

DATA NOTE

Open Access



# Data of whole-genome sequencing of Karakul, Zel, and Kermani sheep breeds

Leila Mohammadipour Saadatabadi<sup>1</sup>, Mohammadreza Mohammadabadi<sup>1\*</sup>, Zeinab Amiri Ghanatsaman<sup>2</sup>, Olena Babenko<sup>3</sup>, Ruslana Volodymyrivna Stavetska<sup>3</sup>, Oleksandr Mikolayovich Kalashnik<sup>4</sup>, Volodymyr Afanasenko<sup>5</sup>, Oleksandr Anatoliiovych Kochuk-Yashchenko<sup>6</sup>, Dmytro Mykolaiovych Kucher<sup>6</sup> and Hojjat Asadollahpour Nanaei<sup>7</sup>

## Abstract

**Objective** The data provided herein represent the whole-genome sequencing data associated with three sheep breeds of Iranian native breeds. Sheep are the first domesticated animals that, during the long path of the evolution process, have created gene variants with desirable phenotypic effects, so they can be suitable models for biomedical research. In addition, sheep have a vital role in providing protein to a notable part of the human population around the world.

**Data description** Ten blood samples were taken from three Iranian native sheep breeds, the Zel, Karakul, and Kermani kinds. Blood samples genomes were extracted using the salting-out technique. The Illumina NovaSeq 6000 platform was used to carry out sequencing of the whole genome in a laboratory in China. All sequence information is available through the NCBI database in the sequence read archive (SRA) format under the accession number PRJNA904537. The dataset presented here can provide a useful resource for genome analysis of livestock breeds adapted to hot and dry regions.

**Keywords** Whole-genome sequencing, Sheep, Iran

## Objective

One of the first animals to be domesticated was the sheep. They most likely tamed Asian Mouflons (*Ovis Orientalis*) around 11,000 years ago (BP), possibly into the Zagros Mountains and/or southeast Anatolia, in the Fertile Crescent [1]. Sheep are farmed all over the world and contribute considerably to the production of animal-based protein used in human nutrition. They have an important contribution to the agricultural economy too. In the western region of Asia, the country of Iran, with more than 52 million sheep classified into 27 different ecotypes, can be one of the significant regions of genetic reserves in sheep. Iran is a hot and dry country due to its location on the Africa-Asia desert cordon, so about 90% of its regions are dry and water-scarce [2, 3]. Conserving variety is critical for increasing production efficiency

\*Correspondence:

Mohammadreza Mohammadabadi  
mrm@uk.ac.ir

<sup>1</sup>Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran

<sup>2</sup>Department of Animal Science, Fars Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education & Extension Organization (AREEO), Fars, Iran

<sup>3</sup>Bila Tserkva National Agrarian University, Bila Tserkva, Ukraine

<sup>4</sup>Sumy National Agrarian University, Sumy, Ukraine

<sup>5</sup>National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine

<sup>6</sup>Department of Breeding, Animal Genetics and Biotechnology, Polissia National University, Zhytomyr, Ukraine

<sup>7</sup>Reproductive Biotechnology Research Center, Avicenna Research Institute, ACECR, Tehran 1983969412, Iran



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

and boosting adaptation to climate-changing conditions. The discovery of related genes, as well as the detection of effective mechanisms for responding to heat stress and immunological reactions, can increase livestock products and also help to maintain genetic diversity in these areas [4]. The data set presented here can provide useful and practical resources for studies related to animal species suitable for arid environments as well as for researching more genetic analysis-related hypotheses.

### Data description

Blood was collected from three breeds of sheep native to Iran, including Karakul (4 samples), Zel (4 samples), and Kermani (2 samples) (from each of the samples, 5 ml of jugular vein blood). Karakul sheep, one of leather breeds, are native to the North Khorasan province of Iran. This province is located in the northeast of Iran and is 300 m above sea level (AMSL: 300 m). The prevailing climate of this region is hot and dry due to its proximity to the central desert of Iran. The Zel breed, tailed sheep, is mostly bred in Mazandaran province, Iran. This province is located in the north of Iran, near the Caspian Sea, and is 2 m above sea level (AMSL: 2 m). The climate of this region is divided into two types: humid and mountainous, due to the presence of the sea, mountains, and forests. Kermani sheep are bred in the southeastern parts of Iran, especially in Kerman province. This wool breed is fully compatible with the hot and dry climate of this province, which has an altitude of 1755 m above sea level (AMSL: 1755 m).

Using the salting-aut technique, the whole genome was obtained from blood samples. Then genome sequencing was done using the Illumina NovaSeq 6000 platform in China. The FastQC program was used to assess the quality of all genomic data. The samples were aligned using the Burrows-Wheeler Aligner (BWA Mem Version 0.7.10) to the sheep genome reference ([https://www.ncbi.nlm.nih.gov/assembly/GCF\\_00274215.1/](https://www.ncbi.nlm.nih.gov/assembly/GCF_00274215.1/)) [5].

SAM (.sam) and BAM (.bam) files were created using the SAMtools program, and SAMtools was also used for reading files, sorting them, and indexing them [6]. To limit the probability of false-positive variant calling, using the Picard toolkit, potential PCR duplicates were eliminated (<http://broadinstitute.github.io/picard>). To improve alignment accuracy, base quality score recalibration (BQSR) and local realignment around indels were performed using tools from the Genome Analysis Toolkit (GATK) [7]. Final variants (SNPs, single nucleotide polymorphisms) were called and filtered using the GATK program. We analyzed indigenous Iranian sheep's genetic information using fixation index and nucleotide diversity ( $\theta\pi$ ) statistical assessments for the detection of probable genes associated with heat adaptation and immunological response, as well as to compare the genetic structure of indigenous and non-indigenous sheep populations. Our findings may help comprehend the molecular mechanisms of heat and dry climate adaptation in small ruminants [8]. The whole-genome sequencing data described in the current paper has been uploaded to the NCBI database in SRA (sequence read archive) format (<https://identifiers.org/ncbi/bioproject:PRJNA904537>) with the accession number PRJNA904537. For more information and data connections, please check Table 1 and the references [9–19].

Whole-genome sequencing data was uploaded to the NCBI SRA Database with the accession number PRJNA904537. For the three different breeds of Karakul (4 individuals), Zel (4 individuals), and Kermani (2 individuals), a total of ten whole genome sequencing files were generated. This table displays the link for the bio-project, in addition to links for each sheep.

### Limitations

The lack of a reference genome from Iranian sheep during alignment is one of the limitations of our study and similar studies. In generating this data, short sequences and

**Table 1** Summary of the whole-genome sequencing data of ten Iranian sheep

lable	Name of data file/data set	File type (file extension)	Data repository and identifier (DOI or accession number)
Bioproject	Whole-genome sequencing of three breeds of Iranian sheep		<a href="https://identifiers.org/ncbi/bioproject:PRJNA904537">https://identifiers.org/ncbi/bioproject:PRJNA904537</a> . [9]. NCBI BioProject
Data set 1	KARAKUL_a11	Fastq (fq.gz)	<a href="https://identifiers.org/ncbi/insdc.sra:SRR25570510">https://identifiers.org/ncbi/insdc.sra:SRR25570510</a> . [10]. NCBI SRA Database
Data set 1	KARAKUL_a12		<a href="https://identifiers.org/ncbi/insdc.sra:SRR25570509">https://identifiers.org/ncbi/insdc.sra:SRR25570509</a> [11]. NCBI SRA Database
Data set 1	KARAKUL_a13		<a href="https://identifiers.org/ncbi/insdc.sra:SRR25570508">https://identifiers.org/ncbi/insdc.sra:SRR25570508</a> [12]. NCBI SRA Database
Data set 1	KARAKUL_a14		<a href="https://identifiers.org/ncbi/insdc.sra:SRR25570507">https://identifiers.org/ncbi/insdc.sra:SRR25570507</a> [13]. NCBI SRA Database
Data set 2	ZEL_a1	Fastq (fq.gz)	<a href="https://identifiers.org/ncbi/insdc.sra:SRR25570514">https://identifiers.org/ncbi/insdc.sra:SRR25570514</a> [14]. NCBI SRA Database
Data set 2	ZEL_a2		<a href="https://identifiers.org/ncbi/insdc.sra:SRR25570513">https://identifiers.org/ncbi/insdc.sra:SRR25570513</a> [15]. NCBI SRA Database
Data set 2	ZEL_a3		<a href="https://identifiers.org/ncbi/insdc.sra:SRR25570512">https://identifiers.org/ncbi/insdc.sra:SRR25570512</a> [16]. NCBI SRA Database
Data set 2	ZEL_a4		<a href="https://identifiers.org/ncbi/insdc.sra:SRR25570511">https://identifiers.org/ncbi/insdc.sra:SRR25570511</a> [17]. NCBI SRA Database
Data set 3	KRM_a1	Fastq (fq.gz)	<a href="https://identifiers.org/ncbi/insdc.sra:SRR25570516">https://identifiers.org/ncbi/insdc.sra:SRR25570516</a> [18]. NCBI SRA Database
Data set 3	KRM_a2		<a href="https://identifiers.org/ncbi/insdc.sra:SRR25570515">https://identifiers.org/ncbi/insdc.sra:SRR25570515</a> [19]. NCBI SRA Database

the Illumina approach were used. But using the emerging long-read sequencing (LRS) technologies can be used to improve the quality of sequencing and increase the accuracy genome evaluation studies.

#### Abbreviations

AMSL	Above mean sea level
BAM	Binary alignment map
BWA	Burrows wheeler aligner
GATK	Genome analysis toolkit
Θπ	Nucleotide diversity
GCTA	Genome-wide complex trait analysis
NCBI	National Center for Biotechnology Information
SNP	Single-nucleotide polymorphism
SRA	Sequence Read Archive

#### Acknowledgements

The authors of the paper thank all the personnel of Karakul Sarakhs Sheep Breeding Station in North Khorasan, Iran, and the Livestock Gene Bank in Babol (Zel Breeding Station), the north of Iran, as well as the staff of the University of Kerman Shahid Bahonar, Kerman, Iran.

#### Authors' contributions

MM and HAN conceived the study. LMS, OB, and RVS performed the sampling and carried out the DNA extraction. HAN, ZAG, OMK, VA, and OAK-Y generated and evaluated the genome resequencing data. The manuscript was prepared by LMS, MM, and DMK. The final manuscript was reviewed and approved by all authors.

#### Funding

The Vice Chancellor for Research and Technology of Kerman's Shahid Bahonar University provided funding for this project (Grant number: G-311/8720). The study's design, data collection, analysis, interpretation, and paper preparation were all supported by the funding bodies.

#### Availability of resources and data

The whole-genome sequencing data described here has been uploaded to the NCBI database in sequence read archive (SRA) format with the accession number PRJNA904537 (<https://identifiers.org/ncbi/bioproject:PRJNA904537>). For more information and connections to the data, please refer to Table 1 and the references [9–19].

#### Declarations

##### Ethics approval and consent to participate

The ARRIVE guidelines 2.0 were followed for conducting this study (<https://arriveguidelines.org/>). The animal science ethics council of Shahid Bahonar University in Kerman, Iran, gave its approval to all experimental protocols and blood collection methods (No. 96/47561, dated 23 September 2018). There were neither dead animals nor injured ones. The applicable rules and regulations of the Livestock Gene Bank at Babol, Sheep breeding facilities Karakul in Sarakhs, and Kerman's Shahid Bahonar University in Kerman, Iran, were adhered to in this study. All methods were performed in accordance with the relevant guidelines and regulations.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

Received: 6 October 2023 / Accepted: 20 November 2023

Published online: 27 November 2023

#### References

1. Zeder MA. Animal domestication in the Zagros: an update and directions for future research. *MOM Ed.* 2008;49:243–77.
2. Pourkhorsandi H, Gattacceca J, Rochette P, d'Orazio M, Kamali H, de Aville R, Letichevsky S, Djamali M, Mirnejad H, Debaille V, Jull AT. Meteorites from the Lut Desert (Iran). *Meteorit Planet Sci.* 2019;54(8):1737–63.
3. Nouri M, Homae M. Drought trend, frequency and extremity across a wide range of climates over Iran. *Meteorol Appl.* 2020;27(2):e1899.
4. Mohamadipoor Saadatabadi L, Mohammadabadi M, Amiri Ghanatsaman Z, Babenko O, Stavetska R, Kalashnik O, Kucher D, Kochuk-Yashchenko O, Asadollahpour Nanaei H. Signature selection analysis reveals candidate genes associated with production traits in Iranian sheep breeds. *BMC Vet Res.* 2021;17(1):1–9.
5. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform bioinformatics. 2009;25(14):1754–60.
6. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9. 1000 Genome Project Data Processing Subgroup.
7. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–303.
8. Saadatabadi LM, Mohammadabadi M, Nanaei HA, Ghanatsaman ZA, Stavetska RV, Kalashnyk O, Kochuk-Yashchenko OA, Kucher DM. Unraveling candidate genes related to heat tolerance and immune response traits in some native sheep using whole genome sequencing data. *Small Ruminant Research.* 2023 Jun;12:107018.
9. NCBI Bioproject. <https://identifiers.org/ncbi/bioproject:PRJNA904537>. (2023).
10. NCBI SRA Database. <https://identifiers.org/ncbi/insdc.sra:SRR25570510>. (2023).
11. NCBI SRA Database. <https://identifiers.org/ncbi/insdc.sra:SRR25570509>. (2023).
12. NCBI SRA Database. <https://identifiers.org/ncbi/insdc.sra:SRR25570508>. (2023).
13. NCBI SRA Database. <https://identifiers.org/ncbi/insdc.sra:SRR25570507>. (2023).
14. NCBI SRA Database. <https://identifiers.org/ncbi/insdc.sra:SRR25570514>. (2023).
15. NCBI SRA Database. <https://identifiers.org/ncbi/insdc.sra:SRR25570513>. (2023).
16. NCBI SRA Database. <https://identifiers.org/ncbi/insdc.sra:SRR25570512>. (2023).
17. NCBI SRA Database. <https://identifiers.org/ncbi/insdc.sra:SRR25570511>. (2023).
18. NCBI SRA Database. <https://identifiers.org/ncbi/insdc.sra:SRR25570516>. (2023).
19. NCBI SRA Database. <https://identifiers.org/ncbi/insdc.sra:SRR25570515>. (2023).

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.