

Tracing the Genetic Legacy of the Tibetan Empire in the Balti

Xing-Yan Yang,^{†,1,2} Allah Rakha,^{†,3,4} Wei Chen,^{5,6} Juzhi Hou,⁷ Xue-Bin Qi,¹ Quan-Kuan Shen,^{1,8} Shan-Shan Dai,^{1,8} Xierzhatijiang Sulaiman,¹ Najmudinov Tojiddin Abdulloevich,⁹ Manilova Elena Afanasevna,⁹ Khudoidodov Behruz Ibrohimovich,⁹ Xi Chen,^{10,11} Wei-Kang Yang,^{10,11} Atif Adnan,¹² Ruo-Han Zhao,¹ Yong-Gang Yao,^{4,8,13} Bing Su,^{1,8,14} Min-Sheng Peng,^{*,1,8,13} and Ya-Ping Zhang^{*,1,2,8,13,14}

¹State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

²State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, School of Life Sciences, Yunnan University, Kunming, China

³Department of Forensic Sciences, University of Health Sciences, Lahore, Pakistan

⁴Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

⁵College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming, China

⁶State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming, China

⁷Key Laboratory of Alpine Ecology (LAE), Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China

⁸Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, China

⁹E.N. Pavlovsky Institute of Zoology and Parasitology, Academy of Sciences of Republic of Tajikistan, Dushanbe, Tajikistan

¹⁰Research Center for Ecology and Environment of Central Asia, Chinese Academy of Sciences, Urumqi, China

¹¹Key Laboratory of Biogeography and Bioresource in Arid Land, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China

¹²Department of Human Anatomy, School of Basic Medicine, China Medical University, Shenyang, China

¹³KIZ-CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

¹⁴Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China

[†]These authors contributed equally to this work.

*Corresponding authors: E-mails: zhangyp@mail.kiz.ac.cn; pengminsheng@mail.kiz.ac.cn.

Associate editor: Evelyne Heyer

Abstract

The rise and expansion of Tibetan Empire in the 7th to 9th centuries AD affected the course of history across East Eurasia, but the genetic impact of Tibetans on surrounding populations remains undefined. We sequenced 60 genomes for four populations from Pakistan and Tajikistan to explore their demographic history. We showed that the genomes of Balti people from Baltistan comprised 22.6–26% Tibetan ancestry. We inferred a single admixture event and dated it to about 39–21 generations ago, a period that postdated the conquest of Baltistan by the ancient Tibetan Empire. The analyses of mitochondrial DNA, Y, and X chromosome data indicated that both ancient Tibetan males and females were involved in the male-biased dispersal. Given the fact that the Balti people adopted Tibetan language and culture in history, our study suggested the impact of Tibetan Empire on Baltistan involved dominant cultural and minor demic diffusion.

Key words: Balti, Tibetan, admixture, genome, mtDNA, X chromosome.

Since 7th century AD, the arising and expansion of Tibetan Empire established by the ancient Tibetan people is regarded as a high point in ancient civilization of the Qinghai-Tibet Plateau (Uray 1979; Beckwith 1993). By conducting series of military campaigns, the Tibetan Empire controlled territory of 4 million km², spanning parts of East Asia, Central Asia, and South Asia (fig. 1). More than 10 million people were under rule (Denwood 2008). The historical and archaeological records indicated massive population migrations along with series of military conquests by the ancient Tibetan Empire

(Ming 2007; Denwood 2008), which could leave a profound legacy on the cultural and genetic diversity of human populations (Beckwith 1993; Walter 2009; Owen-Smith and Hill 2013). Nevertheless, little is known about the genetics of this cultural and military expansion of ancient Tibetans.

The Balti people of Baltistan in Pakistan, an ethnic group proposed to be admixed of Tibetan immigrants with Indo-European (especially, Dardic) descents (Bakshi 1997), is a good model to trace the impact of ancient gene flow of Tibetans. There are two proposed sources for Tibetan contribution to

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

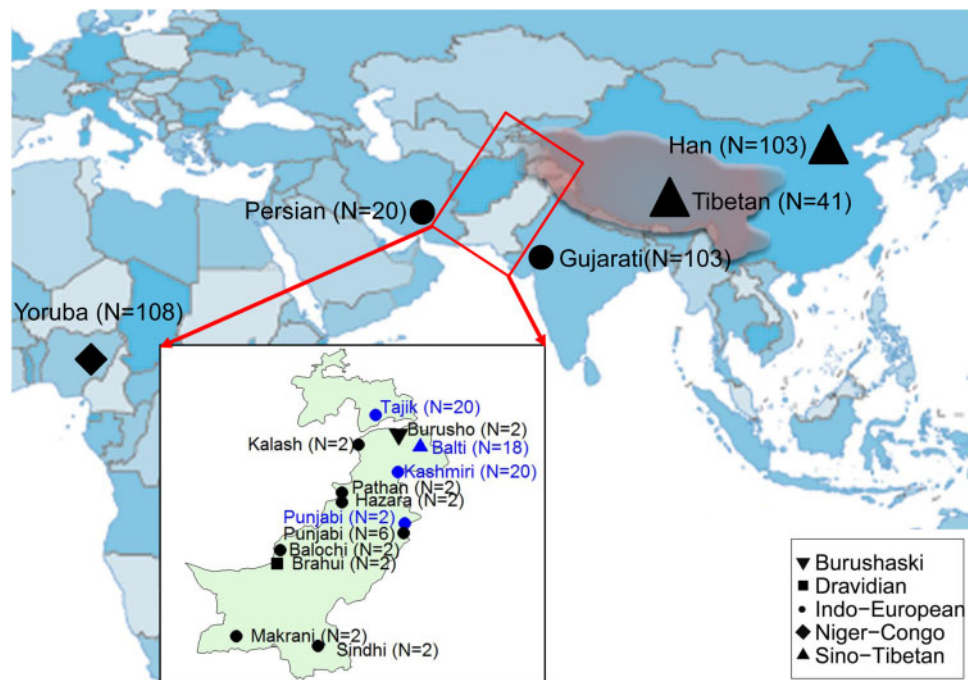


Fig. 1. Whole-genome sequencing of four ethnic populations in Pakistan and Tajikistan. The populations sequenced in this study are labeled blue and previously published data are labeled in black. The WGS data set includes 455 individuals. The shape of the labels indicated the language family used in each population. The territory of the Tibetan Empire in its prime time during AD 780–800 is marked in shallow red. The detailed information is described in [supplementary tables S1 and S2, Supplementary Material online](#).

Balti: the immigrants from Zhangzhung (i.e., Shangshung) which was an ancient kingdom locating in western and northwestern Tibet of China during 500 BC to AD 625 (Aldenderfer and Yinong 2004), and/or the immigrants as recent military conquerors after the 7th century AD due to the expansion of Tibetan Empire (Uray 1979). In contrast to historical evidence, previous Y-chromosomal analyses revealed that the Balti individuals resembled their Indo-European speaking neighbors in northern Pakistan and no paternal gene flow from Tibetans was detected (Qamar et al. 2002). Recently, population genomic analysis has been proved as a powerful approach to explore human admixture history (Hellenthal et al. 2014; Brucato et al. 2017; Fernandes et al. 2019). However, the Balti and its neighboring populations were underrepresented in population genomic researches. Thus, the origin of Balti people is still opaque. In this study, we aimed to dissect the recent admixture derived from Tibetans into the Balti people during the historic period by analyzing whole-genome sequencing (WGS) data of the Balti and its neighboring populations.

Population Genomic Variation

We contextualized our newly generated data of 60 genomes including 18 Baltis from Gilgit-Baltistan, 20 Kashmiris from Azad Jammu and Kashmir, two Punjabis from Punjab in Pakistan, and 20 Pamiri Tajiks from Gorno-Badakhshan Autonomous Region in Tajikistan (fig. 1 and [supplementary table S1, Supplementary Material online](#)), with 395 published genomes from 14 modern populations ([supplementary table S2, Supplementary Material online](#)) (1000 Genomes Project

[Consortium. 2015; Lu et al. 2016; Mallick et al. 2016; Yang et al. 2018; Charati et al. 2019](#)). The details for sampling, sequencing, SNP calling, data merging, and filtering were described in supplementary text, [Supplementary Material online](#). A total of 4,252,528 autosomal SNPs for 455 individuals from 17 populations ([supplementary table S3, Supplementary Material online](#)) were used in the subsequent analyses.

To check the batch effect of joint SNP calling as well as potential bias in reference population sampling and data merging, we independently performed single-sample SNP calling for each of newly generated genomes as described before (Mallick et al. 2016). We then merged them with 753 high-depth sequencing genomes of globally diverse populations (Mallick et al. 2016; Bergström et al. 2020), to get a total of 61,126,197 autosomal SNPs for 879 individuals ([supplementary table S3, Supplementary Material online](#)). We applied analyses of the additional data set to validate the results of population structure and genetic admixture based on the WGS data of 455 individuals.

Population Genetic Structure

We applied principal component analysis (PCA) (Patterson et al. 2006) and ADMIXTURE (Alexander et al. 2009) based on the pruned WGS data of 455 individuals ([supplementary table S3, Supplementary Material online](#)) to discern the clustering pattern of the Balti to other populations and potential admixture in this population, respectively. The PCA under the continental level showed the separation between African (i.e., Yoruba) and other no-African populations in PC1, whereas the Balti individuals projected near other South Asians

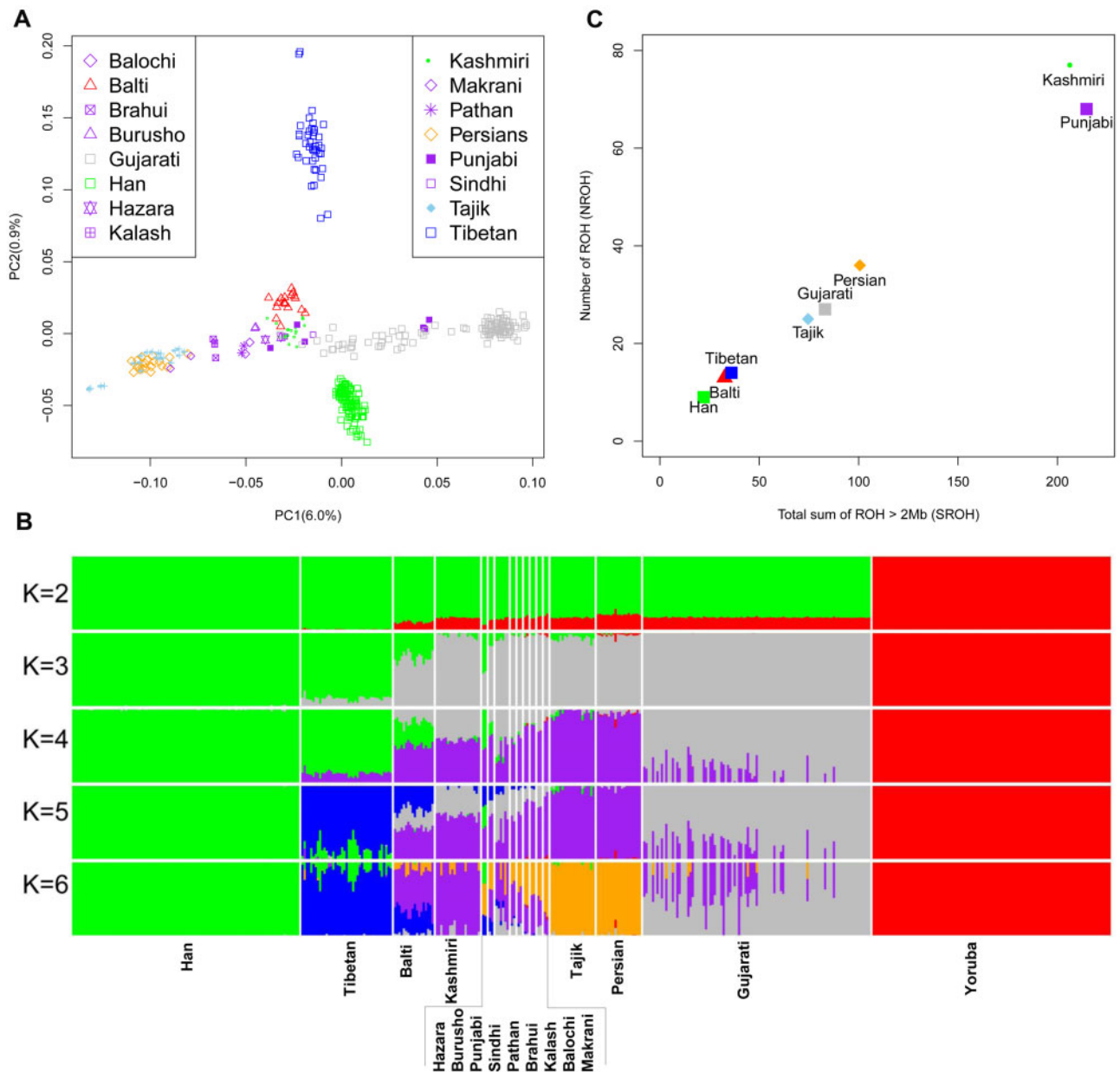


Fig. 2. Population structure of the Balti and other populations. (A) PCA of the individuals from 16 Asian populations. (B) ADMIXTURE analyses for K values of 2–6 showing different contributions of ancestry. The analyses were based on the 472,755 SNPs in 17 populations after pruning of LD. The details for populations and data sets were described in [supplementary tables S2 and S3, Supplementary Material](#) online. (C) The mean sum total length of ROH (SROH) and number of ROH (NROH) in the Balti and other neighboring populations based on 5,408,870 SNPs.

([supplementary fig. S1, Supplementary Material](#) online). The PCA on a finer scale revealed that the Balti clustered with the regionally close Kashmiri and differentiated from Tibetan and Han populations ([fig. 2A](#)). The pattern could be well replicated in ADMIXTURE. Overall, when $K = 3$ ([supplementary fig. S2A, Supplementary Material](#) online), the genetic makeup of Balti population resembles that of East (“green”) and South/Central/West (“gray”) Asians ([fig. 2B](#)). Starting from $K = 4$, ADMIXTURE identified a West and Central Asian genetic component (“purple”), which was distinguished from the South Asian component (“gray”). The Tibetan-specific component (“blue”) in the Balti, Burusho, Hazara, and Punjabi populations was also well visible at $K = 5$, which

has the best cross-validation index among all tests. The Iranian language-speakers’ component (“orange”) was inferred at $K = 6$. We then repeated PCA and ADMIXTURE in the additional data sets ([supplementary tables S2 and S3, Supplementary Material](#) online). All the results ([supplementary figs. S3 and S4, Supplementary Material](#) online) were supported by the patterns observed in WGS data of 455 individuals ([fig. 2B](#)).

We next used PLINK 1.9 ([Chang et al. 2015](#)) to scan the genomes for short (2–8 Mb) runs of homozygosity (ROHs) ([Ceballos et al. 2018](#)) to learn more about admixture as well as endogamy in the Balti population. The Kashmiri and Punjabi populations from Pakistan presented the highest mean

number and sum of ROHs (fig. 2C), most likely due to a high level of consanguinity (Jabeen and Malik 2014). Similar to the Tibetan and Han populations, the Balti population carried low mean number and sum of ROH. The distribution of ROHs in the Persian, Gujarati, and Tajik populations was between the groups of Balti–Tibetan–Han and Kashmiri–Punjabi.

Detection of Gene Flow

We investigated the gene flow between the Balti and nearby populations using the following approaches: f_3 and f_4 (Patterson et al. 2012), TreeMix (Pickrell and Pritchard 2012), and identical by descent (IBD) sharing statistics (Bai et al. 2018). By using f_3 (Balti; South/Central/West Asian, X), we found significant negative scores ($Z \leq -3$) where X represented Tibetan, Han, or Hazara (supplementary fig. S5A, Supplementary Material online). As Hazara has been said to receive gene flow from East Asian (Lawson et al. 2012), we then explored f_3 (Balti; Tibetan/Han, South/Central/West Asian) to show potential gene flow from East Asian (i.e., Tibetan or Han) to the Balti population (supplementary fig. S6A, Supplementary Material online). This process was also replicated by the f_4 statistics (supplementary figs. S5B and S6B, Supplementary Material online). Referring to the context of population tree, we conducted TreeMix to show potential gene flow from Tibetan rather than Han to the Balti population (fig. 3A and supplementary fig. S7, Supplementary Material online). The Tibetan IBD segments (≥ 2 cM) shared in the Balti population is higher ($W = 0.28$) than that in Han Chinese ($W = 0.19$), whereas the pattern is not detected in the other neighboring populations (fig. 3B). The Han Chinese IBD segments (≥ 2 cM) were scarcely shared in the Balti population (fig. 3C). Therefore, Tibetan is proposed as one of candidate sources to the Balti population.

Inference of Admixture Scenarios

Referring to the reported strategy (Busby et al. 2015), we first used CHROMOPAINTER and fineSTRUCTURE (Lawson et al. 2012) to characterize haplotype similarity and population assignment for the phased 455 genomes. The Balti, Tajik, Kashmiri, Tibetan, and Han individuals were grouped according to their population labels (supplementary fig. S8, Supplementary Material online). Some individuals from Punjabi, Persian, and Gujarati were clustered with other ethnic individuals, indicating the intricate genetic structure and admixture in those populations (Metspalu et al. 2011; Narasimhan et al. 2019). Excluding Hazara and Burusho with East Asian ancestry (Lawson et al. 2012), we assigned 14 populations as donors for the Balti population as receipt and re-ran CHROMOPAINTER. We then applied SOURCEFIND (Chacon-Duque et al. 2018) to characterize the Tibetan and Kashmiri ancestry components in the Balti population (supplementary fig. S9, Supplementary Material online). In addition, the Han ancestry component was detected in some runs. It was likely ascribed to the gene flows between Han and Tibetan populations after their divergence (Zhang et al. 2017). We then adopted GLOBETROTTER

(Hellenthal et al. 2014) to infer the admixture scenario for the Balti population (supplementary table S4, Supplementary Material online). Our results showed that the Balti population originated from one-way admixture of the Tibetan and Kashmiri populations around 29 generations ago (fig. 3D and supplementary table S4, Supplementary Material online). The genetic contribution from the Tibetan to the Balti population was $\sim 26\%$ (95% CI: 24.7–27.3%). Furthermore, we assigned 47 populations (especially including 15 populations from China and ten from Pakistan; supplementary table S2, Supplementary Material online) as donors and applied CHROMOPAINTER, SOURCEFIND, and GLOBETROTTER for the Balti population. The results supported the scenario of one-way admixture of the Tibetan and Kashmiri populations for the Balti dated to 27 generations ago (supplementary fig. S9 and table S4, Supplementary Material online). We adopted linkage disequilibrium-based methods ALDER (Loh et al. 2013) to estimate admixtures date of 21.31 ± 3.2 generations in the Balti using Tibetan and Kashmiri populations as proxies for the ancestral sources (fig. 3D and supplementary table S4, Supplementary Material online).

We applied qpAdm and qpGraph (Patterson et al. 2012) to model the admixture scenario inferred above for the Balti population (supplementary text, Supplementary Material online). The qpAdm analyses supported the candidate ancestry sources and their proportions as Kashmiri (76%) and Tibetan (24%) (supplementary table S5, Supplementary Material online). The qpGraph analysis showed the graph fitted recent Tibetan-ancestry gene flow into the Balti population (supplementary fig. S10, Supplementary Material online). When the published ancient DNA data from South and Central Asia (Narasimhan et al. 2019) were included for outgroup f_3 and f_4 statistics, we found no evidence of Tibetan-derived gene flow into the ancient samples (supplementary fig. S11, Supplementary Material online).

Analyses Based on Local Ancestry Inference

To further characterize the demographic history of specific ancestries, we intersected the results of local ancestry inference in the Balti genomes with referring to Tibetan and Kashmiri as candidate source populations in PCAdmix (Brisbin et al. 2012) (supplementary fig. S12A, Supplementary Material online) and RFMix (Maples et al. 2013) (supplementary fig. S12B, Supplementary Material online). We then calculated the length distribution of the ancestral tracts detected with PCAdmix to infer the admixture in the Balti population under various scenarios (fig. 3D and supplementary table S4, Supplementary Material online). The results of *AdmixInfer* (Ni et al. 2016) under a general admixture model suggested a continuous gene flow model for the Tibetan (22.6%)–Kashmiri (77.4%) admixture occurred about 39 generations ago. Further estimation using *MultiWaver* (Ni et al. 2019) under multiple-wave admixtures model also indicated one wave of the Tibetan (22.6%)–Kashmiri (77.4%) admixture taking place about 36 generations ago, rejecting the scenario of multiple-wave admixtures.

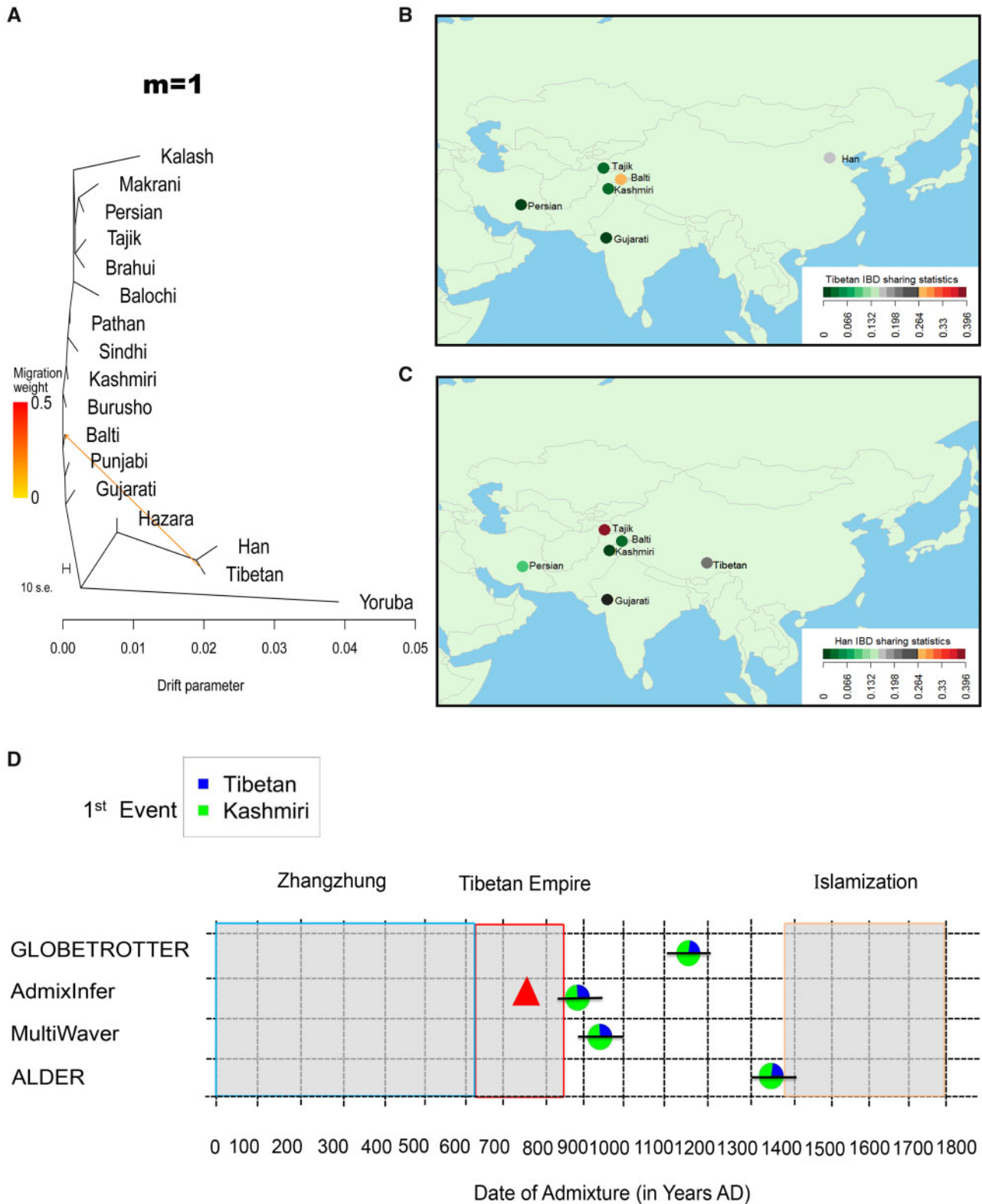


Fig. 3. Gene flow, IBD sharing, and genetic dating. (A) The gene flow from the Tibetan into the Balti people inferred by using TreeMix for migration edges $m = 1$. The degree of recent gene flow measured by shared IBD segments between the Balti and Tibetan (B) and between the Balti and Han Chinese (C). The IBD segments with length ≥ 2 cM and significant score $< 10^{-5}$ were included. (D) Admixture event in the Balti population. The pie charts reflected the admixture proportions of ancestries and the lines represent the 95% confidence intervals of admixture dates. The durations for the Zhangzhung kingdom, the Tibetan Empire, and the Islamization of Baltistan are noted. The triangle refers to the time for Tibetan Empire conquered Baltistan in history.

We retrieved the selective genes for high-altitude adaptation in Tibetans reported in published references and then scanned their enrichment in the Balti genomes (supplementary tables S6 and S7, Supplementary Material online). We adopted the threshold of genome-wide mean + 3SD (Bryc et al. 2015) to indicate an excess of local Tibetan ancestry in the RFMix (Maples et al. 2013) inferred tracts (supplementary fig. S12C and D, Supplementary Material online). We found that few Tibetan-specific adaptive genes were significantly enriched (>3 SD) (supplementary tables S6 and S7, Supplementary Material online). Nevertheless, the potential sampling bias due to small sample size ($n = 18$) might limit power to test the postadmixture selection hypothesis (Tang et al. 2007), which generally requires huge sample size (Jin et al. 2012; Bhatia et al. 2014).

Evidence for Sex-Biased Gene Flow

To check the roles of both sexes in the admixture, we analyzed the mitochondrial DNA (mtDNA), Y, and X chromosomal variation retrieved from the WGS data (supplementary text, Supplementary Material online). The mtDNA gene pool (supplementary table S1, Supplementary Material online) of the 18 Balti individuals comprised the haplogroups being predominant in West Eurasians (4/18): H6a1b, HV, R0a2, and U4b1a1a1 (Pala et al. 2012), South Asians 6/18: M5a2a, M65a, M65b, T2c, U2a1a, and U2b2 (Palanichamy et al. 2004; Sun et al. 2006), and East Asians 8/18: A, B4a4, F1d, M70, M9a1a1cb1a, and Z7 (Kong et al. 2003; Tanaka et al. 2004). In particular, mtDNA haplogroups M70 (Kong et al. 2011; Kang et al. 2013), M9a1a1cb1a (Peng et al. 2011), and Z7 (Chandrasekar et al. 2009) were characterized in Tibetan, Sherpa, and Monpa populations. The dissection of Y chromosomal profile (supplementary table S1, Supplementary Material online) revealed that most paternal lineages (9/10) belong to haplogroups J, L, and R in the Balti males are likely from South Asians (Mondal et al. 2017; Ullah et al. 2017), or the neighboring West Eurasians (Lacau et al. 2012). The Tibetan-predominant haplogroup D1 (Shi et al. 2008) occurred in one Balti male. The Tibetan-ancestry proportions inferred by ADMIXTURE (Ongaro et al. 2019) were significantly higher (Wilcoxon test, $P < 0.05$) in autosomes than X chromosome (supplementary fig. S13, Supplementary Material online), suggesting higher contribution of Tibetan-ancestry males than females into the gene pool of Balti population.

Demic and Cultural Diffusion

Our population genomic analyses showed that the Balti population is a consequence of genetic admixture of exotic Tibetan (22.6–26%) and indigenous Kashmiri ancestries (74–77.4%) (fig. 3D). The gene flow from the Tibetan into the Balti population was dated around 39–21 generations ago, corresponding to AD 869–1391, assuming a generation time of 29 years (Langergraber et al. 2012), with multiple dating approaches (fig. 3D). The historical records report the Baltis' homeland—Baltistan (known as Bruzha)—was conquered by the armed forces of Tibetan Empire at AD

757 (Denwood 2008) and then was controlled by Tibetans until 10th century (Dani 1998). The collapse of Tibetan Empire in AD 841 and subsequent civil wars drove mass migrations of Tibetans into northwestern Tibet and Kashmir (including Baltistan) regions (Shakabpa and Maher 2009). In particular, the analysis of X chromosome showed the gene flow was male-biased. Our results support the scenario, that is, the Tibetan-ancestry in the modern Balti population could be traced back to the military invasions and subsequent immigrants of Tibetan Empire.

The Tibetan-derived gene flow also brought substantial culture diffusion including language shift and elite dominance. The Balti people are Tibetic language speakers (Owen-Smith and Hill 2013). They historically practiced Tibetan writing Bön, and Tibetan Buddhism before the arrival of Islam around the 15th century AD (Walter 2009). Our analyses of ROHs (fig. 2C) suggested that the Balti population did not adopt the traditional endogamy which was popular in Kashmiris (Jabeen and Malik 2014). Thus, the expansion of ancient Tibetan Empire into Baltistan brought minor demic diffusion accompanied with dominant cultural diffusion. Given the genetic diversity and structure existing in populations of Pakistan (Metspalu et al. 2011) and Tibet of China (Jeong et al. 2017; Zhang et al. 2017), the population genomic investigation with extensive sampling (Cole et al. 2017; Arciero et al. 2018), even accompanying with ancient DNA (Jeong et al. 2016; Narasimhan et al. 2019), are still required to discern the detailed genetic changes during the expansion of ancient Tibetan Empire.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We are grateful to all volunteers and samplers. We thank Rasmus Nielsen and Guanghui Dong for comments and discussions and Shaobin Xu for assistance in computation. This study was supported by grants from the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA20040100 and XDA20090000), the National Natural Science Foundation of China (31621062 to Y.-P. Z.) and the Bureau of Science and Technology of Yunnan Province, the Chinese Academy of Sciences President's International Fellowship Initiative, and the Animal Branch of the Germplasm Bank of Wild Species, Chinese Academy of Sciences (the Large Research Infrastructure Funding). M.-S.P. appreciates the support from the Youth Innovation Promotion Association of the Chinese Academy of Sciences.

Author Contributions

Y.P.Z., M.S.P., and A.R. designed research. A.R., N.T.A., M.E.A., K.B.I., X.C., W.K.Y., and A.A. collected the samples. X.Y.Y. conducted experiments. X.Y.Y. and M.S.P. analyzed the data and wrote the paper. Y.P.Z., W.C., J.H., X.B.Q., Y.G.Y., and B.S. revised the paper. Q.K.S., S.S.D., X.S., and R.H.Z. provided

technical assistance. All authors read and approved the final manuscript.

Data Availability

The newly generated genome-wide sequence data of 60 individuals (FASTQ format) in this study have been deposited in the Genome sequence Archive in Big Data Center (<http://gsa.big.ac.cn/>; accession project numbers PRJCA000456 and PRJCA000457) and Japanese Genotype-Phenotype Archive (<http://trace.ddbj.nig.ac.jp/jga/>; study accession number JGAS00000000238).

References

- 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* 526(7571):68–74.
- Aldenderfer M, Yinong Z. 2004. The prehistory of the Tibetan Plateau to the seventh century AD: perspectives and research from China and the West since 1950. *J World Prehist.* 18(1):1–55.
- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–1664.
- Arciero E, Kraaijenbrink T, Haber M, Mezzavilla M, Ayub Q, Wang W, Pingcuo Z, Yang H, Wang J, Jobling MA, et al. 2018. Demographic history and genetic adaptation in the Himalayan region inferred from genome-wide SNP genotypes of 49 populations. *Mol Biol Evol.* 35(8):1916–1933.
- Bai H, Guo X, Narisu N, Lan T, Wu Q, Xing Y, Zhang Y, Bond SR, Pei Z, Zhang Y, et al. 2018. Whole-genome sequencing of 175 Mongolians uncovers population-specific genetic architecture and gene flow throughout North and East Asia. *Nat Genet.* 50(12):1696–1704.
- Bakshi SR. 1997. Kashmir: history and people. New Delhi: Sarup & Sons.
- Beckwith CI. 1993. The Tibetan Empire in Central Asia: a history of the struggle for the great power among Tibetans, Turks, Arabs, and Chinese during the early middle ages. Princeton: Princeton University Press.
- Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, Chen Y, Felkel S, Hallast P, Kamm J, et al. 2020. Insights into human genetic variation and population history from 929 diverse genomes. *Science* 367(6484):eaay5012.
- Bhatia G, Tandon A, Patterson N, Aldrich MC, Ambrosone CB, Amos C, Bandera EV, Berndt SI, Bernstein L, Blot WJ, et al. 2014. Genome-wide scan of 29,141 African Americans finds no evidence of directional selection since admixture. *Am J Hum Genet.* 95(4):437–444.
- Brisbin A, Bryc K, Byrnes J, Zakharia F, Omberg L, Degenhardt J, Reynolds A, Ostrer H, Mezey J, Bustamante CD. 2012. PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum Biol.* 84(4):343–364.
- Brucato N, Kusuma P, Beaujard P, Sudoyo H, Cox MP, Ricaut F-X. 2017. Genomic admixture tracks pulses of economic activity over 2,000 years in the Indian Ocean trading network. *Sci Rep.* 7(1):2919.
- Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. 2015. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am J Hum Genet.* 96(1):37–53.
- Busby GB, Hellenthal G, Montinaro F, Tofanelli S, Bulayeva K, Rudan I, Zemunik T, Hayward C, Toncheva D, Karachanak-Yankova S, et al. 2015. The role of recent admixture in forming the contemporary west Eurasian genomic landscape. *Curr Biol.* 25(19):2518–2526.
- Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. 2018. Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet.* 19(4):220–234.
- Chacon-Duque JC, Adhikari K, Fuentes-Guajardo M, Mendoza-Revilla J, Acuna-Alonzo V, Barquera R, Quinto-Sanchez M, Gomez-Valdes J, Martinez PE, Villamil-Ramirez H, et al. 2018. Latin Americans show wide-spread Converso ancestry and imprint of local native ancestry on physical appearance. *Nat Commun.* 9(1):1–13.
- Chandrasekar A, Kumar S, Sreenath J, Sarkar BN, Urade BP, Mallick S, Bandopadhyay SS, Barua P, Barik SS, Basu D, et al. 2009. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. *PLoS One* 4(10):e7447.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience.* 4(1):s13742-13015-10047-13748.
- Charati H, Peng MS, Chen W, Yang XY, Jabbari Ori R, Aghajanzpour-Mir M, Esmailzadeh A, Zhang YP. 2019. The evolutionary genetics of lactase persistence in seven ethnic groups across the Iranian plateau. *Hum Genomics.* 13(1):7.
- Cole AM, Cox S, Jeong C, Petousi N, Aryal DR, Droma Y, Hanaoka M, Ota M, Kobayashi N, Gasparini P, et al. 2017. Genetic structure in the Sherpa and neighboring Nepalese populations. *BMC Genomics.* 18(1):102.
- Dani AH. 1998. The western Himalayan states. *Hist Civilizations Central Asia.* 4:215–225.
- Denwood P. 2008. The Tibetans in the west, part I. *J Inner Asian Art Archaeol.* 3:7–21.
- Fernandes V, Brucato N, Ferreira JC, Pedro N, Cavadas B, Ricaut FX, Alshamali F, Pereira L. 2019. Genome-wide characterization of Arabian peninsula populations: shedding light on the history of a fundamental bridge between continents. *Mol Biol Evol.* 36(3):575–586.
- Hellenthal G, Busby GB, Band G, Wilson JF, Capelli C, Falush D, Myers S. 2014. A genetic atlas of human admixture history. *Science* 343(6172):747–751.
- Jabeen N, Malik S. 2014. Consanguinity and its sociodemographic differentials in Bhimber district, Azad Jammu and Kashmir, Pakistan. *J Health Popul Nutr.* 32(2):301–313.
- Jeong C, Ozga AT, Witonsky DB, Malmstrom H, Edlund H, Hofman CA, Hagan RW, Jakobsson M, Lewis CM, Aldenderfer MS, et al. 2016. Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc Natl Acad Sci U S A.* 113(27):7485–7490.
- Jeong C, Peter BM, Basnyat B, Neupane M, Beall CM, Childs G, Craig SR, Novembre J, Di Rienzo A. 2017. A longitudinal cline characterizes the genetic structure of human populations in the Tibetan plateau. *PLoS One* 12(4):e0175885.
- Jin W, Xu S, Wang H, Yu Y, Shen Y, Wu B, Jin L. 2012. Genome-wide detection of natural selection in African Americans pre- and post-admixture. *Genome Res.* 22(3):519–527.
- Kang L, Zheng HX, Chen F, Yan S, Liu K, Qin Z, Liu L, Zhao Z, Li L, Wang X, et al. 2013. mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands. *Mol Biol Evol.* 30(12):2579–2587.
- Kong QP, Sun C, Wang HW, Zhao M, Wang WZ, Zhong L, Hao XD, Pan H, Wang SY, Cheng YT, et al. 2011. Large-scale mtDNA screening reveals a surprising matrilineal complexity in East Asia and its implications to the peopling of the region. *Mol Biol Evol.* 28(1):513–522.
- Kong Q-P, Yao Y-G, Sun C, Bandelt H-J, Zhu C-L, Zhang Y-P. 2003. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet.* 73(3):671–676.
- Lacau H, Gayden T, Regueiro M, Chennakrishnaiah S, Bukhari A, Underhill PA, Garcia-Bertrand RL, Herrera RJ. 2012. Afghanistan from a Y-chromosome perspective. *Eur J Hum Genet.* 20(10):1063–1070.
- Langergraber KE, Prüfer K, Rowney C, Boesch C, Crockford C, Fawcett K, Inoue E, Inoue-Muruyama M, Mitani JC, Muller MN, et al. 2012. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc Natl Acad Sci U S A.* 109(39):15716–15721.
- Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. *PLoS Genet.* 8(1):e1002453.
- Loh PR, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, Berger B. 2013. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* 193(4):1233–1254.
- Lu D, Lou H, Yuan K, Wang X, Wang Y, Zhang C, Lu Y, Yang X, Deng L, Zhou Y, et al. 2016. Ancestral origins and genetic history of Tibetan highlanders. *Am J Hum Genet.* 99(3):580–594.

- Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, et al. 2016. The Simons genome diversity project: 300 genomes from 142 diverse populations. *Nature* 538(7624):201–206.
- Maples BK, Gravel S, Kenny EE, Bustamante CD. 2013. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet.* 93(2):278–288.
- Metspalu M, Romero IG, Yunusbayev B, Chaubey G, Mallick CB, Hudjashov G, Nelis M, Magi R, Metspalu E, Remm M, et al. 2011. Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *Am J Hum Genet.* 89(6):731–744.
- Ming Y. 2007. "Tibet-Bolor Path" in China-West traffic in the Tang Dynasty. *West Reg Stud.* (2):12.
- Mondal M, Bergström A, Xue Y, Calafell F, Laayouni H, Casals F, Majumder PP, Tyler-Smith C, Bertranpetit J. 2017. Y-chromosomal sequences of diverse Indian populations and the ancestry of the Andamanese. *Hum Genet.* 136(5):499–510.
- Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, Lazaridis L, Nakatsuka N, Olalde I, Lipson M, et al. 2019. The formation of human populations in South and Central Asia. *Science* 365(6457):eaat7487.
- Ni X, Yang X, Guo W, Yuan K, Zhou Y, Ma Z, Xu S. 2016. Length distribution of ancestral tracks under a general admixture model and its applications in population history. *Sci Rep.* 6(1):20048.
- Ni X, Yuan K, Liu C, Feng Q, Tian L, Ma Z, Xu S. 2019. MultiWaver 2.0: modeling discrete and continuous gene flow to reconstruct complex population admixtures. *Eur J Hum Genet.* 27(1):133–139.
- Ongaro L, Scliar MO, Flores R, Raveane A, Marnetto D, Sarno S, Gnechiruscone GA, Alarcón-Riquelme ME, Patin E, Wangkumhang P, et al. 2019. The genomic impact of European colonization of the Americas. *Curr Biol.* 29(23):3974–3986.
- Owen-Smith T, Hill N. 2013. Trans-Himalayan linguistics: historical and descriptive linguistics of the Himalayan area. Berlin: Walter de Gruyter.
- Pala M, Olivieri A, Achilli A, Accetturo M, Metspalu E, Reidla M, Tamm E, Karmin M, Reisberg T, Hooshiar Kashani B, et al. 2012. Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am J Hum Genet.* 90(5):915–924.
- Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri TK, Palla V, Zhang YP. 2004. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am J Hum Genet.* 75(6):966–978.
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics* 192(3):1065–1093.
- Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet.* 2(12):e190.
- Peng MS, Palanichamy MG, Yao YG, Mitra B, Cheng YT, Zhao M, Liu J, Wang HW, Pan H, Wang WZ, et al. 2011. Inland post-glacial dispersal in East Asia revealed by mitochondrial haplogroup M9a'b. *BMC Biol.* 9(1):2.
- Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8(11):e1002967.
- Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ. 2002. Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet.* 70(5):1107–1124.
- Shakabpa W, Maher D. 2009. One hundred thousand moons: an advanced political history of Tibet. Leiden: BRILL.
- Shi H, Zhong H, Peng Y, Dong YL, Qi XB, Zhang F, Liu LF, Tan SJ, Ma RZ, Xiao CJ, et al. 2008. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol.* 6(1):45.
- Sun C, Kong QP, Palanichamy MG, Agrawal S, Bandelt HJ, Yao YG, Khan F, Zhu CL, Chaudhuri TK, Zhang YP. 2006. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Mol Biol Evol.* 23(3):683–690.
- Tanaka M, Cabrera VM, González AM, Larruga JM, Takeyasu T, Fuku N, Guo L-J, Hirose R, Fujita Y, Kurata M. 2004. Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Res.* 14(10a):1832–1850.
- Tang H, Choudhry S, Mei R, Morgan M, Rodriguez-Cintron W, Burchard EG, Risch NJ. 2007. Recent genetic selection in the ancestral admixture of Puerto Ricans. *Am J Hum Genet.* 81(3):626–633.
- Ullah I, Olofsson JK, Margaryan A, Ilardo M, Ahmad H, Sikora M, Hansen AJ, Shahid Nadeem M, Fazal N, Ali M, et al. 2017. High Y-chromosomal differentiation among ethnic groups of Dir and Swat districts, Pakistan. *Ann Hum Genet.* 81(6):234–248.
- Uray G. 1979. Khrom: administrative units of the Tibetan Empire in the 7th–9th centuries. Westminster: Aris and Phillips.
- Walter ML. 2009. Buddhism and Empire: the political and religious culture of early Tibet. Leiden and Boston: Brill.
- Yang XY, Dai SS, Liu HQ, Peng MS, Zhang YP. 2018. The uncertainty of population relationship and divergence time inferred by the multiple sequentially Markovian coalescent model. *J Hum Genet.* 63(6):775–777.
- Zhang C, Lu Y, Feng Q, Wang X, Lou H, Liu J, Ning Z, Yuan K, Wang Y, Zhou Y, et al. 2017. Differentiated demographic histories and local adaptations between Sherpas and Tibetans. *Genome Biol.* 18(1):115.