scientific reports

Check for updates

OPEN

Ancient mitogenomes from Neolithic, megalithic and medieval burials suggest complex genetic history of Kashmir valley, India

Aparna Dwivedi[1,2,6], Lomous Kumar[1,6], Snigdha Konar[1,2], Nagarjuna Pasupuleti[3], Sanjay Kumar Singh Gahlaud[1,2], Richa Rajpal[1,2], Mohammad Ajmal Shah[5], Mumtaz A. Yatoo[5], Sachin Kumar[1,2], Shiv Kumar Patel[4], Ningombam Somorjit Singh[4] & Niraj Rai[1,2]✉

South Asia is rich in cultural and genetic diversity; however, it is hardly represented in the blooming field of archaeogenetics. The Neolithic site of Burzahom is of high cultural value and archaeological importance and is one of the earliest human settlements in the Kashmir Valley with numerous evidence of migration and cultural assimilation. In our current study, we have reconstructed for the first time the complete mitogenomes of Neolithic, megalithic and medieval individuals from the Burzahom archaeological site in Kashmir. Our findings suggest that Neolithic and Megalithic periods were characterized by predominantly local genetic influence on the maternal gene pool, with some evidence of genetic contact with the Iron Age Swat Valley. While medieval populations showed clear signs of genetic contacts with Swat Valley historical and Central Asian Bronze age populations. Interestingly, Bayesian evolutionary analysis suggests an affinity of one of the medieval samples with a medieval sample from Roopkund Lake; the finding will be more conclusive with more sample evidence. In summary, we propose that the genetics of Neolithic, megalithic and medieval Kashmir agree well with the archaeological evidence of cultural contacts with the Swat Valley and Central Asia.

Keywords Neolithic, Swat valley, Mitogenomes, Bayesian, Kashmir

The Neolithic site of Burzahom (Fig. 1a) (34° 10 'N; 74° 52 'E) evidences one of the earliest human habitations in Kashmir valley, India. The site is situated in the valley between Himalayas and the Pir Panjal range. The valley of Kashmir finds its first mention in the Nilamatpurana, in which valley was first mentioned as a vast lake known as Satisar[1]. Kalhana's Rajatrangi mentions that the valley was named after sage Kashyap as Kashmir[2]. Geological evidence suggests a major earthquake about 85,000 years ago that created a gorge that drained the lake and the lake sediments are now seen as karewas[3–6]. Burzahom archaeological site was discovered by Helmut de Terra in 1935 and T.T. Paterson of the Yale-Cambridge expedition when they investigated the Himalayan glaciation and evidence of early humans in Kashmir. However, an extensive excavation took place under Mr. T. N. Khazanchi between 1960 and 1971[7]. The Neolithic evidences in Kashmir exhibit similarities with several adjacent regions viz., Sarai Kola in Potwar Plateau, Ghalghai and Leobnar in Swat Valley and Karua in the Tibetan Plateau[8,9]. The word "Burzahom" is derived from a tree species called "birch" in Kashmiri language that grows in the Himalayas. The traces of trees were found in excavated dwelling areas as roofing material, indicating its presence even in pre-historic Neolithic periods[6,7].

The site of Burzahom showcases the stages of development of farming, thus identified as one of the rarest and richest Neolithic cultures in South Asia. Spanning from around 2900 BCE to around fifth century CE, Burzahom archeological site exhibits four phases of human habitation, Neolithic I (Aceramic), Neolithic II (Ceramic), Megalithic and Early Historic[10]. The differences between the Neolithic of Kashmir and the plains of India have

[1]DST-Birbal Sahni Institute of Palaeosciences (BSIP), 53, University Road, Lucknow, Uttar Pradesh 226007, India. [2]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India. [3]CSIR-Centre for Cellular and Molecular Biology (CCMB), Hyderabad 500007, India. [4]Anthropological Survey of India (AnSI), Kolkata 700091, India. [5] Centre of Central Asian Studies, University of Kashmir, Srinagar, Kashmir 190006, India. [6]Aparna Dwivedi and Lomous Kumar have contributed equally ✉email: nirajrai@bsip.res.in

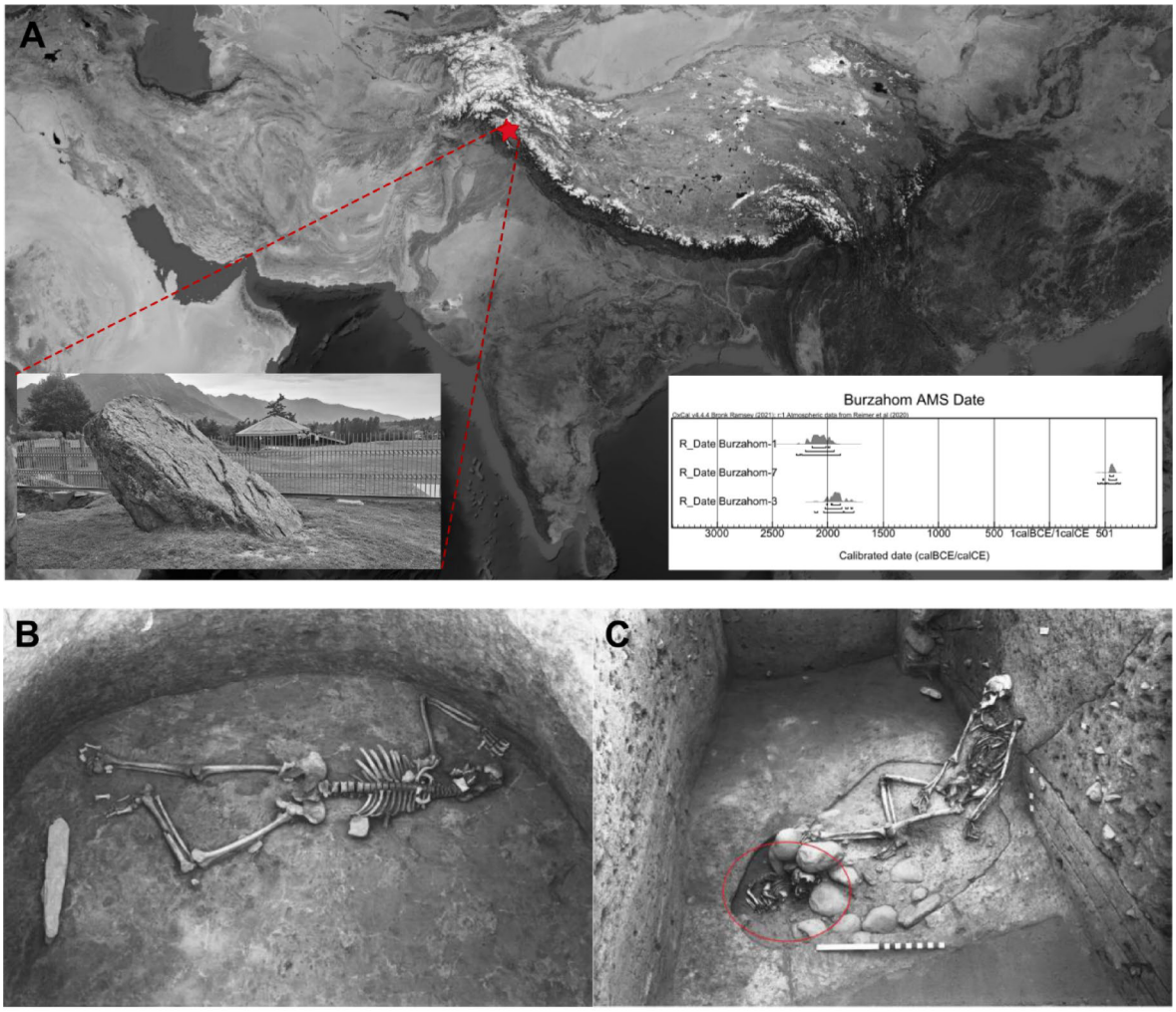

Fig. 1. Sampling location: (**A**) Archaeological site of Burzahom in Kashmir, India (left panel inset is the image of actual excavation site, right panel is calibrated AMS dates), (**B**) Skeletal remains from one of the burial, **C.** Skeleton of a child (red circle) alongside the female skeleton.

always been under hot debate[11–13]. The site shares cultural traits with Europe, West Asia, and Central Asia[11,14]. The Neolithic material culture of Kashmir has also been compared to that of Northern China, Siberia, Mongolia, and Manchuria[10,11,15]. The material culture of recently unearthed sites in the Baramulla District of Kashmir is comparable to that of sites found in northern Pakistan, as well as sites of Burzahom, Gufkral, and Kanispora in Kashmir. The similarities are mostly in ceramics, stone tools, and plant remnants from pits[11,16]. Resemblance can also be found in the types of pottery, decorations, mat impressions, graffiti, perforated pottery, small burnished or grey ware pots, stone tools, rectangular and oval double-notched harvesters, pits for dwellings, schist discs, terracotta bobbins, spindle whorls, and other artefacts found in the material cultures of Neolithic Kashmir and the Swat Valley in Pakistan. These commonalities suggest a shared cultural background and might be attributed to elements such as commerce, exchange, and seasonal migration[11]. Situated in the middle of a larger region, Baramulla might serve as a link between the homologous materials discovered in Burzahom, Gufkral, and Pakistan[11,17]. It is a key possibility that people travelled over the Himalayas due to its strategic location on the Jhelum Valley trade route and contributed to the development of separate cultural complexes in Kashmir, Pakistan, China, and Central Asia[11]. The Indus Valley Civilization (IVC) mature phase flourished in India from around 2600 to 1900 BCE[18]. Burzahom exhibits several similarities as well as differences from this contemporary culture. The Kashmir Neolithic pottery is dated to the last quarter of the fourth millennium BCE, corresponding with the Kot Dijian period at Harappa, according to recent findings at Kanishkapura (Kanispora) in Kashmir. The "horned deity", typical of early Harappan sites (Period I at Rehman Dheri) is depicted on pots during this era. Dish-on-Stand, perforated jars, blades, carnelian beads, copper-bronze objects, Kot Dijian type pottery, female terracotta figurines with thin waist and broad hips (Gumla and Rehman Dheri) reveal similarity between the Neolithic periods of Kashmir and the Harappans indicating cultural exchange in the fourth millennium BCE. Pit houses and polished tools are examples of Neolithic features seen at early and mature Harappan sites, such as Rohira, Kunal, and Bhirrana. These features are comparable to those seen in Burzahom's aceramic and early ceramic phases[19]. The first burial practice from Kashmir came into light through Burzahom (Fig. 1b, c).

Here the most interesting part is that the animals were buried along with humans, the burials are both primary and secondary. The skeletons were balmed with red ochre. The burials reveal trepanation in the skull. There were seven complete and four incomplete indications of cranial surgery, known as trepanation[7]. Burzahom revealed ten human burials, three from the Period III- Megalithic period and seven belonging to period II of the Neolithic period.

Although there are enough archaeological studies and ample evidence pointing to a rich cultural heritage of the Burzahom site as well as its contacts and exchanges with the Harappan culture, Pakistan, Tibet and the West, genetic evidence is overall lacking. Previous genetic studies of contemporary human populations in the region have primarily suggested their cosmopolitan nature, with both local and West Eurasian lineages predominating[20–23]. When it comes to the overall representation of South Asia in the field of ancient genomics, this vast region with its rich cultural diversity still lags behind Africa, Europe and other Asian regions. Despite the harsh climatic conditions for obtaining high-quality DNA from ancient remains, only recently a handful of genetic studies have come to light[24,25], most studies from India deal with ancient mitochondrial genomes due to the high copy number[26,27]. The Neolithic site of Burzahom in the Kashmir Valley of India is an archaeological site of great importance in terms of cultural continuity, diversity and external contacts. This study is the first of its kind that attempts to uncover the genetic architecture of this site from the Neolithic to the Middle Ages. In this study, we recovered ten human skeletal remains from the Neolithic site of Burzahom and two from medieval sites and successfully generated good quality mitogenomes from four (one Neolithic, one megalithic and two medieval) remains. Our study will further pave the way for the advancement of the field of archaeogenetics in South Asia, with insightful results and many more questions to be addressed.

Material and methods

DNA extraction

We carried out the ancient DNA investigation on twelve human skeletal samples. Out of twelve samples, six were petrous, three were long bone and three were tooth (Supplementary Table S1). We successfully extracted DNA from all samples. We processed all samples completely in a clean room in the ancient DNA laboratory at DST-Birbal Sahni Institute of Palaeosciences, Lucknow, India. We did the DNA extraction process in two batches to assure the authenticity. We carried out all the processes with extreme precautions to prevent contamination. Each extraction was done with control samples to take care of errors involved in processing. After decontaminating the human bone samples with UV treatment and removing the upper layer of 3 mm by sandpaper, we prepared a fine bone powder of up to 50 mg. In case of cranial fragment, we isolated the petrous and milled in tissue-lyser to avoid contamination. We did the DNA extraction using bone powder by silica bead method, which is efficient to retrieve ancient DNA by minimizing the co-extraction of PCR inhibitors[20,21]. For extraction, we first washed the bone powder in the extraction buffer, then dissolved it in a 4 ml extraction buffer and incubated overnight (approx. 12 h) at 42 °C. The extraction buffer contained 0.5 M EDTA pH 8.0, 0.5% N-laurylsarcosine, and 0.2 mg/ml proteinase K. We prepared the silica beads to capture the DNA fragments mixing it with a binding buffer. The binding buffer contained 5 M GuHCl, 30% 2-propanol, 5 M NaCl and 3 M Na Acetate. After incubating it for one hour, we collected the pellet and washed it twice with freshly prepared 80% chilled ethanol. Finally, the ethanol was dried and DNA dissolved in Tris EDTA buffer. Ancient DNA library preparation steps involved the end-repair followed by purification of products using silica based columns, adapter ligation again followed by purification step, fill-in reaction and indexing/barcoding[22]. The adapters incorporated are universal to both ends of any DNA molecules and specific to the sequencer. These adapters provide binding sites for the primers that are used for PCR amplification and sequencing on the NGS platforms. Samples were bar-coded/indexed to assign the correct identifying markers to each sample during the bioinformatics analysis downstream while preparing the library[22]. As the quantity of DNA extracted from the sample is marginal and cannot be visualized by any analytical method, an enzymatic reaction (Polymerase Chain Reaction or PCR) was run to 'amplify' the DNA fragment of interest (marker). Then amplified DNA was purified and quantified using bio-analyzer to check their size distribution as well as quantity in each sample. Sequencing was carried out on an Illumina Novaseq 6000 platform at CSIR-CCMB, Hyderabad.

Sequence mapping, consensus calling and mitochondrial DNA haplogroup determination

We used EAGER v1.92.50 with default parameters, a pipeline specially designed to deal with ancient DNA data[23] to process the raw reads. Quality assessment was performed with FastQC software[24]. The adapters were trimmed with AdapterRemoval v2.2.0[25]. Reads shorter than 30 bp were disregarded and subsequently BWA v0.7.12[26] was used to align the reads to the Revised Cambridge Reference Sequences (rCRS)[24] with seed disabled (–l 16,500). The duplicate reads were removed by the DeDup v0.12.7[23]. Ancient DNA authenticity was done using damage pattern inferred from mapDamage2[27] tool along with posterior estimates using MCMC algorithm. To retrieve the mitochondrial consensus sequence from BAM files, we used the Schmutzi pipeline[28]. First, BAM files were sorted using SAMtools[29], followed by MD tagging using the *Calmd* function. Subsequently, MD-tagged BAM files were indexed using the SAMtools index[29]. We first used the contDeam.pl script to estimate initial contamination and endogenous deamination rates. Subsequently, the main iterative procedure of the Schmutzi pipeline[28] was performed first without prediction of contaminants and then with prediction to finally obtain the consensus sequences of both contaminants and endogenous DNA. The final output of the pipeline provides both endogenous and contaminant consensus sequences along with a log file containing information about the log likelihood of each base calculated using the Bayesian method. We obtained the final endogenous Fasta sequence at four quality cutoff scores (–q 10, –q 20, –q 30 and –q 50).

Mitochondrial haplogroup determination and geographical distribution

To determine the mitochondrial haplogroup, the reconstructed fasta sequences at each quality factor were used in HaploGrep2 (v2.4.0)[30]. The geographical distribution of mitochondrial haplogroups found in individuals in ancient Kashmir was determined using a custom R script and using interpolation with the Autokrige function of the Automap package[31] and sf package[32] in R. The isofrequency map for individual haplogroups was plotted in ggplot2[33] in R[34].

Phylogenetic tree and haplotype network

To construct Maximum-Likelihood tree using combined alignment fasta of ancient Kashmir and reference sequences of individual haplogroups we used ML method implemented in MEGA 11[35]. The variants located in regions 16519, 16180–16193 and 310–315 and the AC indels in regions 515–522 were not included in the subsequent analysis. The reference sample sequences used for Phylogenetic tree construction are listed in the supplementary table (Table S2). MUSCLE algorithm implemented in MEGA 11 was used for alignment. A phylogenetic tree for all observed mtDNA haplogroups was constructed using the Maximum- Likelihood approach and 500 bootstrap replicates for testing the inferred tree. A Hasegawa-Kishino-Yano model with Gamma distributed rates and invariable sites (HKY + I + G) were used as the nucleotide substitution model and Maximum Parsimony (MP) tree was used as the initial tree. A Median-Joining (MJ) haplotype network of mtDNA haplogroups were constructed using the POPART[36] package in R[34].

Evolutionary analysis using Bayesian approach

Sequence alignment of ancient Kashmiri individuals and modern and ancient references using MUSCLE algorithm was performed in MEGA11[35]. The alignment file was cleaned for gaps and missing sites, for all the haplogroup alignments, and was used to infer the Bayesian phylogenetic tree in BEASTv2.6.7[37]. We first determined the best partitioning scheme for our combined data using PartitionFinder v2.1.1[38] and fitted the substitution model using jModelTest2[39]. We used a strict coalescent constant population molecular clock as a tree operation and performed three individual runs of 50,000,000 MCMC iterations, with the first 5,000,000 iterations discarded as burn-in and a sampling frequency of 10,000. Tracer v1.7[40] was used to check the parameter traces for convergence of MCMC runs and effective sample size. Log files and tree files from three individual runs were combined in LogCombiner v2.6.7[41]. The summary tree was created using TreeAnnotator[41]. Figtree was used to represent the final phylogenetic tree.

AMS dating of samples and estimation of molecular divergence time

AMS dating of bone samples was performed on bone pieces by collagen extraction and purification[42]. To do this, we cut the well-preserved bones into small pieces of up to 0.5 g after cleaning it with sandpaper to remove impurities. We then placed the bone linings in 0.2 M HCl at room temperature for gradual demineralization and renewed the acid every day until the bone transformed into pseudomorphs that appeared translucent and soft, indicating complete decalcification. The duration of decalcification varies depending on bone size and density. We then rinsed the samples with water to remove the acid. Next, we treated the bone with 0.1 M NaOH to remove lipids, humic acids, and other base-soluble contaminants. The samples were then soaked in water daily for a week to neutralize the pH. Finally, we freeze-dried the samples. Only highly pure deionized MilliQ water was used here. The quality of the samples was checked and samples with a C:N ratio between 2.8 and 3.6 were further processed for AMS radiocarbon measurement[43]. The prepared samples were weighed and packed into tin boats and then introduced to elemental analyzer (EA), where the sample was flash combusted and the evolved gas passed through a sequence of chemicals. The pure N_2, CO_2 gases were separated by gas chromatographic column and finally measured on the thermally coupled detector. The measured gases moved to coupled graphitization unit for graphite target preparation for AMS radiocarbon measurement. The CO_2 were loaded into reactors, where it reacted with H_2 gas at 550 °C to form graphite pellets[44]. The samples were graphitized using 'Automated Graphitization Equipment 3' (Ionplus AG, Switzerland) coupled with Vario Isotope Select' Elemental Analyzer (Elementar Analysensysteme GmbH, Germany) present in the Radiochronology and Isotopic Characterization Laboratory of Birbal Sahni Institute of Palaeosciences (BSIP), Lucknow. The graphitized samples were analyzed by Isotoptech Zrt., Debrecen, Hungary. For the accuracy and precision of the radiocarbon dates, prepared graphite of international standards (IAEA-C3 (Cellulose) and IAEA-C5 (Two Creek Wood) were also measured which was observed within the range of the consensus values.

Molecular dating for evolutionary events was performed using TMRCA (Time to Most Recent Common Ancestor) method of haplotypes. TMRCA of the haplotypes observed in ancient and medieval Kashmir samples were calculated using the Rho(ρ) statistic implemented in the NETWORK tool (https://www.fluxus-engineering.com/sharenet.htm). Time was calculated from one of the ancestral nodes in the haplotype networks and two or more descending haplotypes using a generation time of 30 years. The variance was calculated using the Sigma(σ) statistic as defined by Torroni et al. (1998). Principal component analysis with mitochondrial haplogroup distribution was done using R package *prcomp*[34].

Results

Ancient and medieval mitogenomes and haplogroup distribution

To reconstruct the mitochondrial phylogeny and evolutionary history of ancient and medieval human populations of Kashmir (Fig. 1a) we analyzed 12 samples ranging from Neolithic, early historic and medieval periods. Based on the sequence quality (Supplementary Table S1), we could successively reconstruct complete mitogenomes of 4 out of 12 samples (one Neolithic period, one Megalithic and two Medieval individuals). The alignments for these four samples contained 44077–92052 unique mapped reads, mean coverage of 125.33X-247.15X and having a mean mapping quality of 35.86–36.91 (Supplementary Table S1). Damage patterns derived from mapDamage2

analysis show increased C-to-T substitution frequencies at read ends (i.e., terminal cytosine deamination) of authenticated recovered aDNA sequences (Supplementary Figs. S1–S4). Contamination estimates as inferred by contDeam for Neolithic sample Burzahom3 (BR3) was 0.035 (95% CI 0.03–0.04), for megalithic Burzahom7 (BR7) was 0 (95% CI 0–0.005) (Supplementary Table S4). Final contamination estimates by Schmutzi for both neolithic BR3 and megalithic BR7 samples was 0.01 (95% CI 0–0.02). The raw contamination estimates (by contDeam.pl) for medieval samples Med1 and Med2 were 0.105 (95% CI 0.085–0.125) and 0.145 (95% CI 0.125–0.165) respectively. The final Schmutzi estimates for both Med1 and Med2 were 0.01(95% CI 0–0.02). The radiocarbon dates (AMS based) of four samples (one Neolithic, One Megalithic and two Medieval individuals) are listed in supplementary table S6.

The mitochondrial haplogroups were determined for neolithic, megalithic and medieval samples at four quality cutoff values of consensus calling by Schmutzi (Q10, Q20, Q30 and Q50). We could obtain consistent mitochondrial haplogroups at four quality values. For neolithic sample BR3, the mtDNA haplogroup assigned was M65a + @16311 with haplogrep quality of 0.945 and for megalithic sample BR7 the inferred haplogroup was U2b2 with quality score of 0.85. The inferred mitochondrial haplogroups for medieval sample Med1 was M30 + 16234 (score 0.9479) and for sample Med2 was W4. Although we also obtained the haplogroup (R8a1) of another neolithic sample BR2(Burzahom2) at Schmutzi quality cutoff of Q10, but its haplogrep score was too low (0.5385). Hence, we didn't include this neolithic sample in all downstream analyses (Supplementary Table S3).

In the Principal Component Analysis (PCA) with mtDNA haplogroup distribution the Neolithic sample (Burzahom3) clustered with Scheduled caste Kashmir, Khatri and Tajik samples (Supplementary Fig. S6a). The Megalithic sample (Burzahom7) was near the cluster of Kashmiri Muslim, Balochi, Brahmin_UP and Reddy samples from South Asia and Tajik sample from Central Asia (Supplementary Fig. S6b). The first Burzahom Medieval sample (Med1) was near the cluster having Scheduled caste Kashmir, Rajput_Kashmir, Gujjar, Brahmin_UP, Palestinian and Roopkund_Medieval samples (Supplementary Fig. S6c). Whereas, the second Medieval sample was lying near Sikh_Jatt and Sindhi samples (Supplementary Fig. S6d).

Evolutionary phylogeny and divergence pattern of neolithic sample

We included in our analysis 35 modern (Supplementary Table S2) and 7 ancient individuals affiliated to mitochondrial haplogroup M65 in order to infer genetic relationship of BR3 (Burzahom neolithic) in context of both modern and ancient populations. In terms of prevalence, mtDNA haplogroup M65 is predominantly present in Kashmir (0.23), followed by Pakistan (0.2) and Maharashtra in India (0.13) (Fig. 2a). It is also prevalent with noticeable frequency among Tajiks of Pamir plateau (0.10). We further evaluated this quantitative distribution of M65 haplotypes in terms of genetic relationship using Maximum-likelihood based Phylogenetic approach. In the inferred ML phylogeny (with 1000 bootstrap replicates; HKY + I substitution model) BR3 sample shared a deeply rooted major clade which have modern individuals from Kashmir and Tajikistan along with Iron age individuals from Swat valley in northern Pakistan (SPGT). At subclade level BR3 shared nodes only with SC_Kashmir (Scheduled Caste from Kashmir) and Brokpa tribe (Fig. 2b). In the Bayesian phylogenetic tree, Burzahom3 individual shared clade with Brokpa tribe (Fig. 2c). Their lineage diverged from each other at around 6000 years ago and with other Indo-Europeans and Dravidians from India much earlier (~ 13,000–15,000 years ago). In the Median-Joining network, Burzahom3 shared haplotype with Brokpa individual (Fig. 2d). The coalescent age estimate of Burzahom3 M65 haplotype is around 7041 ± 1190, which is in agreement with ML estimate from Bayesian analysis (~ 6000 YBP) (Supplementary Table S5).

Evolutionary phylogeny and divergence pattern of megalithic sample

For mtDNA haplotype U2b observed in megalithic sample Burzahom7, we included 38 modern and 7 ancient published reference samples to infer the phylogeny (Supplementary Table S2). In terms of its geospatial distribution mtDNA haplogroup U2b, this is highly prevalent in Central Asia (Tajikistan) (0.25), followed by Pakistan (0.17) and also among Tibetans (0.14). U2b has also noticeable presence in Thailand (0.07), Kashmir (0.07) and Uyghurs (0.07) population (Fig. 3a). In the Maximum-Likelihood tree with Mega, The Burzahom7 individual represents a deeply rooted separate clade, but shares the major clade with SPGT (Swat valley Iron age) samples, north Indian Brahmin and single Uyghur individual. Hence, in terms of sample representation, this major clade is heterogenous and majorly occupied by SPGT samples (Fig. 3b). The Bayesian phylogenetic tree was rooted to Tyumen_HG (Western Siberian Hunter-Gatherer) and Burzahom7 sample was present in separate deep-rooted branch and sharing major clade populated with SPGT individuals (major representation), north Indian Brahmin and Uyghur individuals (Fig. 3c). The ML estimates for the divergence of Burzahom7 haplotype goes back to ~ 9000 YBP (8976 YBP). In the Median-Joining network the megalithic Burzahom7 sample shares node with the SPGT sample and Brahmin (Indo-European) sample (Fig. 3d). Divergence time of the Burzahom7 sample from SPGT haplotype based on Rho (ρ) statistics was 8449.514 ± 1303.788 (Supplementary Table S5), which is similar to ML-based estimates from Bayesian analysis (8976 YBP).

Evolutionary phylogeny and divergence pattern of Medieval samples

For the megalithic samples Med1 and Med2 from Kashmir the mtDNA haplogroups obtained were M30 and W4 respectively. For the downstream analysis with Med1 sample, we included 88 modern published M30 reference samples (Supplementary Table S2) and 17 ancient samples ranging from Bronze age to medieval period from South Asia. The worldwide frequency distribution of M30 haplogroup indicates its highest prevalence in Pakistan (0.11) and Kashmir (0.09) (Fig. 4a). Its presence is also noticeable in the Maharashtra (0.07) and Karnataka (0.07) in India. In the ML phylogenetic tree inferred from Mega, Med1 sample shares a major clade with Roopkund_medieval samples, Kashmir, Central Asia, Middle Eastern and Iron age and historical samples from Swat valley in Pakistan (Fig. 4b). In the Bayesian phylogenetic tree, however, Med1 sample shares clade through common node with Roopkund_medieval individual (Fig. 4c). The coalescent age estimate for this node

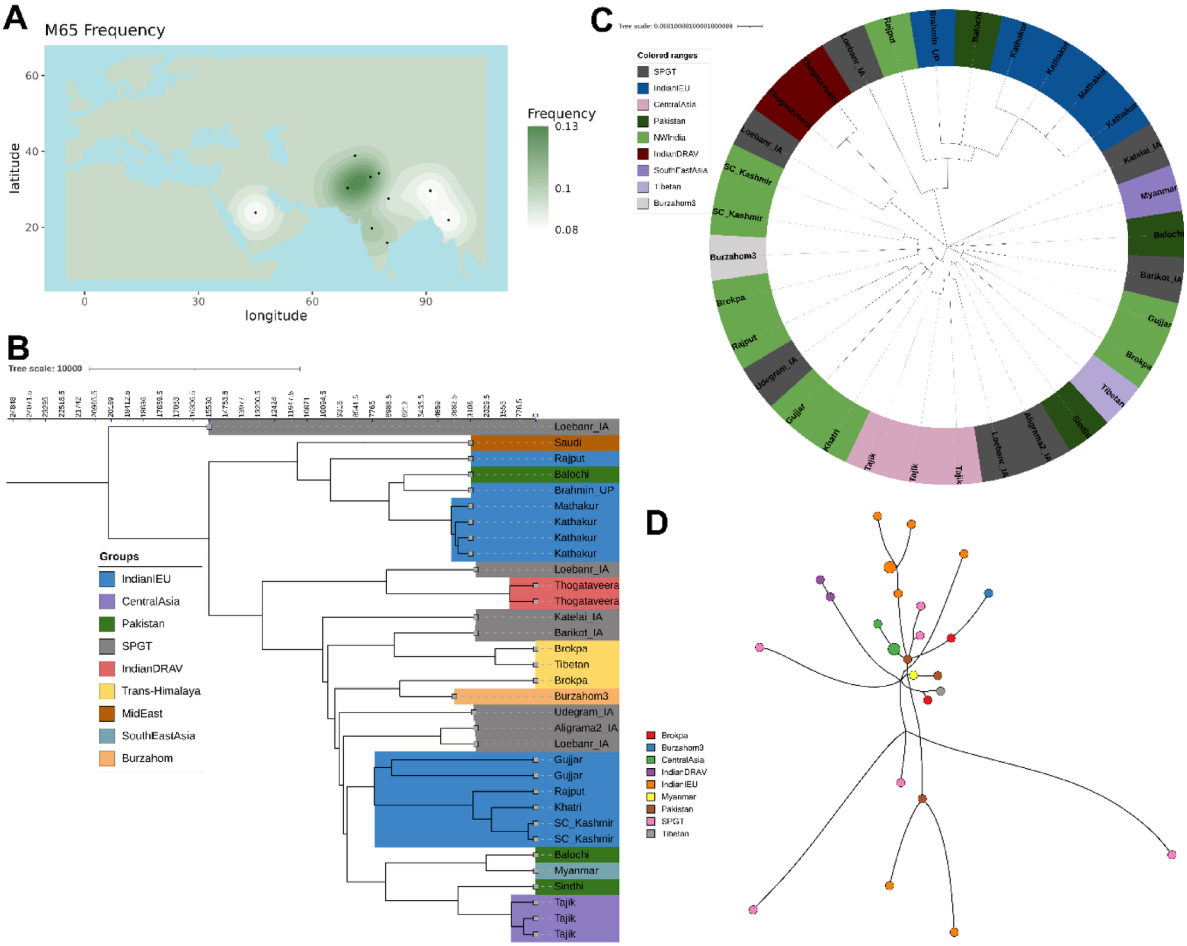

Fig. 2. Mitochondrial analysis results for Burzahom3 (Neolithic) sample; (**A**) Isofrequency map of mitochondrial haplogroup M65, (**B**) Bayesian evolutionary phylogeny for M65, (**C**) ML phylogeny reconstruction for mtDNA haplotype M65, (**D**) MJ evolutionary haplotype network of M65.

is approximately 3500 YBP (3493 YBP) (Supplementary Table S5). While their divergence from Swat Iron age samples was 6210 YBP based on ML estimates from Bayesian analysis. In the MJ haplotype network medieval sample Med1 shares node with sample from Central Asia and Roopkund medieval, SPGT and Swat_H sample haplotype extends from Med1 haplotype (Fig. 4d). The TMRCA of Med1 and Roopkund_med haplotypes from Central Asian haplotype was 2212.968 ± 667.235.

The worldwide distribution of mtDNA haplogroup W4 (found in Med2) suggests its major occurrence in European countries like Italy (0.28), Finland (0.17) and Poland (0.07). In South Asia W4 is present with highest frequency in Pakistan (0.14) and with minor presence in Kashmir in India (0.03) (Fig. 5a). In the evolutionary tree based on Maximum-likelihood approach in MEGA, medieval sample Med2 falls in a clade with Tepe_Hissar_C and BMAC (Bactria-Margiana-Archaeological-Complex), but not with any modern-day samples from South Asia (Fig. 5b). However, in the Bayesian evolutionary phylogeny, Kashmir Med2 sample forms a separate branch among the ancient samples (like Sappali_Tepe_BA, Gonur1_BA and Tepe_Hissar_C) (Fig. 5c). Divergence time based on likelihood estimates of Med2 was more than 5000 YBP from all modern Eurasians. Again, in Median-Joining haplotype network the Med2 haplotype (Yellow circle) shows clear evolutionary link with BMAC haplotype (Red circle) (Fig. 5d). The TMRCA of Med2 haplotype from BMAC (Gonur1_BA) haplotype is ~ 14,000 YBP (14,484.88 ± 1707.06 YBP) (Supplementary Table S5).

Discussion

The state of Jammu and Kashmir lies at the crossroads of Eurasia, with China and Tibet on one side and Afghanistan and Pakistan on the other. This region is considered a transit point for various population resettlements[45]. The current population structure of the region is influenced by various migrations and invasions from Central Asia and China[46]. The influence of China and Tibet in Kashmir has been undeniable since ancient times and is reflected in the Northern Neolithic culture seen in tools and burial practices in the region. The material culture of Burzahom site (Fig. 1a) bears strong similarities to that of the Swat Valley in terms of pottery, matt printing and perforations on pottery, double-notched rectangular and oval harvest vessels, and the presence of pits. These similarities with Swat and the six new sites of Baramulla indicate a connection between Kashmir,

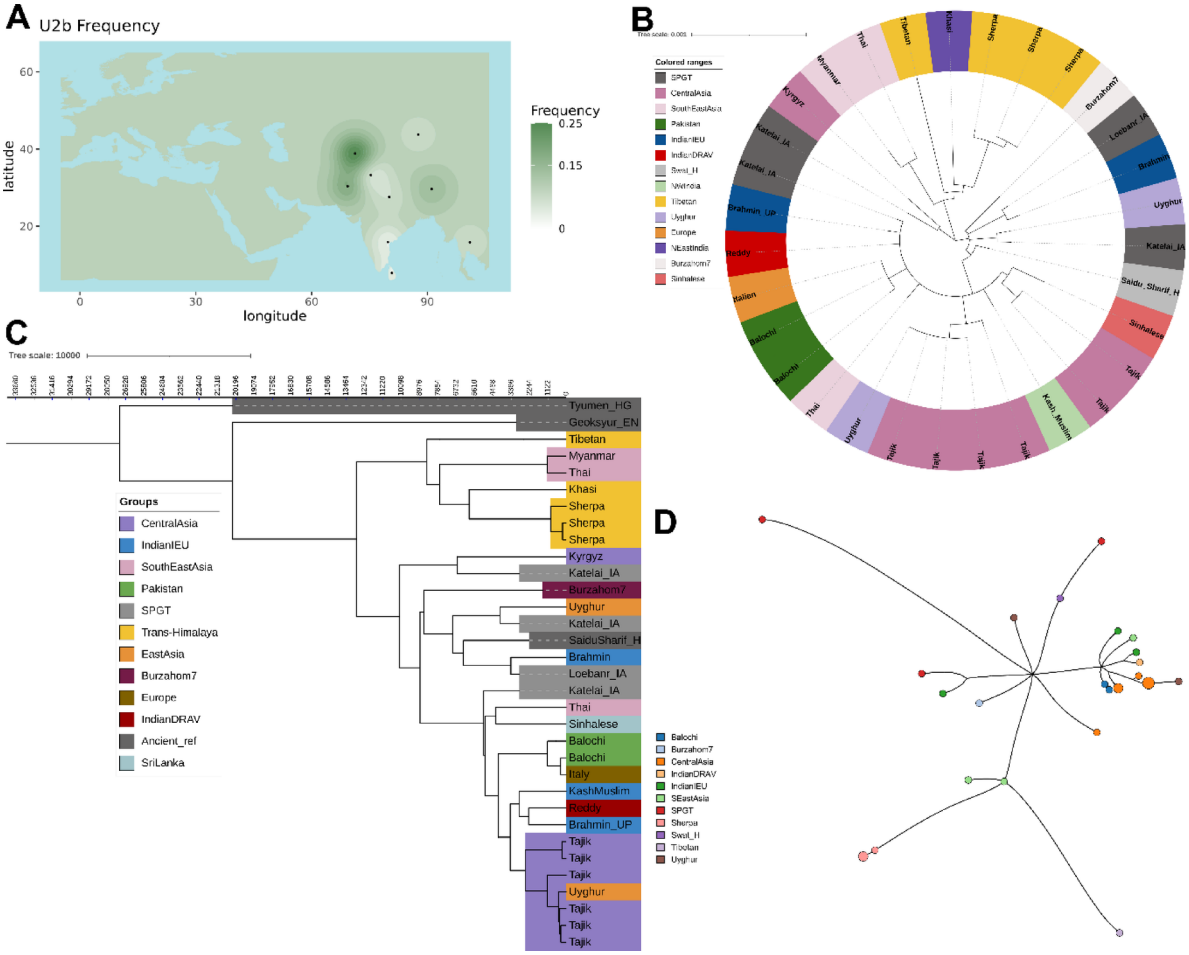

Fig. 3. Mitochondrial analysis results for Burzahom7 (Megalithic) sample; (**A**) Isofrequency map of mitochondrial haplogroup U2b, (**B**) Bayesian evolutionary phylogeny for U2b, (**C**) ML phylogeny reconstruction for mtDNA haplotype U2b, (**D**) MJ evolutionary haplotype network of U2b.

Swat and Baramulla[47]. Although cultural and archaeological aspects suggest a complex influence of migrations, assimilation and demographic changes in the region, detailed genetic aspects were missing. The maternal macro haplogroups found in this region are mostly deeply rooted in India viz. Macrohaplogroup M, suggesting very early settlement of India during the first waves of modern humans from Africa[48,49]. Genetic studies using contemporary Kashmiri mitogenomes suggested a prevalence of mtDNA haplogroups M65a and its subclades, as well as South Asian specific M30, U2a and U2b, as well as Tibetan A21. Haplogroup M65a, present in Tibetan, Himalayan, Indian, and Pakistani populations, is thought to have originated in the Upper Palaeolithic on the Indian subcontinent[50]. Most genetic studies covering the Central Asia (Tajik), Gilgit-Baltistan, Kashmir and Ladakh regions indicate genetic continuity in terms of mtDNA lineage heterogeneity. This is evidenced by the presence of South Asian (M30, U2a and U2b), West Eurasian (subgroups of H and W), East Asian and Tibetan (A and its subgroups, M9) in this vast region, once well connected by the silk route[51].

The present study further extends the quest to uncover the genetic continuity of this region by incorporating temporal genetic aspects with samples from the Neolithic to the Middle Ages. The Neolithic sample (AMS date 2009–2002 calBCE) (Supplementary Table S6) was recovered from the Burzahom site in Kashmir and genetically belongs to the mtDNA lineage M65 (observed haplogroup M65a + @16311), which is also present in the modern mitochondrial gene pool of the Kashmir region. This reflects maternal genetic continuity in this region over the millennia. Both the ML tree and the Bayesian phylogeny indicate its close relationship with local groups, particularly the Brokpa tribe. The coalescence time of the Burzahom3-M65 haplotype in both the ML-based and Rho statistics-based methods suggests an earlier divergence (~7000 YBP), clearly indicating a deep-rooted maternal ancestry of this region. Furthermore, the only megalithic (547–580 calCE) (Supplementary Table S6) sample found (Burzahom7) represents the U2b2 maternal lineage, which is widespread in Kashmir, Tajikistan and Pakistan and was also found in a Bronze Age sample from Rakhigarhi (I6113). The Bayesian phylogeny and ML tree indicate a relationship of Burzahom7 with SPGT (Iron Age Swat Valley) and North Indian Brahmins, but clearly at a much earlier coalescence period (~8000 YBP). This suggests its largely local origin and deep-rooted phylogeny in Kashmir region. The prevalence of both mtDNA haplotypes M65 and U2b among Iron age

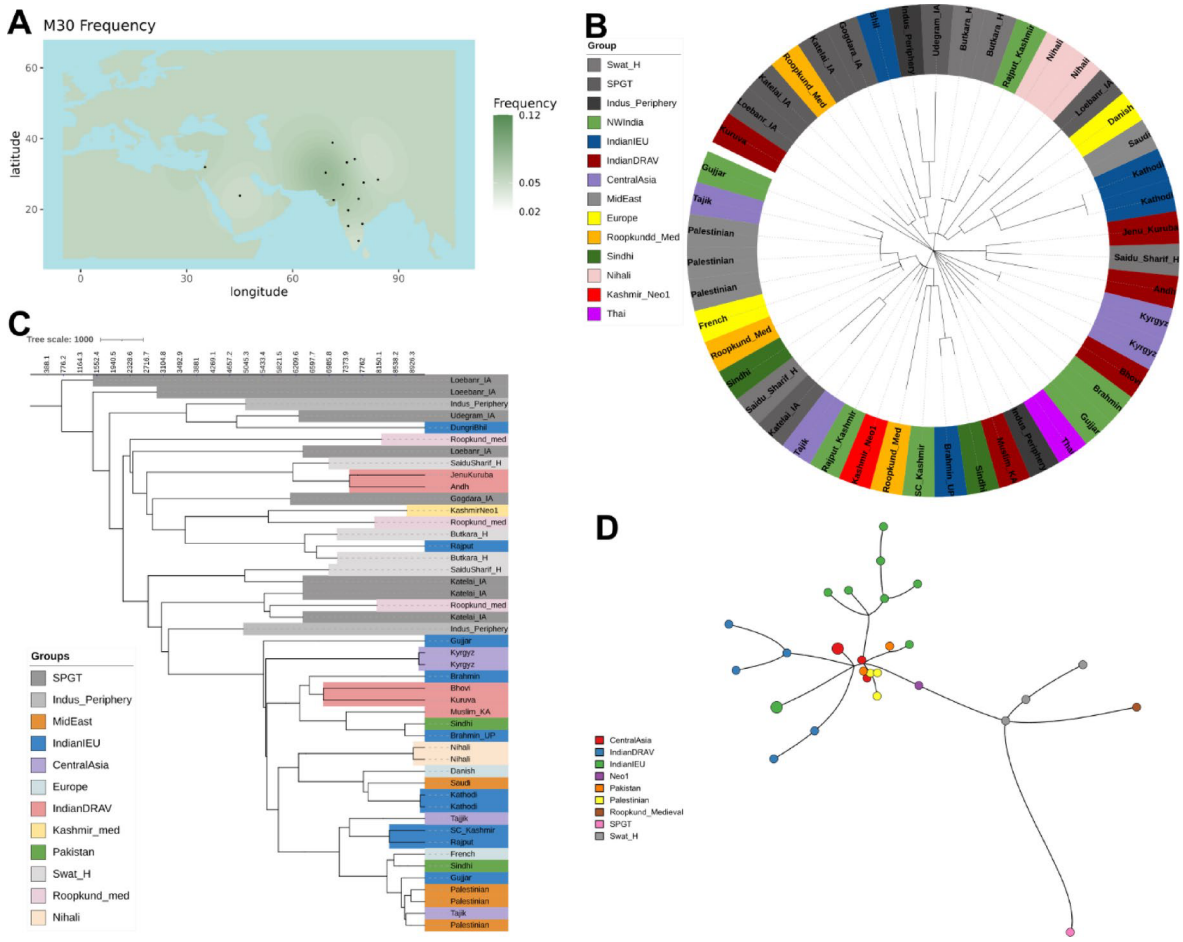

Fig. 4. Mitochondrial analysis results for Med1(Neo1) (Medieval) sample; (**A**) Isofrequency map of mitochondrial haplogroup M30, (**B**) Bayesian evolutionary phylogeny for M30, (**C**) ML phylogeny reconstruction for mtDNA haplotype M30, (**D**) MJ evolutionary haplotype network of M30.

samples from Swat valley and also evolutionary phylogenetic clustering with Neolithic and Megalithic Burzahom samples further validates the archaeological evidences of cultural similarity with SPGT (Swat valley Iron age).

On the contrary, in the medieval period samples (Med1 and Med2) from Kashmir, genetic evidences of links with Swat valley as well as Central Asia suggests large scale demographic shifts in the region. Higher prevalence of mtDNA haplogroup M30 in Northwest and Pakistan as well as its evolutionary phylogeny, in which Med1 clusters with SPGT and Swat_H samples, indicates again the close connection between Swat and Kashmir. This link with Swat valley, which has enough archaeological evidences, extends from Neolithic to medieval timeframe, at least in maternal genetic aspects. Also, in Bayesian phylogeny Med1 sample forms clade with Roopkund_Med (Roopkund medieval) sample, providing some evidence in support of link of Roopkund lake migrants with Kashmir. Although, number of samples is the limitation to this hypothesis and more medieval samples from Kashmir will be conclusive evidence. However, another medieval sample (Med2) belongs to more west Eurasian mtDNA lineage (W4) having origin outside South Asia. The W4 haplotype in Kashmir_Med2 clearly links with Bronze age Central Asia (BMAC culture). This haplogroup is highly prevalent in Europe, with minor presence in Northwest India and Pakistan. The ML phylogeny and Bayesian evolutionary tree clearly suggest the affinity of Med2 sample with copper age and Bronze age Central Asia. Haplotype network of W4 mtDNA haplogroup suggests the origin of Med2 haplotype from BMAC rather than Europe. The coalescent age of M30 haplotype is comparatively much recent (2212 YBP) compared to W4 haplotype (~ 14,000 YBP), which clearly indicates recent lineage of Med1 individual in Kashmir and much earlier lineage of W4 individual in the valley and probable migration from BMAC (Central Asia) site.

In conclusion, this study is the first ever attempt to cover maternal genetic continuity in South Asia using ancient DNA technique and ancient mitogenomes. It clearly indicates the genetic continuity in the Kashmir region along with Pakistan and Central Asia in terms of deeply rooted mitochondrial lineages such as M65, U2b and M30. Our genetic analysis establishes and further validates the archaeological evidences of close tie of Kashmir with Swat valley from Neolithic to medieval periods. The W4 lineage may represent human migration from Central Asia either through trade route links or part of frequent incursions (Mughal invasions). Although Bayesian evolutionary analysis indicates some link of Kashmir_Med1 (medieval) haplotype with Roopkund_

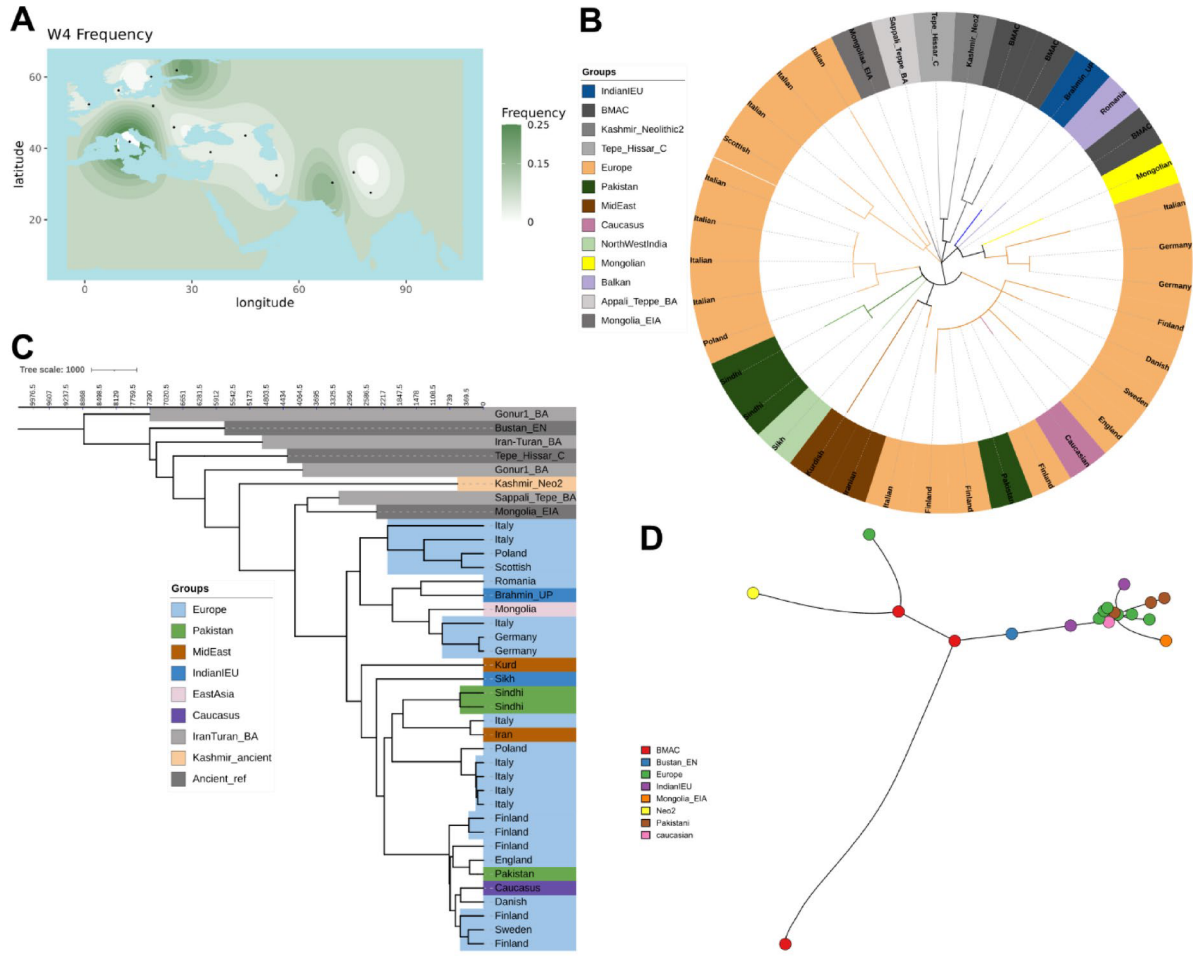

Fig. 5. Mitochondrial analysis results for Med2 (Neo2) (Medieval) sample; (**A**) Isofrequency map of mitochondrial haplogroup W4, (**B**) Bayesian evolutionary phylogeny for W4, (**C**) ML phylogeny reconstruction for mtDNA haplotype W4, (**D**) MJ evolutionary haplotype network of W4.

med migrants, the evidence is still inconclusive given the resolution of ancient data and further inclusion of more samples will be helpful.

Data availability

The datasets generated and/or analysed during the current study are available in the Zenodo repository, https://doi.org/10.5281/zenodo.14232374. The complete mitogenome fasta files are also available from Genbank access ion IDs PV391148, PV415134, PV424087 and PV424088.

Received: 10 November 2024; Accepted: 29 April 2025
Published online: 09 October 2025

References

1. Sharma, M. The region of Kashmir in ancient literature with special mention to tribes. *Pramana Res. J.* **9**, 703–707 (2019).
2. Stein, M. A. *Kalhana's Rajatarangini: A chronicle of the kings of Kashmir*. Vol. 11 (Motilal Banarsidass, 1989).
3. Dar, R. A., Romshoo, S. A., Chandra, R. & Ahmad, I. Tectono-geomorphic study of the Karewa Basin of Kashmir Valley. *J. Asian Earth Sci.* **92**, 143–156 (2014).
4. Agrawal, D. P. Man and environment in India through ages: An interdisciplinary study of the Indian quaternary with focus on north-west. (1992).
5. Kusumgar, S., Agrawal, D., Juyal, N. & Sharma, P. Palaeosols within loess: Dating palaeoclimatic events in Kashmir. *Radiocarbon* **28**, 561–565 (1986).
6. Betts, A. et al. The northern Neolithic of the western Himalayas: New research in the Kashmir Valley. *Archaeol. Res. Asia* **18**, 17–39 (2019).
7. Bhat, S. A. & Dubey, M. Burzahom (Burzohama) a Neolithic Site in Kashmir. *Education*, 111 (2008).
8. Thapar, B. Fresh light on the Neolithic cultures of India. *Archaeological Perspective of India since Independence*, 37–43 (1985).
9. Yatoo, M. A. & Bandey, A. A. Relations of neolithic Kashmir with south and central asia a comparative analysis of material culture from new sites in Kashmir. *J. Central Asian Stud.* **21**, 37 (2014).
10. Khazanchi, T. Our earliest ancestors. Kashmir and it's people: studies in the evolution of Kashmiri society **4**, 1 (2004).

11. Yatoo, M. A. Kashmir and Swat during neolithic times-a comparative analysis of material culture between the sites of two distinct regions. *Ancient Asia* **10** (2019).
12. Allchin, B. & Allchin, F. R. The birth of Indian civilization: India and Pakistan before 500 BC. (1968).
13. Ghosh, A. *An encyclopaedia of Indian archaeology*. (Brill, 1990).
14. Bandey, A. A. *Prehistoric Kashmir: Archaeological history of Palaeolithic & Neolithic cultures*. (Dilpreet Publishing House, 2009).
15. Fairservis, W. A. *The roots of ancient India*. (University of Chicago Press, 1975).
16. Spate, M., Zhang, G., Yatoo, M. & Betts, A. New evidence for early 4th millennium BP agriculture in the Western Himalayas: Qasim Bagh, Kashmir. *J. Archaeol. Sci.: Rep.* **11**, 568–577 (2017).
17. Yatoo, M. A. Iron age material culture in South Asia-analysis and context of recently discovered slag sites in northwest Kashmir (Baramulla District) in India. *Ancient Asia* **6**, 3–3 (2015).
18. Shinde, V. et al. An ancient Harappan genome lacks ancestry from steppe pastoralists or Iranian farmers. *Cell* **179**, 729–735 (2019).
19. Mani, B. Kashmir Neolithic and early Harappan: A linkage. *Pragdhara* **18**, 229–247 (2008).
20. Dabney, J. & Meyer, M. Extraction of highly degraded DNA from ancient bones and teeth. *Ancient DNA: Methods and Protocols*, 25–29 (2019).
21. Korlević, P. et al. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).
22. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010**, pdb. prot5448 (2010).
23. Peltzer, A. et al. EAGER: Efficient ancient genome reconstruction. *Genome Biol* **17**, 60. https://doi.org/10.1186/s13059-016-0918-z (2016).
24. Andrews, R. M. et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147. https://doi.org/10.1038/13779 (1999).
25. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes* **9**, 88. https://doi.org/10.1186/s13104-016-1900-2 (2016).
26. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint* arXiv:1303.3997 (2013).
27. Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684, https://doi.org/10.1093/bioinformatics/btt193 (2013).
28. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol* **16**, 224. https://doi.org/10.1186/s13059-015-0776-0 (2015).
29. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008, https://doi.org/10.1093/gigascience/giab008 (2021).
30. Weissensteiner, H. et al. HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res* **44**, W58-63. https://doi.org/10.1093/nar/gkw233 (2016).
31. Hiemstra, P. H., Pebesma, E. J., Twenhöfel, C. J. W. & Heuvelink, G. B. M. Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Comput. Geosci.* **35**, 1711–1721. https://doi.org/10.1016/j.cageo.2008.10.011 (2009).
32. Pebesma, E. & Bivand, R. *Spatial Data Science: With Applications in R*. (CRC Press, 2023).
33. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer International Publishing, 2016).
34. Team, R. C. in *R Foundation for Statistical Computing*, Vienna, Austria (2021).
35. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol Biol Evol* **38**, 3022–3027. https://doi.org/10.1093/molbev/msab120 (2021).
36. Leigh, J. W., Bryant, D. & Nakagawa, S. Popart: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116. https://doi.org/10.1111/2041-210x.12410 (2015).
37. Bouckaert, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **15**, e1006650, https://doi.org/10.1371/journal.pcbi.1006650 (2019).
38. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol* **34**, 772–773. https://doi.org/10.1093/molbev/msw260 (2017).
39. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat Methods* **9**, 772. https://doi.org/10.1038/nmeth.2109 (2012).
40. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* **67**, 901–904, https://doi.org/10.1093/sysbio/syy032 (2018).
41. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214. https://doi.org/10.1186/1471-2148-7-214 (2007).
42. Sealy, J., Johnson, M., Richards, M. & Nehlich, O. Comparison of two methods of extracting bone collagen for stable carbon and nitrogen isotope analysis: Comparing whole bone demineralization with gelatinization and ultrafiltration. *J. Archaeol. Sci.* **47**, 64–69 (2014).
43. Guiry, E. J. & Szpak, P. Improved quality control criteria for stable carbon and nitrogen isotope measurements of ancient bone collagen. *J. Archaeol. Sci.* **132**, 105416 (2021).
44. Agnihotri, R. et al. Radiocarbon measurements using new automated graphite preparation laboratory coupled with stable isotope mass-spectrometry at Birbal Sahni Institute of Palaeosciences, Lucknow (India). *J. Environ. Radioact.* **213**, 106156 (2020).
45. Sharma, I. et al. Ancient human migrations to and through Jammu Kashmir- India were not of males exclusively. *Sci Rep* **8**, 851. https://doi.org/10.1038/s41598-017-18893-8 (2018).
46. Witas, H. W., Tomczyk, J., Jędrychowska-Dańska, K., Chaubey, G. & Płoszaj, T. mtDNA from the early bronze age to the roman period suggests a genetic link between the indian subcontinent and mesopotamian cradle of civilization. *PLoS ONE* **8**, e73682. https://doi.org/10.1371/journal.pone.0073682 (2013).
47. Yatoo, M. A. J. A. A. Kashmir and Swat During Neolithic Times – A Comparative Analysis of Material Culture Between the Sites of Two Distinct Regions. (2019).
48. Behar, D. M. et al. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet* **90**, 675–684. https://doi.org/10.1016/j.ajhg.2012.03.002 (2012).
49. Sahakyan, H. et al. Origin and spread of human mitochondrial DNA haplogroup U7. *Sci Rep* **7**, 46044. https://doi.org/10.1038/srep46044 (2017).
50. Richards, M. et al. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67**, 1251–1276. https://doi.org/10.1016/S0002-9297(07)62954-1 (2000).
51. Kumar, L. et al. The maternal genetic origin and diversity of the extant populations of the Ladakh region in India. *Mitochondrion* **75**, 101828. https://doi.org/10.1016/j.mito.2023.101828 (2023).

Acknowledgements

NR thanks the Director, Birbal Sahni Institute of Palaeosciences for the financial and scientific support. We thank the Director, Anthropological Survey of India, Kolkata and Ministry of Culture, New Delhi for initiating this project and supervised the work. We thank all the study participants who volunteered in this study. The

research fellowships from University Grants Commission, New Delhi to AD is gratefully acknowledged.

Author contributions

Conceptualisation: N.R.; sample collection and genotyping: A.D., N.P., R.R., S.K., M.A.S., M.Y., S.K.P., N.S.S.; data analysis and visualisation: L.K., A.D.; L.K., A.D., and S.K.S.G. significantly contributed to results interpretation; L.K. wrote the original draft, with contributions from A.P., R.R., S.K., S.K.S.G. and Sn.K.; N.R., L.K., and A.D. revised the manuscript and contributed to the final version.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval

Sample collection and study was approved by Anthropological Survey of India, Kolkata, West Bengal, India.

Informed consent

Not required for ancient DNA study.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-00568-4.

Correspondence and requests for materials should be addressed to N.R.