleotide differences per site between the i^{th} and j^{th} haplotypes, and n is the number of haplotypes. The average nucleotide diversity per chromosome (π_{chrom}) was calculated as the mean over all windows of the respective chromosomes, and total nucleotide diversity (π_{total}) as the mean across all BTAs.

Results

Phylogenetic analysis. Phylogenetic analysis of 69 cattle identified four main clusters that can be assigned to geographical origins (Figure 1a): Northern Europe (green), Central Europe (violet), Jersey and Guernsey islands (red), and Eastern Europe, Central Italy, and Asia (blue). The cluster of Central Europe comprised breeds from Austria, Switzerland, Southern Germany and France. Those are mainly dual-purpose breeds kept in mountainous areas. However, the majority of the breeds including DSN and Holstein formed a cluster based on their origin in Northern European countries. Jersey and Guernsey cattle formed a separate cluster. The fourth cluster included breeds from Asian, Central Italy, and Eastern European countries. Most breeds showed clear clusters without or with very few misassignments of animals. This was not the case for the Modern Danish

Red, Swedish Red, Norwegian Red, Ayrshire Finish, and Modern Angler which showed many misassignments. Regarding DSN, we observed that they clustered closest to Dutch Friesian Red, Dutch Belted, Holstein, Eastern Flanders White Red, and Kholmogory. Those are all breeds originating from Germany, the Netherlands, and Belgium. Kholmogory, a Russian breed, is descendent of animals imported from the Netherlands. Further breeds from the North Sea region, from territories of Germany, the Netherlands, Denmark, Belgium and Scandinavia showed also close relationship to DSN.

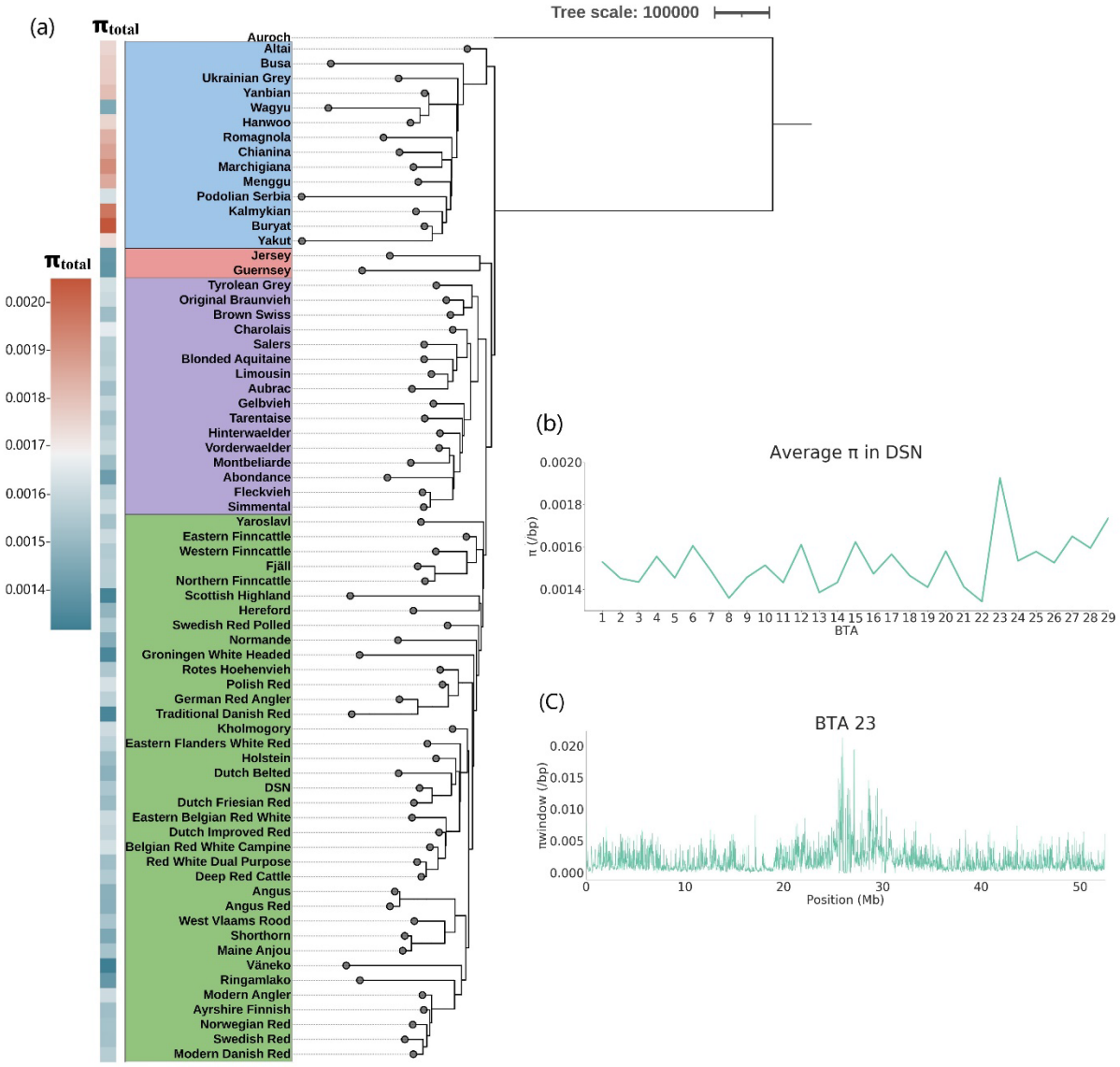

Figure 1: (a) Phylogenetic tree of 69 cattle breeds. Clusters are coloured by their geographical assignment: Central Italy, Eastern European and Asian countries (blue), Jersey and Guernsey (red), Central Europe (violet), Northern European countries (green). Individuals of each breed were collapsed to single clusters (grey spheres). Nucleotide diversity (π_{total}) is listed next to the leaves as gradient (blue=low, red=high). (b) Average nucleotide diversity in DSN per chromosome. (c) Nucleotide diversity in DSN per window on BTA 23.

Genomic diversity. Total nucleotide diversity (π_{total}) for the 69 investigated breeds ranged between 0.001316 in Väneko and 0.002048 in Buryat (Figure 1a). π_{total} was higher in DSN (0.001541) than

in Holstein (0.001505) and 28 other breeds (20 from Northern Europe), but lower than in other European and Asian breeds. The highest π_{total} values were detected in Asian and Eastern European breeds. The other European breeds showed similar π_{total} with a few lower exceptions (<0.0014): Ringamålako (0.001396), Jersey (0.001396), Guernsey (0.001385), Groningen White Headed (0.001345), Traditional Danish Red (0.001339), Scottish Highland (0.001328) and Väneko (0.001316). Regions with particular high nucleotide diversity (π_{window}) in DSN as well as in all other breeds were observed on BTA 4, 10, 12, 15, 29 (Figure 1b), and in particular on BTA23 between 19 and 37 Mb (Figure 1c), which includes the major histocompatibility complex (MHC) and the bovine leukocyte antigen (BoLA).

Discussion

DSN showed similar nucleotide diversity as the other breeds from the same cluster, indicating a good management of the endangered population. As expected, in DSN and all other breeds, BTA 23 contains a highly polymorphic region where the MHC is located. This high diversity around MHC is also seen in other species (Shiina et al., 2017). Asian and Eastern European breeds showed higher nucleotide diversity (π_{total}) likely due to less intensive breeding programs, genetic bottlenecks, or founder effects in comparison to Central and Northern European breeds (Felius et al., 2014). Further migration of these cattle to Southern, Central and finally Northern Europe are consistent with the clusters in our phylogenetic tree (Felius et al., 2014). The lowest nucleotide diversity of Northern European Red cattle populations was evident in Traditional Danish Red and Groningen White Headed, which resulted from recent inbreeding in the small populations (Schmidtmann et al., 2021). Recently, Schmidtmann et al. (2021) have found high levels of admixture between Modern Angler, Ayrshire Finnish, and Norwegian Red, which explains why our clustering between those breeds had many misassignments.

References

Brade, W. and Brade, E. (2013) Berichte über Landwirtschaft 91(2).

Food and Agriculture Organization of the United Nations (FAO). Risk status of livestock breeds. Available at: https://www.fao.org/sustainable-development-goals/indicators/252/en/

Felius, M., Beerling, M.L., Buchanan, D.S., Theunissen, B., et al. (2014) Diversity. 6(4):705-750 https://doi.org/10.3390/d6040705

Hayes, B.J., and Daetwyler, H.D. (2018) Annu. Rev. Anim. Biosci. 2019. 7:89–102 https://doi.org/10.1146/annurev-animal-020518

Kardos, M., Armstrong, E.E., Fitzpatrick, S.W., Hauser, S., Hedrick, P.W., et al. (2021) PNAS. 118(48). https://doi.org/10.1073/pnas.2104642118

Kristensen, T.N., Hoffmann, A.A., Pertoldi, C., Stronen, A. V. (2015) Front. Genet. 6(38) https://doi.org/10.3389/fgene.2015.00038

Letunic, I., Bork, P. (2019). Web Serv. issue Publ. 47. https://doi.org/10.1093/nar/gkz239

Neumann, G.B., Korkuć, P., Arends, D., Wolf, M.J., May, K., et al. (2021) BMC Genomics 22. https://doi.org/10.1186/S12864-021-08237-2

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., et al. (2007). Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795

Schmidtmann, C., Schönherz, A., Guldbrandtsen, B., Marjanovic, J., Calus, M.,et al. (2021) GSE. 53. https://doi.org/10.1186/s12711-021-00613-6

Shiina, T., Blancher, A., Inoko, H., Kulski, J.K., 2017. Immunology 150, 127–138. https://doi.org/10.1111/IMM.12624