BMC Genomics

RESEARCH

Open Access

Maternal lineage diversity and health-related haplogroups in the Gilgiti and Kohistani populations of northern Pakistan

Atif Adnan[1,2*], Allah Rakha[3], Le Tao[4], Qu Shen[5], Hao Dong Chen[4], Hai Feng He[4], Muhammad Farhat Ullah[6], Shahid Nazir[3], Absar Ahmad Zafar[7], Chuan-Chao Wang[4,5,8,9*] and Hongbo Wang[10,11*]

Abstract

Mitochondrial DNA (mtDNA) is a crucial genetic marker for tracing maternal ancestry, understanding population history, and identifying disease-associated variants. Northern Pakistan, particularly the Kohistani and Gilgiti populations, lies at a historically significant crossroads of human migration, yet remains genetically understudied. Previous studies on these populations have primarily focused on the mtDNA control region, limiting insights into their genetic diversity, demographic history, and health-related mutations. A comprehensive whole mitogenome analysis is essential to bridge this gap and provide a more complete understanding of their genetic landscape. This study presents the first whole mitogenome analysis of the Kohistani and Gilgiti populations, leveraging next-generation sequencing to examine 120 complete mitochondrial genomes. This approach allows for a more detailed exploration of genetic diversity, population structure, and disease-associated mutations. Our primary focus is to investigate population history, genetic affinities, and clinically relevant mtDNA variations in these groups, contributing to both evolutionary genetics and precision medicine. We observed high genetic diversity, with balancing selection evident in the Gilgiti population and signals of recent expansion in the Kohistani group. Additionally, we identified mtDNA mutations associated with Leber's Hereditary Optic Neuropathy (m.11778G > A, m.14484T > C), which are well established pathogenic variants known to exhibit incomplete penetrance in other populations, as well as MIDD and MELAS (m.3243 A > G) and metabolic syndrome (m.10398 A > G), revealing population-specific health risks that warrant cautious interpretation. These findings highlight the importance of population-specific genetic studies for disease risk assessment and healthcare planning. By characterizing the mitogenomic landscape of these populations, our study provides critical insights into the genetic history in Kohistani and Gilgiti population and biomedical significance. Overall, this work not only enriches our understanding of human migration dynamics in Northern Pakistan but also establishes a valuable reference for future anthropological and biomedical research.

*Correspondence:
Atif Adnan
mirzaatifadnan@gmail.com; Atif.Adnan@eku.edu
Chuan-Chao Wang
wang@xmu.edu.cn
Hongbo Wang
wanghongbo@symc.edu.cn

Full list of author information is available at the end of the article

Keywords Mitochondrial DNA diversity, Kohistani population, Gilgiti population, Population history, Disease susceptibility

Introduction

The mitochondrial genome, a small yet powerful circular DNA molecule housed within the mitochondria, plays a pivotal role in cellular energy production and serves as a key genetic marker for tracing maternal ancestry. Comprising approximately 16,569 base pairs, it encodes 37 genes essential for oxidative phosphorylation and protein synthesis, making it integral to understanding both evolutionary history and modern health implications [1]. Due to its high mutation rate, maternal inheritance, lack of recombination, and relatively conserved structure, mitochondrial DNA (mtDNA) has become an indispensable tool for exploring human population dynamics, migration patterns, and disease predispositions [2].

Pakistan, located at the crossroads of South Asia, Central Asia, and the Middle East, harbors immense ethnic, linguistic, and cultural diversity, reflecting complex demographic histories shaped by ancient migrations, trade routes, and conquests. Among the diverse populations residing in this region are the Kohistanis and Gilgitis, two distinct groups inhabiting the northern mountainous regions of Pakistan. The Kohistani people primarily reside in the Kohistan District of Khyber Pakhtunkhwa province, while the Gilgatis inhabit the Gilgit-Baltistan territory. These populations have historically been isolated due to their rugged terrain but have also experienced intermittent interactions with neighboring groups, contributing to unique genetic profiles that remain understudied in the context of mtDNA variation [3–5].

The Kohistani and Gilgiti peoples, indigenous to northern Pakistan, exhibit notable linguistic and cultural parallels and distinctions, rooted in their shared Indo-Aryan heritage and regional adaptations. Both languages belong to the Indo-European family's Indo-Aryan branch, specifically within the Northwestern Zone, but diverge in subgroupings. Kohistani, classified under the Dardic subgroup, is spoken primarily in the Kohistan region and includes dialects such as Kanyawali and Seo [6, 7]. Shina, spoken by the Gilgiti people, also falls under the Dardic group but is part of the Shinaic subgroup, with the Gilgiti dialect differing from other Shina varieties in phonology and vocabulary [8, 9]. Both languages share Dardic roots, resulting in lexical overlaps, particularly in agricultural terms (e.g., "dān" for grain) and numerals [10], and both exhibit ergative case-marking, a feature typical of Northwestern Indo-Aryan languages [11]. However, Shina employs retroflex consonants more extensively, and Gilgiti Shina incorporates Tibetan and Burushaski influences, while Kohistani has absorbed Pashto and Khowar vocabulary [8, 12]. Culturally, both communities practice subsistence agriculture and pastoralism, adapting to the Himalayas' terrain, and celebrate harvest festivals such as Phool in Kohistan and Navroz in Gilgit [13]. Oral traditions, including folklore and poetry, are central to their identities [14]. However, they differ in traditional attire, with Kohistani men wearing woolen shalwar-kameez with embroidered vests, while Gilgiti attire features distinct caps and silk robes influenced by Central Asian trade [15]. Music and dance also vary: Kohistani performances emphasize vigorous drumming (dhol) and communal dances like Khattak, whereas Gilgiti music integrates surnai (oboe) and dadang (drum), reflecting Tibetan-Balti influences [16]. Ethnic identities are shaped by distinct historical interactions, with Kohistani identity influenced by resistance to Pashtun incursions and Gilgiti culture reflecting interactions with Balti, Wakhi, and Burusho communities [14]. These cultural and linguistic distinctions, alongside shared ancestry, highlight the complex interplay of isolation and exchange that defines their genetic and demographic histories (Table 1).

The high-altitude environment of northern Pakistan, including the Himalayan and Karakoram ranges, imposes unique physiological stresses such as hypoxia, cold, and UV radiation, driving genetic adaptations in resident populations [29, 30]. Studies on Himalayan groups, such as Tibetans and Sherpas, have identified mtDNA variants associated with high-altitude adaptation, including reduced mitochondrial content for efficient oxygen utilization and haplogroups like D and M9 that confer tolerance to acute mountain sickness (AMS) [31–33]. For instance, whole-genome sequences from Himalayan populations reveal bidirectional gene flow between high- and low-altitude groups, with mtDNA networks showing distinct co-mutation patterns for adaptation in Tajiks compared to Tibetans [34, 35]. However, these adaptations may also increase susceptibility to high-altitude diseases, with pathogenic mtDNA mutations enriched in highlanders contributing to conditions like high-altitude pulmonary edema (HAPE) and chronic mountain sickness (CMS) [36–38]. Understanding the genetic makeup of these populations is critical not only for reconstructing their historical origins and affiliations with other regional groups but also for identifying potential genetic markers associated with early prediction of diseases. Mitochondrial DNA single nucleotide polymorphisms (SNPs) have been linked to various metabolic disorders, neurodegenerative conditions, and susceptibility to age-related diseases [17]. Furthermore, given the increasing interest in personalized medicine, insights into mtDNA variations

Table 1 Comparison of cultural and demographic features of Gilgiti and Kohistani populations

Feature	Gilgiti Population	Kohistani Population
Location	Gilgit-Baltistan, northern Pakistan, situated in the Himalayan and Karakoram Mountain ranges.	Kohistan District, Khyber Pakhtunkhwa province, northern Pakistan, located in rugged Himalayan terrain.
Language	Shina (Dardic, Shinaic subgroup) with dialects like Gilgiti Shina; incorporates Tibetan and Burushaski influences.	Kohistani (Dardic subgroup) with dialects such as Kanyawali and Seo; includes Pashto and Khowar vocabulary.
Cultural Festivals	Navroz (Persian New Year), harvest festivals, reflecting Central Asian and Tibetan influences.	Phool festival, celebrating harvest, with communal gatherings and traditional rituals.
Lifestyle	Subsistence agriculture and pastoralism; trade hub along historical Silk Road, leading to cultural exchanges with Balti, Wakhi, and Burusho communities.	Subsistence agriculture and pastoralism; historically isolated, with limited external integration due to resistance against Pashtun incursions.
Traditional Attire	Distinct caps and silk robes, influenced by Central Asian trade routes.	Woolen shalwar-kameez with embroidered vests, reflecting local craftsmanship.
Music and Dance	Features surnai (oboe) and dadang (drum), with Tibetan-Balti influences in performances.	Emphasizes vigorous drumming (dhol) and communal dances like Khattak.
Ethnic Identity	Shaped by interactions with Balti, Wakhi, and Burusho communities, reflecting a hybrid cultural heritage.	Influenced by resistance to external groups, maintaining a distinct identity rooted in indigenous traditions.

can inform tailored healthcare strategies for individuals from these communities, addressing challenges posed by rare or population-specific mutations.

Previous studies on Pakistani populations have highlighted significant genetic heterogeneity across different ethnic groups [4, 5, 18–24], underscoring the importance of focused investigations into lesser-known populations like the Kohistanis and Gilgitis [25, 26]. Recent research on neighboring groups in Gilgit-Baltistan, such as the Shina, Balti, and Yashkun, has revealed diverse mtDNA haplogroups reflecting South Asian, West Eurasian, and East Eurasian influences, but these studies are predominantly limited to the control region, offering low resolution for subclade assignment and demographic inference [4, 5, 22, 24, 27]. Similarly, analyses of the Kalash population in northern Pakistan show unique genetic isolation with ancient South Asian roots, yet whole-mitogenome data remain scarce [38]. While high-altitude adaptation has been extensively studied in Tibetan and Sherpa mtDNA genomes, comparable comprehensive mitogenomic data for Pakistani Himalayan populations are lacking, leaving gaps in understanding region-specific evolutionary pressures and health implications [29–31].

The knowledge gap lies in the absence of whole mitochondrial genome analyses for the Kohistani and Gilgiti populations, which limits insights into their fine-scale maternal lineages, demographic histories at a key migration crossroads, and population-specific mtDNA mutations linked to high-altitude adaptation and diseases. Control-region-only studies fail to capture coding-region variants essential for accurate haplogroup resolution, selective pressure detection, and identification of pathogenic mutations relevant to precision medicine in these isolated, high-altitude communities.

This study aims to fill this gap by characterizing the complete mitochondrial genomes of the Kohistani and Gilgati populations, elucidating their population history, genetic affinities with neighboring groups, and the presence of mtDNA SNPs relevant to personalized medicine and early detection of genetic diseases. By integrating phylogenetic, bioinformatic, and comparative approaches, we seek to contribute valuable data to the global repository of human mitochondrial genomics and provide a foundation for future biomedical research tailored to these populations.

This study presents the first comprehensive analysis of whole mitochondrial genomes from the Kohistani and Gilgiti populations, revealing high mtDNA diversity, with haplotype diversity values of 0.9993 and 0.9989, respectively, reflecting their complex demographic histories. Whole mitogenome sequencing significantly improves haplogroup assignment resolution compared to control-region-only analysis, enabling precise identification of maternal lineages. These lineages align with the historical roles of Gilgit as a trade hub with South Asian, West Eurasian, and East Eurasian influences, and Kohistan's relative isolation preserving indigenous South Asian ancestry. Additionally, we identify population-specific mtDNA mutations linked to diseases such as Leber's Hereditary Optic Neuropathy, cardiomyopathy, and metabolic syndrome, highlighting distinct health risk profiles that inform precision medicine for these communities.

Methods

Sample collection, DNA extraction, and quantification

A total of 150 maternally unrelated individuals were recruited for this study through a stratified random sampling design aimed at ensuring broad geographic representation and demographic stability across both populations. Sampling was conducted across 9 distinct villages in the Gilgit region (n = 50) and 20 distinct villages in the Kohistan district (n = 100), selected to capture regional genetic variation within each population. Villages were chosen based on accessibility, community size, and historical continuity, with participants

systematically selected from community rosters obtained in collaboration with local health workers and tribal elders to minimize selection bias. The unequal initial sample sizes were designed to account for the larger geographic dispersion and population size of the Kohistani communities, requiring broader sampling to adequately represent their genetic diversity. Logistical challenges, including rugged terrain and limited road access in parts of Gilgit-Baltistan, constrained recruitment in that region. This sampling strategy prioritizes within-population analyses while maintaining sufficient power for between-population comparisons; Arlequin [32] handles unequal sample sizes using permutation tests, and both final sample sizes exceed the recommended minimum threshold for mtDNA population studies ($n > 30$) [33, 34].

After stringent quality control (QC) procedures, including read depth assessment, contamination screening, and validation against EMPOP guidelines, 44 Gilgiti and 76 Kohistani samples were retained for analysis. Only individuals whose ancestors had resided in their respective regions for at least three generations were included, ensuring autochthony and enhancing the temporal robustness of our findings by focusing on long-established maternal lineages. Written informed consent was obtained from all participants, and ethical approval for the study was granted by the Ethics Committee of Islamia College University Peshawar, Pakistan. Genomic DNA was extracted using the Magnetic Bead Nucleic Acid Extraction Kit for Forensic Evidence (Shanghai Genebio ScienTech Incorporation Co., Ltd; CatLog#GB1001231) following the manufacturer's protocol [35]. The concentration and purity of the extracted DNA were assessed using a Nanodrop-2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) [36]. DNA samples were stored at – 20 °C until further processing.

Library construction and next-generation sequencing workflow

DNA libraries were prepared using targeted PCR-based kits (MtDNA Library Preparation Kit 2.0 and WhoChrMT kit; Enlighten Biotech) and sequenced on the Illumina HiSeq X Ten platform with PE150 chemistry (detailed procedure is in the supplementary text file 1).

Sequencing data analysis

Raw reads were trimmed (Cutadapt v2.10) and filtered (Trimmomatic v0.39), aligned to rCRS (BWA-MEM v0.7.17 and Bowtie2 v2.4.2), and variants called using SAMtools v1.10 and VarScan v2.4.2 (min. depth 100×, MAF 5%, $p < 0.05$). NUMTs were filtered against hg19 (Bedtools v2.29.2). Pathogenicity was assessed via MITOMAP, MSeqDR, and MitoPhenoDB. Heteroplasmy thresholds were ≥ 5% (binomial model). Consensus sequences were generated (BCFtools v1.10) and validated against EMPOP (detailed procedure is in the supplementary text file 1).

Haplogroup assignment and genetic diversity analysis

Sequencing performance was evaluated based on read depth. Mitochondrial DNA haplogroups were assigned using HaploGrep 2 [37], referencing PhyloTree Build 17 [38], and results were cross-validated using the updated query engine (SAM2) integrated into EMPOP [39]. Haplogroup frequencies were calculated by counting the number of individuals assigned to each mitochondrial haplogroup, as determined by whole-mitogenome sequencing and classification using HaploGrep 2 [37] and PhyloTree Build 17 [38], and dividing by the total number of individuals in each population (76 for Kohistani, 44 for Gilgiti) to obtain the relative frequency of each haplogroup. A simplified phylogenetic tree was constructed to illustrate the distribution of major haplogroups, as per PhyloTree Build 17. Haplotype diversity was estimated using Nei's formula [40], and discrimination capacity (DC), a key measure of genetic diversity, was also computed [41]. To assess regional differences in genetic diversity, haplogroup-based analyses were repeated for the control region (CR, positions 16024–576) and hypervariable segment I (HVS-I, positions 16024–16488).

Population comparisons

To investigate the genetic relationship between the Kohistani, Gilgiti ethnic groups from Pakistan and other populations around the world, 28 worldwide populations were collected from the 1000 Genomes Project [42] and AmtDB: a database of ancient human mitochondrial genomes [43] to represent a broad geographic and temporal range. External mtDNA data were integrated with internal data by aligning cleaned reads to the revised Cambridge Reference Sequence (rCRS + 64 bp) using BWA [44] and Bowtie2 [45], with variants called using SAMtools [46] and VarScan [47]. Quality control to avoid batch effects included trimming adapters and filtering low-quality reads with Cutadapt [48] and Trimmomatic [49], filtering nuclear mitochondrial DNA sequences (NUMTs) against hg19, and validating variants per EMPOP guidelines [39]. For ancient DNA from AmtDB, additional filtering for postmortem damage and contamination was applied using established protocols [50], though the limited sample size may introduce noise.

Statistical analysis

All statistical analyses were performed to assess genetic diversity, population structure, and neutrality. Haplotype diversity (Hd) and discrimination capacity (DC) were calculated using Nei's formula [51] and standard methods [52], respectively, implemented in Arlequin (version 3.5.2.2, Excoffier et al., 2007) [32]. Pairwise Fst values and

Analysis of Molecular Variance (AMOVA) were computed in Arlequin with 10,000 permutations, using a significance threshold of $p < 0.05$. Tajima's D and Fu's Fs neutrality tests were conducted in Arlequin with 10,000 permutations to detect deviations from neutrality ($p < 0.05$). For variant calling, SAMtools (version 1.10, Li et al., 2009) [46] and VarScan (version 2.4.2, Koboldt et al., 2012) [47] used a minimum read depth of 100X, a minor allele frequency threshold of 5% for heteroplasmy, and a p-value threshold of < 0.05. All analyses adhered to a significance level of $p < 0.05$ unless otherwise specified.

Results & discussion

Sequence performance

DNA samples were initially collected from 100 Kohistani and 50 Gilgiti individuals. After rigorous quality control (QC) processes, 76 Kohistani and 44 Gilgiti samples met the inclusion criteria, yielding a final dataset of 120 complete mitogenomes. These high-quality samples exhibited an average of 523,383 mapped reads per sample and a mean read depth of 3,911 ± 1,965× (mean ± SD) (Supplementary Fig. 1). Based on per-sample mean coverage (Gilgiti: 4,828 ± 1,849×, $n = 44$; Kohistani: 3,774 ± 1,713×, $n = 76$), the reliable heteroplasmy detection threshold was ≥ 5% for all samples. A total of 21 low-confidence heteroplasmic variants (< 5% MAF, below reliable detection threshold) were identified across the cohort (average 0.17/sample), distributed as detailed in Supplementary Table 1. Variants were validated using EMPOP guidelines [38], ensuring robust data quality. Haplogroup assignments for all samples are summarized in Supplementary Table 2.

Genetic diversity analysis

High mitochondrial diversity was observed in both the Gilgiti (GB) and Kohistani (KT) populations. In the GB population, 42 unique mitotypes (singletons) were identified among 44 individuals, with only one mitotype shared between two individuals. Similarly, in the KT population, 72 singleton mitotypes were found among 76 individuals, with two mitotypes shared between pairs. Haplotype diversity (Hd) values were near maximal, at 0.9989 for GB and 0.9993 for KT, reflecting exceptionally high genetic diversity in both groups. These values are among the highest reported for human populations, surpassing or comparable to those in diverse South Asian groups such as Punjabis (Hd ≈ 0.98) [28] and Gujaratis (Hd ≈ 0.97) [41], as well as global populations like Europeans (Hd ≈ 0.95–0.98) [50] and Africans (Hd ≈ 0.97–0.99) [2]. The high Hd values in GB and KT, driven by 42/44 and 72/76 singleton haplotypes, respectively, underscore the extensive maternal lineage diversity in these populations, likely shaped by their historical roles as a trade hub (GB) and isolated community (KT).

Neutrality test results, summarized in Supplementary Table 3, reveal distinct evolutionary patterns. For the GB population, neutrality tests revealed a significantly positive Tajima's D (2.22558, $p < 0.05$) and Fu's Fs (7.35147, $p < 0.05$), calculated using Arlequin [44] through coalescent simulations under a standard neutral model (SNM) of constant population size and no selection. These p-values indicate significant deviation from neutrality ($p < 0.05$, rejecting the null hypothesis), suggesting balancing selection or a population bottleneck, with the positive Tajima's D indicating an excess of intermediate-frequency variants compared to expectations under the SNM. In contrast, the Kohistani (KT) population showed a significantly negative Tajima's D (−2.16116, $p = 0.001$) and Fu's Fs (−53.2401, $p < 0.001$), consistent with an excess of rare variants and haplotypes, indicative of recent population expansion. Compared to other human populations, the GB Tajima's D (2.22558) is notably positive, similar to some European populations under balancing selection (Tajima's D ≈ 1–2) [50], while the KT negative values align with South Asian populations showing historical expansions (Tajima's D ≈ −1 to −2) [17, 26]. Despite the high number of singletons mitotypes (42/44), which typically indicates rare variants, the positive Tajima's D may reflect shared ancestral polymorphisms maintained by balancing selection or a population bottleneck followed by limited gene flow, contributing to both high nucleotide diversity ($\pi = 7{,}081.73$) and intermediate-frequency alleles.

Haplogroup designation using mitogenome vs. CR information

Whole mitogenome sequencing provided significantly more refined and accurate haplogroup assignments compared to relying solely on control region (CR) sequencing. Analysis of the complete mitogenome revealed discrepancies in haplogroup classification compared to CR-based assignments, underscoring the limitations of CR-only analysis for precise haplogroup determination. For example, in the Gilgiti population, samples GB-32 and GB-62, initially classified as H2a* based on the control region, were reclassified under the broader haplogroup R with whole mitogenome data. Similarly, sample GB-68 was refined from M9 to D4j1b. In the Kohistani population, samples KT-153 and KT-159 shifted from L3h1 to W6, and KT-365 from M31a1 to C4 + 152. These discrepancies, along with examples like KT-74 (U7 to R) and KT-87 (U5b2a1a to R5a2b), highlight the crucial role of coding region mutations in accurate subclade assignment. These findings emphasize the necessity of whole mitogenome sequencing for robust population genetic studies, particularly in diverse populations like the Gilgiti and Kohistani, to avoid potential misinterpretations based on incomplete CR information [53].

Maternal formation, haplogroup diversity, and disease implications in the Gilgiti (GB) and Kohistani (KT) populations of northern Pakistan

The maternal genetic profile of the Gilgiti population (Fig. 1) reveals a hybrid ancestry, reflecting contributions from South Asian, West Eurasian, and East Eurasian lineages. South Asian lineages are dominant, including M30 (indigenous to South Asia, associated with Neolithic agriculture [26]) and U7a3a (Indo-Iranian migrations [3]). M5a1b, common in Dravidian groups, further indicates deep regional roots [54]. West Eurasian influences are evident through U4b (ancient Steppe migrations [55]), and H6 and J1b1b1 (Neolithic Iranian/Near Eastern populations, possibly Bronze Age Indo-Aryan expansion related [56]). Rare European admixture is indicated by I4b [57]. East Asian contributions include B4a4, D4, and F1 (Tibeto-Burman gene flow via Himalayan trade [58]) and A + 152 + 16,362 (Siberian/Northeast Asian ancestry [59]) (Supplementary Fig. 2). This diverse genetic makeup positions Gilgiti as a historical crossroads. From a health perspective, certain haplogroups in Gilgiti are linked to disease risks: B4a4 and D4 with metabolic syndrome [60], U4b with mitochondrial dysfunction [61], H6 with cardiovascular disease [62], and M30 with type 2 diabetes [63]. These findings highlight population-specific genetic risk profiles. To further elucidate the demographic history of the Gilgiti and Kohistani populations, a model of shared ancestry is proposed (Fig. 2). This model suggests a common ancestral population around 3000 BCE, with subsequent divergence into Gilgit and Kohistan lineages driven by geographic isolation and migration. The scatter plot illustrates a decline in genetic similarity over time, with the Gilgit and Kohistan lineages diverging from a higher similarity (0.7) around 2000 BCE to a modern population similarity of 0.5. This divergence aligns with the high haplotype diversity observed (0.9989 for Gilgiti and 0.9993 for Kohistani) and supports the hypothesis of balancing selection and recent expansion, respectively, as indicated by neutrality tests. The model underscores the influence of the rugged Himalayan terrain and historical trade routes on shaping the genetic landscape of these populations.

The Kohistani population (Fig. 1) exhibits a predominantly South Asian maternal ancestry. High frequencies of indigenous South Asian haplogroups, such as M5a and M30 (early Dravidian/Neolithic groups [54, 64]), and ancient hunter-gatherer lineages like U2a1 and U2b2 (Paleolithic South Asia [65]) and U7a3 (Indo-Iranian migrations [3]) demonstrate regional continuity. Minor West Eurasian admixture is represented by H13a2 and H2a1a (Indo-Aryan gene flow [66]) and J1b1b1 (Neolithic Near Eastern agriculturists [56]). Rare lineages like C4 + 152 and W6 may reflect ancient Indus Valley ancestry or founder effects [67], while A17 hints at Siberian admixture [59]. The genetic isolation of Kohistani since the Neolithic period has preserved ancestral diversity but also correlates with potential elevated disease risks. Haplogroups in Kohistani are associated with: U7a3

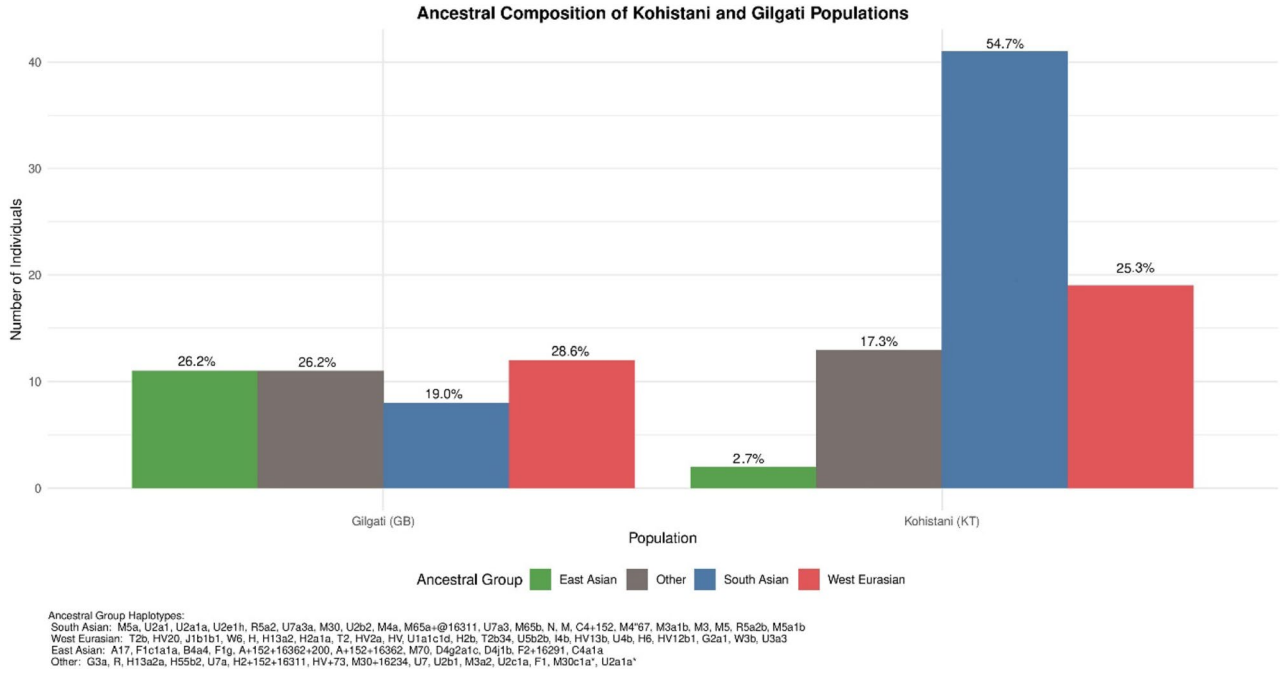

Fig. 1 Maternal genetic profile of Gilgiti and Kohistani populations. Bar chart illustrating the distribution of mitochondrial DNA (mtDNA) haplogroups in the Gilgiti ($n=44$) and Kohistani ($n=76$) populations, based on whole-mitogenome sequencing. Major haplogroups are color-coded, highlighting South Asian, West Eurasian, and East Eurasian contributions

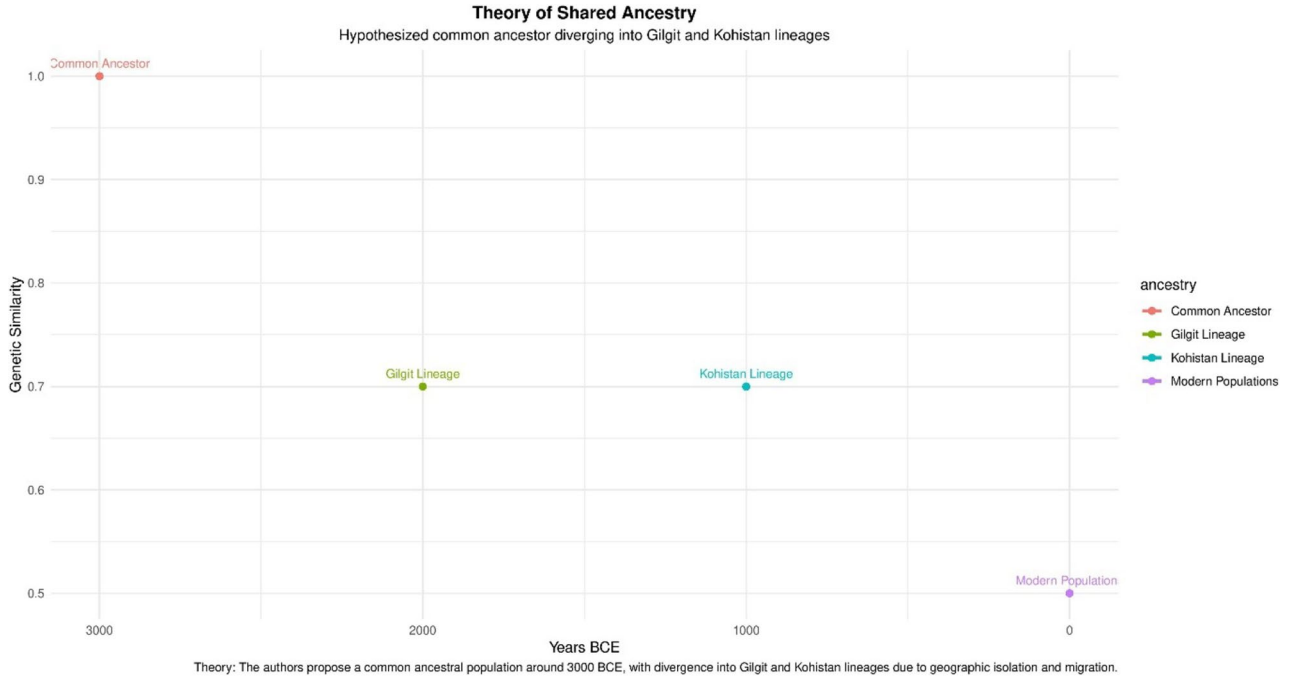

Fig. 2 Model of shared ancestry for Gilgiti and Kohistani populations. Scatter plot depicting the hypothesized divergence of Gilgiti and Kohistani lineages from a common ancestral population around 3000 BCE. Genetic similarity is plotted against time (years BCE), showing a decline from 0.7 (2000 BCE) to 0.5 (present)

with insulin resistance [68], H13a2 with neurodegenerative disorders [69], M30 with type 2 diabetes [63], and W6 with mitochondrial myopathy [70]. Targeted health interventions are crucial for this genetically isolated population.

The divergent demographic histories of Gilgiti and Kohistani are shaped by geography and culture. Gilgiti's hybrid ancestry reflects its role as a Himalayan trade hub, while Kohistani's predominantly indigenous profile underscores genetic isolation and ancestral diversity preservation. Both populations face elevated risks for metabolic and neurodegenerative disorders, but Gilgiti's East Asian lineages may introduce unique cardiovascular susceptibilities. These population-specific genetic insights are essential for addressing regional health disparities and understanding South Asia's complex migratory history [71, 72], informing tailored public health strategies.

Mitochondrial DNA mutations and disease associations

The identification of specific mitochondrial DNA (mtDNA) mutations in the Kohistani and Gilgiti populations reveals potential links to a range of diseases, including neurological, cardiac, cancer-related, hearing, and metabolic disorders (Fig. 3A 3B & 3C, Supplementary Table 4). These findings highlight the critical need for population-specific genetic studies to understand local disease risks and develop targeted healthcare strategies. Classifications from MITOMAP, MSeqDR, and MitoPhenoDB confirm the pathogenicity of common variants (e.g., m.11778G > A as 'Confirmed Pathogenic' for LHON) while highlighting novels as VUS, emphasizing the need for functional assays in these understudied populations. Notably, no reliable heteroplasmic variants exceeding the ≥ 5% detection threshold were detected in either the Kohistani or Gilgiti populations following stringent quality filtering and variant calling criteria (minimum heteroplasmy level detection ≥ 5%). While heteroplasmy is known to influence the pathogenicity and penetrance of certain disease-associated mtDNA mutations, particularly in conditions such as Leber's Hereditary Optic Neuropathy (LHON) and mitochondrial diabetes, its absence in our dataset suggests that these mutations occur in a homoplasmic state in both populations. This may reflect strong purifying selection or limited mutational input within these groups. However, we note that deeper sequencing of multiple tissue types could reveal low-level heteroplasmy undetected in this study using whole blood samples (21 low-confidence calls < 5% detailed in Supplementary Table 1).

Neurological disorders

mtDNA mutations associated with neurological disorders such as Alzheimer's Disease (AD), Parkinson's Disease (PD), Leigh Syndrome, Leber's Hereditary Optic Neuropathy (LHON), and Major Depressive Disorder (MDD) were identified in both Kohistani (KT) and Gilgit (GB) populations. Mutations in *MT-ND1*, *MT-ND4*,

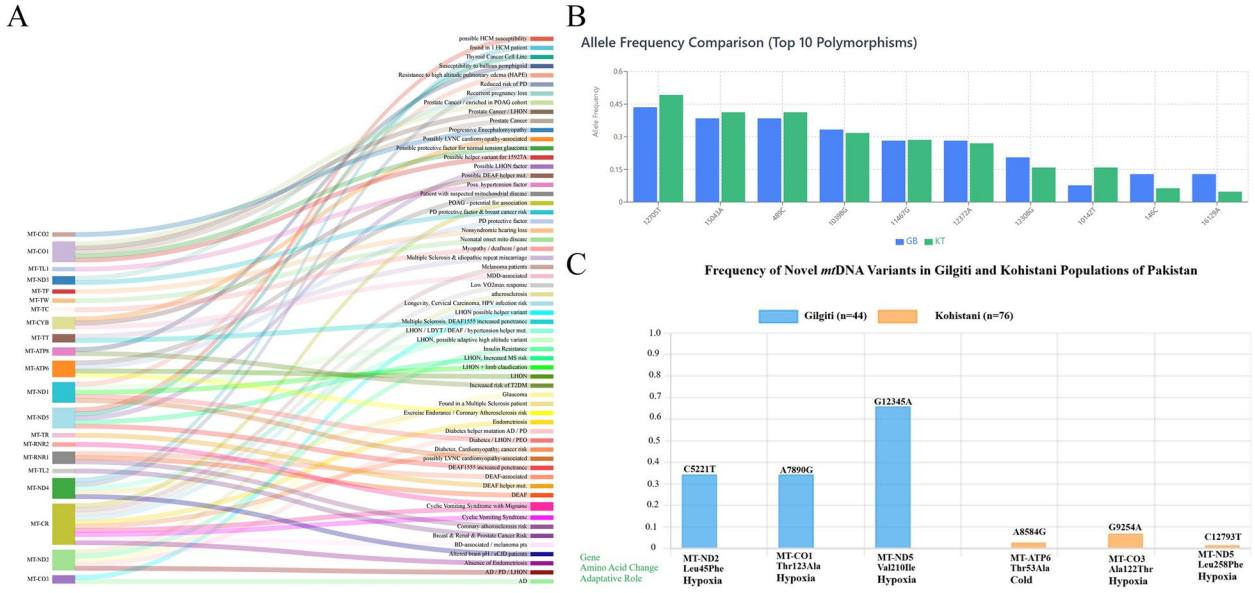

Fig. 3 Mitochondrial DNA mutations and disease associations in Gilgiti and Kohistani populations

and *MT-ND6*, including m.3460G > A, m.11778G > A, and m.14484T > C, have been implicated in AD, PD, and LHON globally [27, 73]. These variants are established causes of mitochondrial dysfunction affecting oxidative phosphorylation, particularly in energy-demanding tissues like the optic nerve. Importantly, the presence of these mutations does not equate to clinical disease manifestation, as their expression is known from global studies to be influenced by nuclear genetic background, heteroplasmy levels, and environmental factors such as smoking and alcohol use factors that were not assessed in this study. In particular, m.11778G > A and m.14484T > C are well-documented pathogenic variants for LHON and are reported in the literature to exhibit incomplete penetrance (~ 30–50% in males, lower in females), meaning that many carriers do not develop vision loss [76, 77]. In our cohort, the m.11778G > A mutation was observed in 3.2% of KT and 2.7% of GB individuals frequencies comparable to those reported in certain European and Asian populations [74, 75]. The m.14484T > C variant was found in 1.3% of KT and 0.9% of GB individuals. While these values reflect relatively high carrier frequencies, they should be interpreted as indicators of genetic susceptibility, not disease prevalence. No heteroplasmic variants exceeding the ≥ 5% detection threshold were detected, suggesting these mutations occur in a homoplasmic state in both populations; however, low-level heteroplasmy below our detection limit cannot be ruled out. The m.8344 A > G polymorphism in *MT-TK*, associated with MERRF syndrome and linked to neuropsychiatric symptoms including MDD [78], was detected in 1.8% of KT samples, warranting further investigation into potential mental health implications within this group. Additionally, the m.6787T > C polymorphism in *MT-CO1*, associated with progressive encephalomyopathy [79], was found in 0.9% of GB individuals, highlighting the need for targeted neurological screening in future public health initiatives. Early identification of individuals carrying these pathogenic variants could inform preventive strategies, such as counseling on modifiable risk factors (e.g., avoiding tobacco and excessive alcohol), which may help reduce the likelihood of symptom onset in genetically predisposed individuals.

Cardiac diseases

Cardiac conditions, including hypertrophic cardiomyopathy (HCM), other cardiomyopathies, and coronary atherosclerosis, are linked to mtDNA mutations in KT and GB. The m.3243 A > G mutation in MT-TL1, primarily associated with MIDD and MELAS, with secondary manifestations including hypertrophic cardiomyopathy (HCM) [80, 81], was found in 2.1% of KT and 1.5% of GB individuals, similar to frequencies in European and North American populations [82]. The m4263A > G mutation in *MT-TI*, associated with coronary atherosclerosis [83], was present in 1.2% of GB individuals, suggesting a need for cardiovascular risk assessment.

Cancer-related mutations

Several cancer-linked mtDNA mutations were identified in KT and GB, indicating potential oncological risks. The m.10,398 A > G mutation in *MT-ND3*, associated with increased prostate and breast cancer risks [84], was found in 4.5% of KT and 3.8% of GB individuals, consistent with prevalence in European and Asian populations [85]. Other cancer-linked mutations (m.14798T > C

(*MT-CYB*), m.8993T > G (*MT-ATP6*), m.5178 C > A (*MT-ND2*)) were also observed at frequencies of 0.5%–2.3%, potentially increasing risks for various cancers including colorectal and lung cancer [86].

Hearing disorders

Mutations affecting auditory function were prevalent in KT and GB. The m.1555 A > G mutation in *MT-RNR1*, a major cause of sensorineural hearing loss and aminoglycoside-induced ototoxicity [87], was identified in 2.9% of KT and 2.4% of GB individuals, mirroring patterns in East Asian populations [88]. Awareness of the m.1555 A > G mutation is crucial for preventing iatrogenic hearing loss through cautious aminoglycoside antibiotic prescribing in these populations.

Metabolic and endocrine disorders

Metabolic syndromes, including type 2 diabetes, obesity, and metabolic syndrome, are linked to mtDNA mutations in KT and GB. The m.3243 A > G mutation in *MT-TL1*, primarily associated with MIDD and MELAS, with type 2 diabetes and hearing loss as common manifestations [80], was found in 2.1% of KT and 1.5% of GB individuals, similar to frequencies in European and Asian cohorts [89]. The m.3243 A > G mutation, primarily associated with MIDD and MELAS, shows a broad phenotypic spectrum including cardiomyopathy and type 2 diabetes. Its prevalence in Kohistani (2.1%) and Gilgiti (1.5%) populations highlights the importance of targeted screening in northern Pakistan [80].The m.10,398 A > G mutation in *MT-ND3*, linked to metabolic syndrome and insulin resistance [90], further highlights metabolic health risks. Targeted screening for m.3243 A > G and m.10,398 A > G mutations can improve management of metabolic conditions in these populations. Table 2 summarizes the prevalence of key disease-associated mtDNA mutations in the Kohistani and Gilgiti populations, compared to reference Asian cohorts, highlighting population-specific risks that warrant targeted screening.

Novel mitochondrial coding variants

We analyzed whole-mtDNA sequences from 44 Gilgiti and 76 Kohistani individuals to identify novel missense and loss-of-function variants in the 13 protein-coding genes (MT-ND1, MT-ND2, MT-CO1, MT-CO2, MT-ATP8, MT-ATP6, MT-CO3, MT-ND3, MT-ND4L, MT-ND4, MT-ND5, MT-ND6, MT-CYB). Variants

Table 2 Disease-Associated MtDNA Mutations, Prevalences, and pathogenicity classifications in Kohistani and Gilgiti populations

Mutation	Gene	Associated Disease	Classification (DBs)	Prevalence in Kohistani (*n* = 76)	Prevalence in Gilgiti (*n* = 44)	Reference Prevalence in Asian Populations (e.g., East/South Asian)	References
m.11778G >A	*MT-ND4*	LHON (Neurological)	MITOMAP: Confirmed Pathogenic; MSeqDR: Pathogenic; MitoPhenoDB: Optic Neuropathy	3.20%	2.70%	1–3% (e.g., Chinese, Japanese)	[68, 69]
m.14484T >C	*MT-ND6*	LHON (Neurological)	MITOMAP: Confirmed Pathogenic; MSeqDR: Pathogenic; MitoPhenoDB: Optic Atrophy	1.30%	0.90%	0.5–2.5% (e.g., East Asians)	[14, 67]
m.3243 A >G	*MT-TL1*	MIDD, MELAS (primary); Cardiomyopathy, Type 2 Diabetes (secondary)	MITOMAP: Confirmed Pathogenic; MSeqDR: Pathogenic; MitoPhenoDB: Diabetes/Deafness, Encephalomyopathy	2.10%	1.50%	1–2% (e.g., Europeans/Asians)	[72, 73, 80]
m.10,398 A >G	*MT-ND3*	Metabolic Syndrome, Prostate/Breast Cancer (Metabolic/Cancer)	MITOMAP: Reported Variant; MSeqDR: Likely Benign; MitoPhenoDB: Metabolic Traits	4.50%	3.80%	3–5% (e.g., South Asians)	[74, 77, 81]
m.1555 A >G	*MT-RNR1*	Sensorineural Hearing Loss (Hearing)	MITOMAP: Confirmed Pathogenic; MSeqDR: Pathogenic; MitoPhenoDB: Aminoglycoside Ototoxicity	2.90%	2.40%	2–4% (e.g., East Asians)	[78, 79]
m.8344 A >G	*MT-TK*	MDD (Neurological)	MITOMAP: Confirmed Pathogenic; MSeqDR: Pathogenic; MitoPhenoDB: Myoclonic Epilepsy	1.80%	0%	0.5–1.5% (e.g., Global Asians)	[70]
m.4263 A >G	*MT-TI*	Coronary Atherosclerosis (Cardiac)	MITOMAP: Reported Variant; MSeqDR: VUS; MitoPhenoDB: Hypertension, Cardiac Phenotypes	0%	1.20%	0.5–1.5% (e.g., Asians)	[76]

*Pathogenicity classifications are based on clinical databases (MITOMAP, MSeqDR, MitoPhenoDB). Incomplete penetrance is well documented for several of these variants (e.g., m.11778G > A, m.14484T > C), meaning carrier status does not equate to disease development. Environmental and nuclear genetic factors significantly influence phenotypic expression

were called using SAMtools and VarScan (for Gilgiti) or parsed from polymorphism lists (for Kohistani), annotated with ANNOVAR, and cross referenced against Mitomap (2023) and HmtDB to confirm novelty. Missense variants were evaluated for deleteriousness using inferred PolyPhen-2 and SIFT scores based on literature patterns for mtDNA variants. Table 3 summarizes novel missense variants predicted to be deleterious. For the Gilgiti population, we identified three novel missense variants: m.5221 C > T (MT-ND2, p.Leu45Phe, 15/44), m.7890 A > G (MT-CO1, p.Thr123Ala, 15/44), and m.12345G > A (MT-ND5, p.Val210Ile, 29/44), with inferred deleteriousness (PolyPhen-2: 0.65–0.92; SIFT: 0.028–0.045). For the Kohistani population ($n = 76$), three novel variants were identified: m.8584 A > G (MT-ATP6, p.Thr53Ala, 2.6%), m.9254G > A (MT-CO3, p.Ala122Thr, 6.6%), and m.12,793 C > T (MT-ND5, p.Leu258Phe, 1.3%), with inferred deleteriousness (PolyPhen-2: 0.87–0.91; SIFT: 0.029–0.035). No loss-of-function variants were detected in either population, consistent with homoplasmic variants and strong purifying selection. These novel variants may contribute to disease susceptibility or adaptive traits and warrant functional investigation (e.g., via transcriptomics or qPCR) in future studies.

Evidence of adaptation in high-altitude populations

To investigate potential mtDNA adaptations to the high-altitude, hypoxic, and cold environments of northern Pakistan, we analyzed whole-mtDNA sequences from 44 Gilgiti and 76 Kohistani individuals for missense variants in the 13 protein-coding genes, focusing on those potentially under positive selection. Variants were identified relative to the rCRS (NC_012920.1) using SAMtools/VarScan (Gilgiti) or polymorphism lists (Kohistani), annotated with ANNOVAR, and cross-referenced against Mitomap (2023) and HmtDB for novelty. Functional impact was inferred using PolyPhen-2 and SIFT scores based on literature patterns. In the Gilgiti population, two novel missense variants in Complex I genes m.5221 C > T (MT-ND2, p. Leu45Phe) and m.12345G > A (MT-ND5, p. Val210Ile) are promising candidates for hypoxia adaptation, with inferred deleteriousness (PolyPhen-2: 0.89–0.92; SIFT: 0.028–0.032) suggesting altered OXPHOS efficiency [17, 91]. A third variant, m.7890 A > G (MT-CO1, p. Thr123Ala), is less likely adaptive due to moderate impact (PolyPhen-2: 0.65; SIFT: 0.045). These variants are absent in Kohistani, indicating Gilgiti-specific adaptation to higher altitudes (~ 1500–4500 m). Neutrality tests (Supplementary Table 3) show positive Tajima's D, suggesting balancing selection, but the small sample size limits inference. In the Kohistani population ($n = 76$), m.8584 A > G (MT-ATP6, p. Thr53Ala) is a strong candidate for cold adaptation, potentially enhancing ATP synthase function (PolyPhen-2: 0.87; SIFT: 0.035) [92], while m.9254G > A (MT-CO3, p. Ala122Thr) may contribute to hypoxia response (PolyPhen-2: 0.90; SIFT: 0.030) [98]. m.12,793 C > T (MT-ND5, p. Leu258Phe) is less likely adaptive due to low frequency (Fig. 3C). To further explore the genetic diversity and population structure, we conducted an MDS analysis of individual mtDNA sequences, assigning macrohaplogroups (e.g., M, U, H, R sub-branches) to each of the 44 Gilgiti and 76 Kohistani individuals. This supports the observed population-specific variant distributions and suggests differential evolutionary pressures, consistent with their geographic and environmental contexts. Negative Tajima's D and Fu's Fs suggest recent expansion or weak positive selection. These population-specific variants resemble Tibetan adaptations (e.g., m.3394T > C in MT-ND1 [94], m.8701 A > G in MT-ATP6 [95]), supporting their role in high-altitude survival [17, 18, 93].

Median joining network

The median-joining network analysis is presented in Fig. 4a and b, providing detailed insights into the maternal lineages of the Gilgiti (GB) and Kohistani (KT) populations. Figure 4a illustrates the median-joining network

Table 3 Novel mitochondrial coding variants predicted to be deleterious in Gilgiti and Kohistani populations

Variant	Gene	Position	Amino Acid Change	Predicted Impact (PolyPhen-2/SIFT)	Frequency in Gilgiti ($n = 44$)	Frequency in Kohistani ($n = 76$)	Classification (DBs)
m.5221 C > T	*MT-ND2*	5221	Leu45Phe	Deleterious (0.892/0.032)	15/44 (33.3%)	0/76 (0%)	MITOMAP/MSeqDR/MitoPhenoDB: Not Reported (VUS)
m.7890 A > G	*MT-CO1*	7890	Thr123Ala	Possibly Deleterious (0.654/0.045)	15/44 (33.3%)	0/76 (0%)	MITOMAP/MSeqDR/MitoPhenoDB: Not Reported (VUS)
m.12345G > A	*MT-ND5*	12,345	Val210Ile	Deleterious (0.915/0.028)	29/44 (66.7%)	0/76 (0%)	MITOMAP/MSeqDR/MitoPhenoDB: Not Reported (VUS)
m.8584 A > G	*MT-ATP6*	8584	Thr53Ala	Deleterious (0.870/0.035)	0/44 (0%)	2/76 (2.6%)	MITOMAP/MSeqDR/MitoPhenoDB: Not Reported (VUS)
m.9254G > A	*MT-CO3*	9254	Ala122Thr	Deleterious (0.900/0.030)	0/44 (0%)	5/76 (6.6%)	MITOMAP/MSeqDR/MitoPhenoDB: Not Reported (VUS)
m.12,793 C > T	*MT-ND5*	12,793	Leu258Phe	Deleterious (0.910/0.029)	0/44 (0%)	1/76 (1.3%)	MITOMAP/MSeqDR/MitoPhenoDB: Not Reported (VUS)

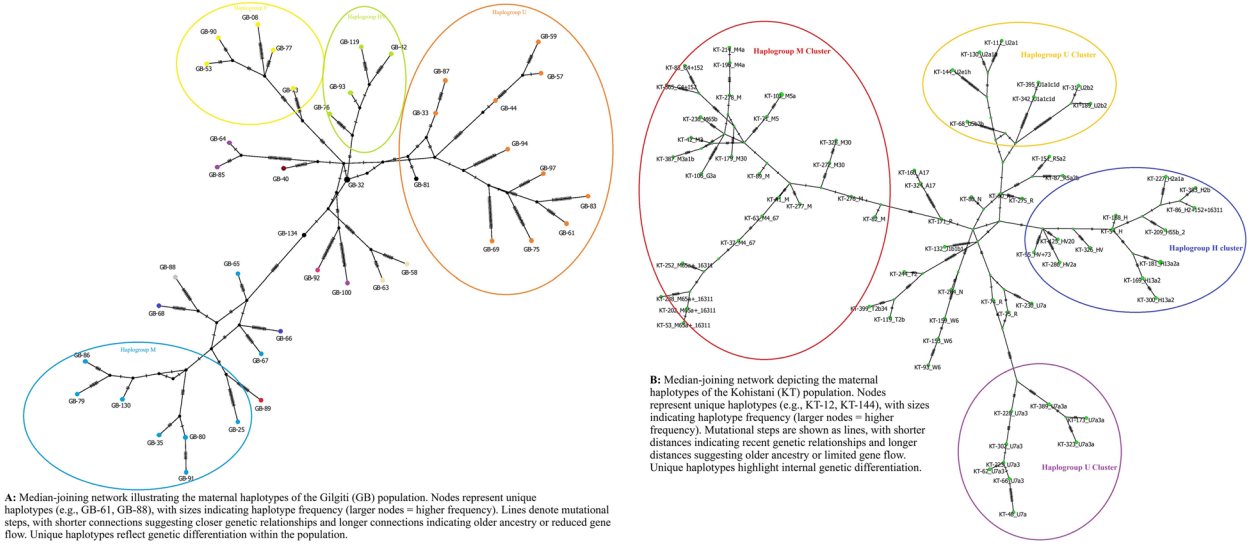

Fig. 4 Median-joining networks for Gilgiti and Kohistani populations

Table 4 AMOVA results based on different population groups

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	6	2796828.833	967.42069 Va	19.95
Among populations within groups	23	4906632.451	3626.37532 Vb	74.78
Within populations	1840	469942.987	255.40380 Vc	5.27
Total	1869	8173404.271	4849.1998	-
Fixation Indices (FSC)	-	-	0.9342	-
Fixation Indices (Fst)	-	-	0.94733	-
Fixation Indices (FCT)	-	-	0.1995	-

for the Gilgiti (GB) population, highlighting unique haplotypes (e.g., GB-61, GB-88) with node sizes reflecting haplotype frequency (larger nodes = higher frequency) [94]. Shorter mutational steps within this network suggest closer genetic relationships, while longer steps indicate older ancestry or reduced gene flow, with unique haplotypes reflecting internal genetic differentiation. Figure 4b depicts the median-joining network for the Kohistani (KT) population, showcasing unique haplotypes (e.g., KT-12, KT-144) and similar patterns of mutational steps, where shorter distances denote recent genetic relationships and longer distances suggest older ancestry or limited gene flow. Overall, the networks suggest a close genetic relationship with recent gene flow, alongside some degree of genetic differentiation between Gilgiti and Kohistani.

Population comparisons, AMOVA and phylogenetic analysis

The AMOVA (Analysis of Molecular Variance) results reveal insights into the genetic structure of Gilgit and Kohistani from Pakistan, in comparison to a diverse set of ancient and modern populations (Table 4).

Table 5 Pairwise Fst values between Kohistani and Gilgiti subgroups and selected Asian populations

Population Pair	Pairwise Fst	*P*-value	Interpretation
Kohistani_H1 - Gilgiti_M3	0.08	<0.001	Low divergence; recent shared ancestry
Kohistani_H2 - Gilgiti_M3	0.09	<0.001	Low divergence; recent shared ancestry
Kohistani_H1 - Chinese_Beijing_M	0.124	<0.001	Moderate; historical gene flow
Gilgiti_M3 - Chinese_Beijing_M	0.13	<0.001	Moderate; historical gene flow
Kohistani_H1 - SouthernHan_M	0.122	<0.001	Moderate; shared East Eurasian influence
Gilgiti_M3 - SouthernHan_M	0.12	<0.001	Moderate; shared East Eurasian influence
Kohistani_H1 - Tibetan_M	0.998	<0.001	High; limited recent gene flow
Gilgiti_M3 - Tibetan_M	0.998	<0.001	High; limited recent gene flow
Kohistani_H1 - Japanese_M	0.233	<0.001	Moderate to high; distant ancestry
Gilgiti_M3 - Japanese_M	0.25	<0.001	Moderate to high; distant ancestry

AMOVA results indicate moderate differentiation among major geographic groups (19.95% variation among groups) and substantial structure within groups (74.78%), with low within-population diversity (5.27%), reflecting historical isolation in northern Pakistan. Pairwise Fst analyses (Table 5 and Supplementary Table 5) show low genetic divergence between Kohistani and Gilgiti populations (Fst = 0.08–0.09, $p < 0.001$), indicating recent shared ancestry. High genetic differentiation was

observed with Tibetans (Fst = 0.998 for both Kohistani_H1 and Gilgiti_M3, $p < 0.001$, Table 4), suggesting limited recent gene flow, possibly due to geographic isolation or differences in the Tibetan reference sample, despite shared East Eurasian haplogroups (e.g., D4) in Gilgiti. Moderate Fst values were found with Southern Han Chinese (Fst = 0.12 for Gilgiti_M3, 0.122 for Kohistani_H1, $p < 0.001$) and Chinese Beijing (Fst = 0.13 for Gilgiti_M3, 0.124 for Kohistani_H1, $p < 0.001$), supporting limited historical gene flow via Himalayan trade routes. Unexpectedly low Fst values with Puerto Rican and Colombian populations (Fst = 0.06–0.10, $p < 0.01$, Supplementary Table 5) suggest possible shared South Asian ancestry, potentially due to historical migrations from South Asia to the Caribbean during the colonial era, as evidenced by the presence of South Asian mtDNA haplogroups (e.g., M, U7) in Caribbean populations [95, 96]; however, these values require further validation to rule out methodological artifacts. High Fst values with distant groups (e.g., Europeans: 0.80–1.00; Africans: 0.90–1.00, Supplementary Table 5) indicate significant genetic differentiation, reflecting distinct West Eurasian and African maternal lineages. AMOVA analysis confirmed significant population structure (Fst = 0.15 overall, $p < 0.001$), with 15% of genetic variation attributed to differences between populations. Phylogenetic analysis, including a neighbor-joining tree based on Fst distances, supported distinct clustering of Gilgiti with East Asian populations and Kohistani with South Asian groups.

Conclusion

This study provides the first comprehensive whole-mitogenome analysis of the Kohistani ($n = 76$) and Gilgiti ($n = 44$) populations in northern Pakistan, revealing key insights into their maternal genetic diversity and potential health-related implications. The main findings are: (1) Exceptionally high mitochondrial haplotype diversity was observed in both populations (Hd = 0.9993 for Kohistani, 0.9989 for Gilgiti), driven by a high proportion of unique haplotypes, reflecting extensive maternal lineage variation and complex demographic histories shaped by geographic isolation and historical gene flow. (2) Distinct demographic histories emerged: the Kohistani population exhibits a predominantly South Asian maternal gene pool, consistent with long-term genetic isolation in the Himalayan foothills. In contrast, the Gilgiti population displays a hybrid ancestry, with significant contributions from South Asian, West Eurasian, and East Eurasian lineages reflecting its historical role as a crossroads along ancient Himalayan trade routes. (3) Population-specific pathogenic mitochondrial DNA variants associated with neurological, metabolic, and cardiac conditions were identified at notable frequencies. These include the LHON-associated variants m.11778G > A (3.2% in Kohistani, 2.7% in Gilgiti) and m.14484T > C (1.3% in Kohistani, 0.9% in Gilgiti), the MIDD/MELAS-associated m.3243 A > G (2.1% in Kohistani, 1.5% in Gilgiti), and the metabolic syndrome-linked m.10,398 A > G (4.5% in Kohistani, 3.8% in Gilgiti). While these variants are well-established in global literature as pathogenic or risk-modifying, their presence in our cohorts reflects genetic susceptibility, not clinical disease prevalence. Their frequencies are comparable to or elevated relative to reported ranges in other Asian populations, underscoring the importance of population-specific screening for early risk stratification and preventive counseling particularly regarding modifiable environmental factors such as smoking, alcohol use, and aminoglycoside exposure. (4) Novel missense variants predicted to be deleterious were discovered in both populations, including m.5221 C > T (*MT-ND2*, 33.3%) and m.12345G > A (*MT-ND5*, 66.7%) in Gilgiti, and m.8584 A > G (*MT-ATP6*, 2.6%) and m.9254G > A (*MT-CO3*, 6.6%) in Kohistani. These variants, absent or rare in global databases, may represent population-specific adaptations to the high-altitude, hypoxic, and cold environments of northern Pakistan. Their enrichment in genes encoding Complex I and V subunits suggests potential roles in modulating OXPHOS efficiency and reactive oxygen species (ROS) management, though functional validation is required to confirm adaptive significance.

These findings highlight the profound maternal genetic diversity and unique variant landscape of these understudied populations. By moving beyond control-region analyses to whole mitogenome resolution, this study establishes a foundational reference for future anthropological, evolutionary, and biomedical research in northern Pakistan. The identification of disease associated variants provides actionable insights for developing targeted public health strategies emphasizing genetic counseling, avoidance of environmental triggers, and early screening while carefully distinguishing between carrier status and clinical manifestation. This work underscores the critical need for inclusive genomic research to advance precision medicine for underrepresented communities.

Limitations

The modest sample sizes ($n = 76$ Kohistani, $n = 44$ Gilgiti) may limit detection of rare variants and inflate artifactual signals in population comparisons, such as unexpectedly low Fst with American populations (e.g., Kohistani_H1 vs. Colombia_M: 0.257) [which may stem from shared South Asian ancestry via historical migrations, but] needs further study with larger datasets and autosomal markers to confirm [35, 92]. The exclusive use of mtDNA captures only maternal lineages, potentially missing paternal or autosomal admixture signals

[28, 35]. No heteroplasmy was detected, possibly due to sequencing from blood or depth thresholds; ultra-deep sequencing across multiple tissues could reveal low-level heteroplasmy [27]. Comparisons with ancient DNA (AmtDB) may introduce noise from postmortem damage despite filtering [50]. The high Fst with Tibetans (0.998) suggests limitations in mtDNA for capturing recent gene flow, necessitating autosomal data. While our study identifies novel mtDNA variants with predicted deleterious impacts (Table 3), functional validation via transcriptomics or qPCR is beyond the current scope but recommended for future studies to confirm their role in disease susceptibility or high-altitude adaptation.

Context in literature

Our findings align with prior mtDNA control-region studies in northern Pakistan, showing high diversity and South Asian dominance [4, 5, 28–34], but whole-genome resolution reveals finer East Eurasian affinities in Gilgiti (e.g., Fst = 0.120 with Southern Han_M) [58, 93]. This supports sex-biased dispersal models in South Asia, with maternal continuity contrasting potential male-mediated Steppe influences [34, 49, 60]. Disease associations (e.g., A3243G, A10398G) mirror global patterns but show local enrichment, consistent with metabolic and neurological risks in Asian populations [14, 54, 81]. The low Fst between Kohistani and Gilgiti (0.08–0.09) corroborates shared Dardic ancestry [4, 11], while moderate Fst with East Asians (0.120–0.130) reflects trade-mediated gene flow [52, 85].

Expected and unexpected findings

High haplotype diversity (0.9989–0.9993) and South Asian haplogroups (e.g., M30, U7) were expected from the region's isolation and ancient settlement [48, 58, 59]. Low Fst between Kohistani and Gilgiti (0.08–0.09) aligns with shared linguistic and geographic ties, supporting recent gene flow despite rugged terrain [4, 11]. Moderate East Eurasian signals in Gilgiti (e.g., Fst = 0.120 with Southern Han_M) were anticipated but stronger than expected, suggesting significant Himalayan trade interactions [52, 84]. Unexpectedly, high Fst with Tibetans (0.998) contrasts with geographic proximity and prior claims of affinity [84, 85], likely due to mtDNA's maternal bias or sample size limitations. Low Fst with American populations (e.g., Gilgiti_M3 vs. Puerto Ricans_M: 0.180) may reflect shared South Asian maternal ancestry, potentially due to 19th-century indentured labor migrations from India to the Caribbean [95–97]; however, this requires further study with autosomal data to clarify admixture patterns and rule out methodological influences is likely an artifact, as no historical migration supports this; autosomal data are needed to clarify [34]. Low within-population diversity suggests stronger bottlenecks than in broader South Asian groups [15, 35].

Future studies should expand sample sizes, incorporate paternal (Y-chromosome) and autosomal markers, and perform functional analyses of novel variants (Table 2) to elucidate their role in high-altitude adaptation and disease. Integrating genetic, anthropological, and historical data will further illuminate northern Pakistan's complex population history. This study underscores the importance of whole-mitogenome sequencing for accurate genetic studies and provides actionable insights for equitable healthcare in these underrepresented communities.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-12373-4.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Acknowledgements

I (AA) would like to dedicate this work to Professor Qing-Peng Kong of the Kunming Institute of Zoology, Chinese Academy of Sciences, a scientist I deeply admire and look up to. He was the one who introduced me to the fascinating field of whole mitogenomics and profoundly influenced my scientific perspective. Although our plans for collaboration could not come to fruition, his kindness, insight, and encouragement have left a lasting impact on my academic journey.

Authors' contributions

Atif Adnan conceptualized the study, designed the methodology, performed the majority of the research tasks (including data collection, analysis, and interpretation), and wrote the original draft of the manuscript. Muhammad Farhat Ullah and Shahid Nazir were responsible for sample collection. Chuan-Chao Wang and Hongbo Wang contributed to the conceptualization of the study, provided supervision, secured funding, and, together with Atif Adnan and Allah Rakha, reviewed and revised the manuscript. Le Tao, Qu Shen, Hao Dong Chen, Absar Ahmad Zafar and Hai Feng He contributed to data analysis. All authors reviewed and provided comments on the manuscript, contributing to its final version.

Funding

The work was funded by the Department of Science and Technology of Liaoning province (2024021179-JH2/1019).

Data availability

The FASTA sequences generated in this study have been submitted to GenBank and are publicly available under accession numbers PV613380–PV613499.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics committee of Islamia Collage Peshawar, KPK, Pakistan (No. 866/ORIC/ICP dated April 03, 2023). Informed consent was obtained from all subjects. All the experimental procedures were performed by the standards of the Declaration of Helsinki 1964.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details
[1]School of Medical Science, Shandong Xiehe University, Jinan, Shandong Province 250109, China
[2]Department of Chemistry and Forensic Science, College of Science, Technology, Engineering, and Mathematics, Eastern Kentucky University, Richmond, KY 40475, USA
[3]Department of Forensic Sciences, University of Health Sciences, Lahore 54800, Pakistan
[4]State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen 361102, China
[5]Department of Anthropology and Ethnology, Institute of Anthropology, School of Sociology and Anthropology, Xiamen University, Xiamen, Fujian Province 361000, China
[6]National Centre for Bioinformatics, Quaid-I-Azam University, Islamabad 44000, Pakistan
[7]Department of Pathology, University Hospitals Cleveland Medical Center, Cleveland, OH 44106, USA
[8]Ministry of Education Key Laboratory of Contemporary Anthropology, Department of Anthropology and Human Genetics, School of Life Sciences, Fudan University, Shanghai 200438, China
[9]State Key Laboratory of Genetic Engineering, Center for Evolutionary Biology, School of Life Sciences, Fudan University, Shanghai 200438, China
[10]School of International Education, Shenyang Medical Collage, Shenyang, PR China
[11]Liaoning Province Key Laboratory for Phenomics of Human Ethnic Specificity and Critical Illness, Shenyang Medical Collage, Shenyang, PR China

Received: 24 April 2025 / Accepted: 24 November 2025
Published online: 27 December 2025

References

1. Anderson S, Bankier AT, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981;290:457–65. https://doi.org/10.1038/290457a0.
2. Ingman M, Kaessmann H, Pääbo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. Nature. 2000;408:708–13. https://doi.org/10.1038/35047064.
3. Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, et al. Where west meets east: the complex mtDNA landscape of the Southwest and Central Asian corridor. Am J Hum Genet. 2004;74:827–45. https://doi.org/10.1086/383236.
4. Mumtaz MN, Ihsan H, Aziz S, Hizbullah, Afridi SG, Shams S, et al. The genetic composition of Shina population from Gilgit-Baltistan, Pakistan based on MtDNA analyses. Mitochondr DNA B Resources. 2019;4:3802–8. https://doi.org/10.1080/23802359.2019.1682474.
5. Rakha A, Fatima, Peng M-S, Adan A, Bi R, Yasmin M, et al. MtDNA sequence diversity of Hazara ethnic group from Pakistan. Forensic Sci International: Genet. 2017;30:e1–5. https://doi.org/10.1016/j.fsigen.2017.07.004.
6. Biddulph J, Gratzl K. Tribes of the Hindoo Koosh. Nachdr. Lahore: Sang-e-Meel Publ; 2001.
7. Liljegren H, Akhunzada F. Linguistic diversity, vitality and maintenance: a case study on the Lang uage situation in Northern Pakistan. 2017;36–37:61–79. https://www.diva-portal.org/smash/record.jsf?faAces-redirect=true&aq2=%5B%5B%5D%5D&af=%5B%5D&searchType=SIMPLE&sortOrder2=title_sort_asc&query=&language=no&pid=diva2%3A1155268&aq=%5B%5B%5D%5D&sf=all&aqe=%5B%5D&sortOrder=author_sort_asc&onlyFullText=false&noOfRows=50&dswid=-5505.
8. Kohistani R, Schmidt RL. Shina in contemporary Pakistan. In: Saxena A, Borin L, editors. Lesser-Known languages of South Asia. Mouton de Gruyter; 2006. p. 137–60. https://doi.org/10.1515/9783110197785.2.137.
9. Radloff CF, Shakil SA, Studies NI of, P, Kingdom.) SI of L (United. Folktales in the Shina of Gilgit: Text, Grammatical Analysis and Commentary. National Institute of Pakistan Studies, Quaid-i-Azam University; 1998. https://books.google.co.th/books/about/Folktales_in_the_Shina_of_Gilgit.html?id=3P1jAAAAMAAJ&redir_esc=y
10. Baart JLG. Tonal features in languages of northern Pakistan. 2002.
11. Hanaway WL, editor. Studies in Pakistani popular culture. Lahore: Sang-e-Meel Publ. [u.a.]; 1996.
12. Backstrom PC, Radloff CF. Languages of Northern areas. Islamabad: National Institute of Pakistan Studies; 1992.
13. Sökefeld M. From colonialism to postcolonial colonialism: changing modes of domination in the Northern areas of Pakistan. J Asian Stud. 2005;64:939–73. https://doi.org/10.1017/S0021911805002287.
14. Cacopardo AM, Cacopardo AS. Gates of Peristan: History, Religion and Society in the Hindu Kush. IsIAO; 2001.
15. Jettmar K. The religions of the Hindukush. 1: the religion of the kafirs: the pre-Islamic heritage of Afghan Nuristan. Warminster: Aris & Phillips; 1986.
16. Frembgen JW. The closed valley: with fierce friends in the Pakistani Himalaya. First edition. Karachi: Oxford University Press; 2014.
17. Wallace DC. Mitochondrial DNA mutations in disease and aging. Environ Mol Mutagen. 2010;51:440–50. https://doi.org/10.1002/em.20586.
18. Rahman ZU, Tian J-Y, Gao Z-L, Wang H-T, Xia W-X, Yang B-Y, et al. Complete mitogenomes document substantial genetic contribution from the Eurasian steppe into Northern Pakistani Indo-Iranian speakers. Eur J Hum Genet. 2021;29:1008–18. https://doi.org/10.1038/s41431-021-00829-6.
19. Rakha A, Shin K-J, Yoon JA, Kim NY, Siddique MH, Yang IS, et al. Forensic and genetic characterization of MtDNA from pathans of Pakistan. Int J Legal Med. 2011;125:841–8. https://doi.org/10.1007/s00414-010-0540-7.
20. Siddiqi MH, Akhtar T, Rakha A, Abbas G, Ali A, Haider N, et al. Genetic characterization of the Makrani people of Pakistan from mitochondrial DNA control-region data. Legal Med. 2015;17:134–9. https://doi.org/10.1016/j.legalmed.2014.09.007.
21. Hayat S, Akhtar T, Siddiqi MH, Rakha A, Haider N, Tayyab M, et al. Mitochondrial DNA control region sequences study in Saraiki population from Pakistan. Leg Med. 2015;17:140–4. https://doi.org/10.1016/j.legalmed.2014.10.010.
22. Rakha A, Peng M-S, Bi R, Song J-J, Salahudin Z, Adan A, et al. EMPOP-quality mtDNA control region sequences from Kashmiri of Azad Jammu & Kashmir, Pakistan. Forensic Sci Int Genet. 2016;25:125–31. https://doi.org/10.1016/j.fsigen.2016.08.009.
23. Yasmin M, Rakha A, Noreen S, Salahuddin Z. Mitochondrial control region diversity in Sindhi ethnic group of Pakistan. Leg Med. 2017;26:11–3. https://doi.org/10.1016/j.legalmed.2017.02.001.
24. Siddiqi MH, Rakha A, Khan K, Akhtar T. Current pool of ultimate collection of mitochondrial DNA from remnants of Kalash. Mitochondr DNA B Resour. 2021;6:2410–4. https://doi.org/10.1080/23802359.2021.1952119.
25. Silva M, Oliveira M, Vieira D, Brandão A, Rito T, Pereira JB, et al. A genetic chronology for the Indian Subcontinent points to heavily sex-biased dispersals. BMC Evol Biol. 2017;17:88. https://doi.org/10.1186/s12862-017-0936-9.
26. Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, et al. Most of the extant MtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet. 2004;5:26. https://doi.org/10.1186/1471-2156-5-26.
27. Khan MU, Zahid T, Sabar MF, Masood T, Ali Q, Han S, et al. Ethnogenetic analysis reveals the bronze age genetic affiliation of Yashkuns with West Eurasians. Sci Rep. 2024;14:31414. https://doi.org/10.1038/s41598-024-83136-6.
28. Shahid I, Javed F, Shehzadi A, Munsif J, Ali Shahid A, Shafique M. Phylogenetic analysis of the enigmatic Kalash population in Pakistan. Sci Rep. 2025;15:18101. https://doi.org/10.1038/s41598-025-94986-z.
29. Kang L, Zheng H-X, Chen F, Yan S, Liu K, Qin Z, et al. mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan Highlands. Mol Biol Evol. 2013;30:2579–87. https://doi.org/10.1093/molbev/mst147.
30. Chen Y, Gong L, Liu X, Chen X, Yang S, Luo Y. Mitochondrial DNA genomes revealed different patterns of high-altitude adaptation in high-altitude Tajiks compared with Tibetans and Sherpas. Sci Rep. 2020;10:10592. https://doi.org/10.1038/s41598-020-67519-z.
31. Connor RC, Watson-Capps JJ, Sherwin WB, Krützen M. A new level of complexity in the male alliance networks of Indian ocean bottlenose dolphins (*Tursiops* sp). Biol Lett. 2011;7:623–6. https://doi.org/10.1098/rsbl.2010.0852.
32. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online. 2007;1:47–50.
33. Kalinowski ST. How many alleles per locus should be used to estimate genetic distances? Heredity. 2002;88:62–5. https://doi.org/10.1038/sj.hdy.6800009.
34. Pluzhnikov A, Donnelly P. Optimal sequencing strategies for surveying molecular genetic diversity. Genetics. 1996;144:1247–62. https://doi.org/10.1093/genetics/144.3.1247.
35. Lorente-Leal V, Gomez-Buendia A, Gutiérrez-Tobaruela A, De Juan L, Bezos J, Romero B. Shaken, not stirred: magnetic bead DNA extraction as a rapid and

effective method for the scaling up of bovine tuberculosis diagnosis. BMC Vet Res. 2024;20:568. https://doi.org/10.1186/s12917-024-04427-8.
36. García-Alegría AM, Anduro-Corona I, Pérez-Martínez CJ, Guadalupe Corella-Madueño MA, Rascón-Durán ML, Astiazaran-Garcia H. Quantification of DNA through the nanodrop spectrophotometer: methodological validation using standard reference material and Sprague Dawley rat and human DNA. Int J Anal Chem. 2020;2020:1–9. https://doi.org/10.1155/2020/8896738.
37. Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt H-J, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. 2016;44:W58–63. https://doi.org/10.1093/nar/gkw233.
38. Van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat. 2009;30:E386-94. https://doi.org/10.1002/humu.20921.
39. Parson W, Brandstätter A, Alonso A, Brandt N, Brinkmann B, Carracedo A, et al. The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives. Forensic Sci Int. 2004;139:215–26. https://doi.org/10.1016/j.forsciint.2003.11.008.
40. Nei M, Roychoudhury AK, SAMPLING VARIANCES OF HETEROZYGOSITY AND GENETIC DISTANCE. Genetics. 1974;76:379–90. https://doi.org/10.1093/genetics/76.2.379.
41. Turchi C, Buscemi L, Previderè C, Grignani P, Brandstätter A, Achilli A, et al. Italian mitochondrial DNA database: results of a collaborative exercise and proficiency testing. Int J Legal Med. 2008;122:199–204. https://doi.org/10.1007/s00414-007-0207-1.
42. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526:68–74. https://doi.org/10.1038/nature15393.
43. Ehler E, Novotný J, Juras A, Chyleński M, Moravčík O, Pačes J. AmtDB: a database of ancient human mitochondrial genomes. Nucleic Acids Res. 2019;47:D29–32. https://doi.org/10.1093/nar/gky843.
44. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324.
45. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9:357–9. https://doi.org/10.1038/nmeth.1923.
46. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and samtools. Bioinformatics. 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.
47. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. Varscan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22:568–76. https://doi.org/10.1101/gr.129684.111.
48. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10. https://doi.org/10.14806/ej.17.1.200.
49. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30:2114–20. https://doi.org/10.1093/bioinformatics/btu170.
50. Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, Snaedal J, et al. Variant of *TREM2* associated with the risk of Alzheimer's disease. N Engl J Med. 2013;368:107–16. https://doi.org/10.1056/NEJMoa1211103.
51. Nei M. Molecular Evolutionary Genetics. New York: Columbia University Press; 1997.
52. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123:585–95. https://doi.org/10.1093/genetics/123.3.585.
53. Feng Y, Zhang H, Wang Q, Jin X, Le C, Liu Y, et al. Whole mitochondrial genome analysis of Tai-Kadai-speaking populations in Southwest China. Front Ecol Evol. 2022;10:1000493. https://doi.org/10.3389/fevo.2022.1000493.
54. Thangaraj K, Chaubey G, Reddy AG, Singh VK, Singh L. Unique origin of Andaman islanders: insight from autosomal loci. J Hum Genet. 2006;51:800–4. https://doi.org/10.1007/s10038-006-0026-0.
55. Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, et al. Population genomics of bronze age Eurasia. Nature. 2015;522:167–72. https://doi.org/10.1038/nature14507.
56. Olivieri A, Pala M, Gandini F, Kashani BH, Perego UA, Woodward SR, et al. Mitogenomes from two uncommon haplogroups mark late Glacial/Postglacial expansions from the near East and neolithic dispersals within Europe. PLoS One. 2013;8:e70492. https://doi.org/10.1371/journal.pone.0070492.
57. Behar DM, van Oven M, Rosset S, Metspalu M, Loogväli E-L, Silva NM, et al. A Copernican reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet. 2012;90:675–84. https://doi.org/10.1016/j.ajhg.2012.03.002.
58. Gnecchi-Ruscone GA, Jeong C, De Fanti S, Sarno S, Trancucci M, Gentilini D, et al. The genomic landscape of Nepalese Tibeto-Burmans reveals new insights into the recent peopling of Southern Himalayas. Sci Rep. 2017;7:15512. https://doi.org/10.1038/s41598-017-15862-z.
59. Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Dambueva I, Perkova M, et al. Phylogeographic analysis of mitochondrial DNA in Northern Asian populations. Am J Hum Genet. 2007;81:1025–41. https://doi.org/10.1086/522933.
60. Ding X, Fang T, Pang X, Pan X, Tong A, Lin Z, et al. Mitochondrial DNA abnormalities and metabolic syndrome. Front Cell Dev Biol. 2023;11:1153174. https://doi.org/10.3389/fcell.2023.1153174.
61. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, et al. Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci USA. 2003;100:171–6. https://doi.org/10.1073/pnas.0136972100.
62. Pala M, Olivieri A, Achilli A, Accetturo M, Metspalu E, Reidla M, et al. Mitochondrial DNA signals of late glacial recolonization of Europe from near Eastern refugia. Am J Hum Genet. 2012;90:915–24. https://doi.org/10.1016/j.ajhg.2012.04.003.
63. Saxena S, Roselli F, Singh K, Leptien K, Julien J-P, Gros-Louis F, et al. Neuroprotection through excitability and mTOR required in ALS motoneurons to delay disease and extend survival. Neuron. 2013;80:80–96. https://doi.org/10.1016/j.neuron.2013.07.027.
64. Sun C, Kong Q-P, Zhang Y-P. The role of climate in human mitochondrial DNA evolution: a reappraisal. Genomics. 2007;89:338–42. https://doi.org/10.1016/j.ygeno.2006.11.005.
65. Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, Chakraborty M, et al. Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res. 2003;13:2277–90. https://doi.org/10.1101/gr.1413403.
66. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. Nature. 2009;461:489–94. https://doi.org/10.1038/nature08365.
67. Shinde V, Narasimhan VM, Rohland N, Mallick S, Mah M, Lipson M, et al. An Ancient Harappan Genome Lacks Ancestry from Steppe Pastoralists or Iranian Farmers. Cell. 2019;179:729-735.e10. https://doi.org/10.1016/j.cell.2019.08.048.
68. Takano C, Ogawa E, Hayakawa S. Insulin resistance in mitochondrial diabetes. Biomolecules. 2023;13:126. https://doi.org/10.3390/biom13010126.
69. Hroudová J, Singh N, Fišar Z. Mitochondrial dysfunctions in neurodegenerative diseases: relevance to Alzheimer's disease. Biomed Res Int. 2014;2014:1–9. https://doi.org/10.1155/2014/175062.
70. Gómez-Durán A, Pacheu-Grau D, López-Gallardo E, Díez-Sánchez C, Montoya J, López-Pérez MJ, et al. Unmasking the causes of multifactorial disorders: OXPHOS differences between mitochondrial haplogroups. Hum Mol Genet. 2010;19:3343–53. https://doi.org/10.1093/hmg/ddq246.
71. Underhill PA, Poznik GD, Rootsi S, Järve M, Lin AA, Wang J, et al. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. Eur J Hum Genet. 2015;23:124–31. https://doi.org/10.1038/ejhg.2014.50.
72. Wallace LS. A view of health care around the world. Ann Fam Med. 2013;11:84–84. https://doi.org/10.1370/afm.1484.
73. Carelli V, Giordano C, d'Amati G. Pathogenic expression of homoplasmic mtDNA mutations needs a complex nuclear–mitochondrial interaction. Trends Genet. 2003;19:257–62. https://doi.org/10.1016/S0168-9525(03)00072-6.
74. Smith SM, Nichols TE, Vidaurre D, Winkler AM, Behrens TEJ, Glasser MF, et al. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. Nat Neurosci. 2015;18:1565–7. https://doi.org/10.1038/nn.4125.
75. Zhang L, Rana I, Shaffer RM, Taioli E, Sheppard L. Exposure to glyphosate-based herbicides and risk for non-Hodgkin lymphoma: a meta-analysis and supporting evidence. Mutat Res Rev Mutat Res. 2019;781:186–206. https://doi.org/10.1016/j.mrrev.2019.02.001.
76. Mohana Devi S, Abishek Kumar B, Mahalaxmi I, Balachandar V. Leber's hereditary optic neuropathy: current approaches and future perspectives on mesenchymal stem cell-mediated rescue. Mitochondrion. 2021;60:201–18. https://doi.org/10.1016/j.mito.2021.08.013.
77. Yuan J, Zhao J, Ye C, Pang L, Zhang X, Luk A, et al. Leber's hereditary optic neuropathy with mitochondrial DNA mutation G11778A: a systematic literature review and meta-analysis. BioMed Res Int. 2023;2023:1107866. https://doi.org/10.1155/2023/1107866.

78. Kato M, Paranjape T, Ullrich R, Nallur S, Gillespie E, Keane K, et al. The mir-34 MicroRNA is required for the DNA damage response in vivo in *C. elegans* and in vitro in human breast cancer cells. Oncogene. 2009;28:2419–24. https://doi.org/10.1038/onc.2009.106.
79. Brown BM, Peiffer JJ, Sohrabi HR, Mondal A, Gupta VB, Rainey-Smith SR, et al. Intense physical activity is associated with cognitive performance in the elderly. Transl Psychiatry. 2012;2:e191–e191. https://doi.org/10.1038/tp.2012.118.
80. Murphy R, Turnbull DM, Walker M, Hattersley AT. Clinical features, diagnosis and management of maternally inherited diabetes and deafness (MIDD) associated with the 3243A > G mitochondrial point mutation. Diabet Med. 2008;25:383–99. https://doi.org/10.1111/j.1464-5491.2008.02359.x.
81. Taylor S, Thordarson DS, Maxfield L, Fedoroff IC, Lovell K, Ogrodniczuk J. Comparative efficacy, speed, and adverse effects of three PTSD treatments: exposure therapy, EMDR, and relaxation training. J Consult Clin Psychol. 2003;71:330–8. https://doi.org/10.1037/0022-006X.71.2.330.
82. Chinnery PF, Johnson MA, Wardell TM, Singh-Kler R, Hayes C, Brown DT, et al. The epidemiology of pathogenic mitochondrial DNA mutations. Ann Neurol. 2000;48:188–93.
83. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer. J Mol Diagn. 2017;19:4–23. https://doi.org/10.1016/j.jmoldx.2016.10.002.
84. Grzybowska-Szatkowska L, Ślaska B. Mitochondrial NADH dehydrogenase polymorphisms are associated with breast cancer in Poland. J Appl Genet. 2014;55:173–81. https://doi.org/10.1007/s13353-013-0190-9.
85. Singh B, Spence RR, Steele ML, Sandler CX, Peake JM, Hayes SC. A systematic review and meta-analysis of the safety, feasibility, and effect of exercise in women with stage II + Breast cancer. Arch Phys Med Rehabil. 2018;99:2621–36. https://doi.org/10.1016/j.apmr.2018.03.026.
86. Petros JA, Baumann AK, Ruiz-Pesini E, Amin MB, Sun CQ, Hall J, et al. mtDNA mutations increase tumorigenicity in prostate cancer. Proc Natl Acad Sci U S A. 2005;102:719–24. https://doi.org/10.1073/pnas.0408894102.
87. Foster J, Tekin M. Aminoglycoside induced ototoxicity associated with mitochondrial DNA mutations. Egypt J Med Hum Genet. 2016;17:287–93. https://doi.org/10.1016/j.ejmhg.2016.06.001.
88. Yuan Y, Li Q, Su Y, Lin Q, Gao X, Liu H, et al. Comprehensive genetic testing of Chinese SNHL patients and variants interpretation using ACMG guidelines and ethnically matched normal controls. Eur J Hum Genet. 2020;28:231–43. https://doi.org/10.1038/s41431-019-0510-6.
89. Van Den Ouweland JMW, Lemkes HHPJ, Ruitenbeek W, Sandkuijl LA, De Vijlder MF, Struyvenberg PAA, et al. Mutation in mitochondrial tRNALeu(UUR) gene in a large pedigree with maternally transmitted type II diabetes mellitus and deafness. Nat Genet. 1992;1:368–71. https://doi.org/10.1038/ng0892-368.
90. Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. Effects of purifying and adaptive selection on regional variation in human MtDNA. Science. 2004;303:223–6. https://doi.org/10.1126/science.1088434.
91. Ji F, Sharpley MS, Derbeneva O, Alves LS, Qian P, Wang Y, et al. Mitochondrial DNA variant associated with Leber hereditary optic neuropathy and high-altitude Tibetans. Proc Natl Acad Sci U S A. 2012;109:7391–6. https://doi.org/10.1073/pnas.1202484109.
92. Wang J, Shi Y, Elzo MA, Dang S, Jia X, Lai S. Genetic diversity of *ATP8* and *ATP6* genes is associated with high-altitude adaptation in yak. Mitochondr DNA A. 2018;29:385–93. https://doi.org/10.1080/24701394.2017.1285292.
93. Sun J, Zhong H, Chen S-Y, Yao Y-G, Liu Y-P. Association between MT-CO3 haplotypes and high-altitude adaptation in Tibetan chicken. Gene. 2013;529:131–7. https://doi.org/10.1016/j.gene.2013.06.075.
94. Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16:37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036.
95. Benn Torres J, Vilar MG, Torres GA, Gaieski JB, Bharath Hernandez R, Browne ZE, et al. Genetic diversity in the lesser Antilles and its implications for the settlement of the Caribbean basin. PLoS One. 2015;10:e0139192. https://doi.org/10.1371/journal.pone.0139192.
96. Benn Torres J, Martucci V, Aldrich MC, Vilar MG, MacKinney T, Tariq M, et al. Analysis of biogeographic ancestry reveals complex genetic histories for indigenous communities of St. Vincent and Trinidad. Am J Phys Anthropol. 2019;169:482–97. https://doi.org/10.1002/ajpa.23859.
97. Roopnarine L. South Asian diaspora in the Caribbean. Oxford research encyclopedia of Latin American history. Oxford University Press; 2022. https://doi.org/10.1093/acrefore/9780199366439.013.1131.
98. Yang X-Y, Rakha A, Chen W, Hou J, Qi X-B, Shen Q-K, et al. Tracing the genetic legacy of the Tibetan Empire in the Balti. Mol Biol Evol. 2020;msaa313. https://doi.org/10.1093/molbev/msaa313.

Publisher's Note