时空三极环境大数据平台

**兴都库什山脉人群的线粒体基因组数据（2020-2021）**

英文标题：Mitochondrial genome data of populations in Hindu Kush mountains (2020-2021)

1、摘要

兴都库什山脉（Hindu kush Mountains）是亚洲中部的高大山脉。平均海拔约5000米，最高峰蒂里奇米尔峰海拔7690米，是发源于青藏高原西南部的印度河和发源于帕米尔高原的阿姆河的分水岭。研究该区域人群的遗传结构，有助于深入理解该地区人群的扩散历史和适应高原环境的遗传基础。在本研究中，我们对分布于兴都库什山脉的5个群体的213个个体进行了母系遗传结构的研究。采用线粒体基因组捕获建库和二代测序（Illumina HiSeq X Ten platform）的方法，获得了213个个体的线粒体全基因组序列（平均深度>1000x）。基于系统发育思想，我们对这些数据进行质量控制，确保没有样本污染等质量问题。以修订后的剑桥标准序列参考，进行突变位点的输出。根据世界范围内的人群的线粒体DNA系统发育树（PhyloTree.org），对每个样本进行单倍型类群划分。综合以往发表的现代人群和古代样本的mtDNA数据，系统研究了该地区人群的起源和扩散历史。结果表明，该地区的印欧语系人群的主要母系遗传组分来源于欧亚西部。单倍群W3a1a和J1b1a1等类群可能与印欧语系人群的迁徙有关，说明印欧语系向南亚的扩散可能不仅仅是简单的文化扩张，也伴随着一部分的人群迁徙。此外，研究还表明，巴基斯坦北部可能在印欧语系向南亚的扩散中起着重要的通道作用。

2、关键词

主题关键词：生物资源,人口
学科关键词：人地关系
地点关键词：兴都库什山脉, 巴基斯坦北部
时间关键词：2020-2021

3、数据细节

1.比例尺：None

2.投影：

3.文件大小：3.43MB

4.数据格式：None

4、空间范围

|  |  |  |
| --- | --- | --- |
| - | 北：36.0 | - |
| 西：71.37 | - | 东：74.84 |
| - | 南：34.0 | - |

5、时间范围2019-12-31 16:00:00+00:00--2021-12-31 03:59:59+00:00

6、引用方式

数据的引用:

孔庆鹏. 兴都库什山脉人群的线粒体基因组数据（2020-2021）. 时空三极环境大数据平台, DOI:10.11888/HumanNat.tpdc.271870, CSTR:18406.11.HumanNat.tpdc.271870, 2021.[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.271870, CSTR:18406.11.HumanNat.tpdc.271870, 2021]

文章的引用:

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、资助项目信息

泛第三极环境变化与绿色丝绸之路建设专项(XDA20000000)

8、数据资源提供者

姓名: 孔庆鹏
单位: 中国科学院昆明动物研究所
电子邮件: kongqp@mail.kiz.ac.cn