A Big Earth Data Platform for Three Poles

**Mitochondrial genome data of populations in Hindu Kush mountains (2020-2021)**

1、Description

The Hindu Kush mountains are tall mountains in Central Asia. The average altitude is about 5000 meters, and the highest peak, tiriqmir peak, is 7690 meters above sea level. It is the watershed between the Indus River in the southwest of the Qinghai Tibet Plateau and the Amu Darya River in the Pamir Plateau. Studying the genetic structure of the population in this area will help to deeply understand the diffusion history of the population and the genetic basis of adapting to the plateau environment. Studying the genetic structure of the population in this area will help to deeply understand the diffusion history of the population and the genetic basis of adapting to the plateau environment. In this study, we studied the maternal genetic structure of 213 individuals from 5 populations distributed in Xingdu Kush mountains. The mitochondrial genome sequences of 213 individuals (average depth > 1000x) were obtained by mitochondrial genome capture library construction and second-generation sequencing (Illumina hiseq x ten platform). Based on the idea of phylogeny, we control the quality of these data to ensure that there are no quality problems such as sample pollution. The modified Cambridge standard sequence was used as a reference for the output of mutation sites. According to the mitochondrial DNA phylogenetic tree (phylotree. ORG) of people around the world, each sample was divided into haplotype groups. Based on the mtDNA data of modern population and ancient samples published in the past, the origin and diffusion history of population in this area were systematically studied. The results show that the main maternal genetic components of Indo European populations in this region come from western Eurasia. Haplotypes w3a1a and j1b1a1 may be related to the migration of Indo European language groups, indicating that the diffusion of Indo European language groups to South Asia may not only be a simple cultural expansion, but also accompanied by the migration of some people. In addition, the study also shows that northern Pakistan may play an important channel role in the diffusion of Indo European languages to South Asia.

2、Keywords

Theme：Biological Resources,Population
Discipline：Human-nature Relationship
Places：Hindu kush Mountains, Northern Pakistan
Time：2020-2021

3、Data details

1.Scale：None

2.Projection：

3.Filesize：3.43MB

4.Data format：None

4、Space scope

|  |  |  |
| --- | --- | --- |
| - | north：36.0 | - |
| west：71.37 | - | east：74.84 |
| - | south：34.0 | - |

5、Time frame:2019-12-31 16:00:00+00:00--2021-12-31 03:59:59+00:00

6、Reference method

References to data:

KONG Qingpeng. Mitochondrial genome data of populations in Hindu Kush mountains (2020-2021). A Big Earth Data Platform for Three Poles, doi:10.11888/HumanNat.tpdc.2718702021

References to articles:

Rahman, Z.U., Tian, J.Y., & Gao, Z.L. et al. (2021). Complete mitogenomes document substantial genetic contribution from the Eurasian Steppe into northern Pakistani Indo-Iranian speakers. Eur J Hum Genet 29, 1008–1018. https://doi.org/10.1038/s41431-021-00829-6

7、Supporting project information

Pan-Third Pole Environment Study for a Green Silk Road-A CAS Strategic Priority A Program

8、Data resource provider

name: KONG Qingpeng
unit: Kunming Institute of Zoology, Chinese Academy of Sciences
email: kongqp@mail.kiz.ac.cn