Recovery of Chloroplast Genomes from Medieval Millet Grains Excavated from the Areni-1 Cave in Southern Armenia

Stephen M. Richards (smr3434@gmail.com)
School of Biological Science, The University of Adelaide

Leiting Li
State Key Laboratory of Plant Molecular Genetics, Shanghai Center for Plant Stress Biology

James Breen
School of Biological Science, The University of Adelaide

Nelli Hovhannisyan
Yerevan State University

Oscar Estrada
School of Biological Science, The University of Adelaide

Boris Gasparyan
National Academy of Sciences of the Republic of Armenia

Matthew Gilliham
Waite Research Institute & School of Agriculture, Food, and Wine

Alexia Smith
University of Connecticut, Department of Anthropology

Alan Cooper
BlueSky Genetics

Heng Zhang
State Key Laboratory of Plant Molecular Genetics, Shanghai Center for Plant Stress Biology

Research Article

Keywords: chloroplast genomes, medieval millet grains, Areni-1 cave, Panicum miliaceum L.

Posted Date: October 27th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-959635/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Broomcorn millet (*Panicum miliaceum* L.) was domesticated in northern China at least 7,000 years ago and was subsequentially adopted as a cereal in many areas throughout Eurasia. One such locale is Areni-1 an archaeological cave site in Southern Armenia, a region that has an important history in crop domestication. The rich botanical material found at Areni-1 includes grains identified by morphology as broomcorn millet that were $^{14}$C dated to the medieval era (873 ± 36 CE and 1118 ± 35 CE). To retrace the demographic history of these broomcorn millet samples, we used ancient DNA extraction and hybridization capture enrichment to sequence and assemble three chloroplast genomes from the Areni-1 grains and then compared these sequences to 50 modern chloroplast genomes. Overall, the chloroplast genomes contained a low amount of diversity and little inference on broomcorn demography could be made. However, in a phylogeny the chloroplast genomes separated into two clades with strong bootstrap support, similar to what has been reported for nuclear DNA from broomcorn millet. In a haplotype network, the chloroplast genomes of two accessions of wild (undomesticated) broomcorn millet contained a relatively large number of variants, 11 SNPs. These SNPs were not present in the domesticated varieties, suggesting these wild accessions may not be directly related to the lineages that underwent domestication or that broomcorn millet may have undergone a domestication bottleneck resulting in lost diversity in the chloroplast genome. These results demonstrate that broomcorn millet from archaeological sites can preserve DNA for at least 1000 years and serve as a genetic resource to study the domestication of this cereal crop.

Introduction

Millet is a generic term that refers to a group of small seeded grasses that were important crops in the past and remain a significant source of food and fodder in many areas of the world today$^{1,2}$. The majority of cultivated millet belong to the Paniceae tribe of the Poaceae family and include broomcorn millet (*Panicum miliaceum* L.), pearl millet (*Pennisetum glaucum* (L.) R.Br.), and foxtail millet (*Setaria italica* (L.) P. Beuvois), while finger millet (*Eleusine coracana* Gaertn.) belongs to the Chlorideae clade of Poaceae$^{3}$. Millet, along with other C4 pathway species such as corn (*Zea mays* L.), sugar cane (*Saccharum officinarum* L.) and sorghum (*Sorghum bicolor* (L.) Moench), are characterised by their ability to be cultivated in a wide range of environments, requiring modest amounts of water and having short growing seasons (60–90 days).

Across southwest Asia, numerous early archaeobotanical finds of millet have been recorded, however the domestication history of millet, and particularly of broomcorn millet remains unclear. Recent research observing charred grain remains in east Asia suggests that broomcorn millet was being farmed in Northern China during the Neolithic as early as 5880 BCE$^{4–8}$. In central Asia, the earliest broomcorn millet grains date to the mid/late 3rd millennium cal BCE$^{9,10}$. Whilst further west in the Caucasus, the earliest broomcorn millet dates to the mid 2nd millennium cal BCE$^{11}$. For broomcorn millet specifically, there has been a debate as to whether a single domestication event took place in northern China or Eastern Europe.
followed by a rapid east-west dispersal or whether multiple independent domestication events took place across space and/or through time. This uncertainty was partially compounded by the difficulty in distinguishing between archaeobotanical remains of *Panicum* and *Setaria* based on grain morphology alone, particularly in the case of charred remains, which can be subject to morphological changes and degradation of the outer seed coat. Recent archaeological and genetic studies support domestication of broomcorn millet solely in China. Archaeological evidence indicates that broomcorn millet was utilized in different ways including being ground to flour/meal and also being used to brew beer.

Archaeobotanical data from campsites suggest that the initial dispersal of domesticated broomcorn millet appears to have been driven (east to west) by mobile pastoralists who migrated along the mountain foothill ecotone of Central Asia and adopted broomcorn millet as a low-investment crop that was grown during summer encampments. The migration route taken by these pastoralists would eventually become one of the Silk Road east-west exchange pathways connecting China and Europe.

Within the Southern Caucasus (which includes Armenia, Georgia, and Azerbaijan), the earliest finds of broomcorn millet come from the Bronze Age, dated to the 2nd millennium BCE. Later examples of broomcorn millet have been found at various sites including Areni-1 (also known as Birds’ Cave: 39° 43’ 53” N, 45° 12’ 13” E) a three-chambered karstic cave site located on the left bank of the Arpa River basin, a tributary of the River Araxes, within the eastern portion of the modern village of Areni in southern Armenia. In the context of crop domestication, Areni-1 and the Caucasus in general are interesting locations where wild relatives of many cereals are widely represented. The proximity to both the Fertile Crescent a region important for the development of agriculture and to Silk Roads exchange routes underlies the importance of these locations to the understanding of the domestication, adaptation, and distribution of many cereal crops.

Excavations at Areni-1 began in 2007, exposing well preserved occupations dating through the Early Chalcolithic (5,260 – 4,960 Cal BCE) to the medieval (4th –18th centuries CE) eras. Brief Bronze and Iron Age occupations were also found, but settlement at the site was largely restricted to the Late Chalcolithic and medieval eras. Chalcolithic finds within the first gallery of the cave include numerous large ceramic storage vessels, some of which contained human remains such as skulls. Grape remains, archeochemical analyses, and the presence of clay wine presses indicate that wine was produced at the cave during the Chalcolithic. Artefacts from Areni-1 suggest that from the end of 5th millennium BCE onwards, local people were using the cave for a variety of additional purposes such as habitation, sheltering livestock, storing crops, and as a base for collecting fruits.

Plant remains recovered from Areni-1 include vast quantities of well-preserved desiccated and charred seeds, fruits, stones (endocarps), and stems of both wild and cultivated plants. Desiccated flowers, leaves, wood, and vine branches were also found in the cave, which is rare as generally, archaeobotanical remains across Southwest Asia are preserved predominantly through charring, which results in the attrition of many plant materials. A diverse range of taxa (over 40 genera) has been identified at Areni-1,
documenting the richness of plant use in antiquity at the site\textsuperscript{23}. While distinct differences are evident between the Late Chalcolithic and the medieval strata, continuity in some aspects of plant use is also apparent; in fact, many of the domesticated species observed in the cave continue to be used in the village of Areni today. Cereals recovered from both the Late Chalcolithic and medieval era occupations include bread wheat (\textit{Triticum aestivum} L.) and barley (\textit{Hordeum vulgare}), which were first domesticated in the Fertile Crescent during the Neolithic Revolution. By the medieval era, sorghum (\textit{Sorghum} spp.), rice (\textit{Oryza} spp.), and millet were a part of the botanical assemblage of the cave\textsuperscript{23,25}. Currently in Armenia, millet is mainly cultivated in the Gegharkunik Province and Ararat valley northwest and north of Areni-1 respectively, but not in the immediate vicinity of the cave.

The excellent preservation of the Areni-1 millet allowed the identification of the grains as broomcorn millet based on morphology alone. Further, this preservation suggested that endogenous DNA could be recovered at levels that would allow us to investigate the demographic history of broomcorn millet by performing genetic comparisons of the Areni-1 millet with modern varieties. For these comparisons we chose the chloroplast genome which is commonly used to study the evolutionary relationship between plant species because this genetic marker is haploid, inherited through a single parent, and has a highly conserved gene content\textsuperscript{26}.

Using ancient DNA (aDNA) extraction techniques, high-throughput sequencing, and hybridization capture enrichment\textsuperscript{27}, we generated three chloroplast genomes from five Areni-1 millet grains assayed. The Areni-1 chloroplast genomes were compared to organelle references from 11 members of the Paniceae tribe of grasses to confirm the morphometric identification and taxonomic placement of the grains. Additionally, to investigate the demographic history of broomcorn millet, the Areni-1 chloroplast genomes were compared to similar organelle sequences from 50 modern varieties of broomcorn millet using both phylogenetic and haplotype network analysis.

**Results**

**Areni-1 millet morphology and identification**

Millet grains from both Trench 1 and 3 of the Areni-1 site (Fig. 1a and Table 1) were well preserved via desiccation, maintaining both the original colour and the husk (palea and lemma) surrounding the grain (Fig. 1b). When morphological features of the grains were visually compared to modern caryopses of a variety of species belonging to the \textit{Panicum} and \textit{Setaria} genera, they overlapped best with broomcorn millet. The morphological similarities between the archaeological specimens and the modern comparative material are shown in Fig. 1b. Both demonstrate similar sizes, the presence of a highly smooth and glossy husk, and a similar outline with a pointed apical end and a relatively blunt proximal/base end. Morphological differences from reference examples of \textit{S. italica}, which are included in Fig. 1b, clearly indicate that the Areni-1 grains are not foxtail millet.
Occupations at Areni-1 have been dated from the Early Chalcolithic (5,260 – 4,960 Cal BCE) to the medieval (400–1,800 Cal CE) eras. To confirm the contextual dating of the Areni-1 grains to the medieval era, $^{14}$C dating was performed on botanical material associated with the grains. The size of the Areni-1 millet did not provide enough mass to perform carbon dating and extraction of aDNA from a single seed, so $^{14}$C dating was done indirectly using associated botanical material. The $^{14}$C dates placed the Areni-1 grains in the Early and High medieval era (Table 1).

### Table 1
Archaeological context and associated radiocarbon dates of millet seeds

| Millet Sample | Location in Areni-1 Cave | Dating Laboratory Number | Material Dated | Date | Calibrated Date 1 sigma (68.2%) | Calibrated Date 2 sigma (95.4%) |
|---------------|-------------------------|--------------------------|----------------|------|-------------------------------|-------------------------------|
| 11294a | Trench 1, Unit 3, Locus 50 (10.01.2009) | AA-107539 | Vitis sp. pedicel | 873 ± 36 BP | 1054 – 1218 CE | 1042 – 1248 CE |
| 11294b | | X30053 | | | | |
| 11292a | Trench 3, Sq O29/O30 | AA-104139 | Panicum miliaceum (pool of 3 grains) | 1118 ± 35 BP | 890 – 980 CE | 780 – 1020 CE |
| 11293a | | X27645 | | | | |
| 11295a | Locus 2, Spit 5 | | | | | |

### Ancient DNA extraction and library construction

After the species of the Areni-1 millet were established via grain morphometric analysis, we then aimed to recover ancient DNA (aDNA) from the grains in order to establish a molecular relationship to other broomcorn millet varieties via the chloroplast genome. Five grains were chosen randomly for extraction: two of the grains (11294a and 11294b) were from soil in Trench 1 located in the first gallery of the cave, whilst the remaining three grains (11292a, 11293a, and 11295a) were excavated near a round structure located under the overhang of the cave in Trench 3 (Fig. 1a and Fig S1). Extraction of aDNA from the Areni-1 grains and negative blank controls was performed in a designated low DNA laboratory using a previously published two-step protocol. Recovered aDNA was initially converted into truncated Illumina sequencing libraries and then processed into full length shotgun and chloroplast DNA (cpDNA) enriched libraries.

### Mapping Areni-1 shotgun data to chloroplast reference genome

In the shotgun libraries, 11292a and 11295a were the only two grains that yielded a significant level of endogenous DNA and produced a large number of unique reads mapping to the chloroplast reference (> 90,000). The remaining shotgun libraries produced more than a thousand-fold fewer unique chloroplast mapped reads (Table 2). The extraction blanks and certain millet samples (11293a and 11294b)
contained a comparatively small number of unique mapped reads in both the shotgun and cpDNA enriched libraries, suggesting that contamination from laboratory sources or between libraries during the processing of the Areni-1 samples was not a substantial issue. In the shotgun libraries, grains 11293a and 11294b both produced fewer mapped reads than the extraction blanks, but these results are likely very stochastic as these libraries contained a very small fraction of mappable reads.

**Mapping of Areni-1 enriched cpDNA data to chloroplast reference genome**

The cpDNA enriched libraries were produced using a hybridization capture protocol with RNA probes designed against barley and wheat organelle reference genomes. All ancient shotgun libraries were taken through two rounds of cpDNA hybridization capture regardless of endogenous DNA content in order to treat all samples in a similar manner and maximize the recovery of target sequences. Of the Areni-1 millet enriched for cpDNA, only three grains (11292a, 11294a, and 11295a) produced libraries that contained a sufficient number of reads that mapped to the chloroplast reference genome (> 19,000) that allowed reconstruction of the Areni-1 chloroplast genomes (Table 2). No reads from the ancient libraries could be mapped to the inverted repeat regions (IRs) of the chloroplast genome, because the short-fragmented sequences in this data could not be accurately placed in these nearly identical loci (Fig. 2). As the IRs do not contain any mapped reads, these loci were removed from all further analysis. Libraries 11292a and 11295a covered > 99.7 % of the reference (sans IRs) with at least one read, whilst the fraction of the reference covered by at least one read in library 11294a was approximately 8% less than the other two libraries.

**Read length distribution and mapDamage analysis of enriched cpDNA libraries**

To support the indirect $^{14}$C dating of the millet grains to the medieval era, sequences from the 11292a, 11294a, and 11295a cpDNA enriched libraries were examined for characteristics of authentic aDNA: short read lengths and the presence of deamination miscoding lesions. All libraries from the three grains contained reads that were fragmented to a short length, generally < 150 bp (Fig. 3a). Grain 11294a produced the shortest reads, suggesting that this sample was the most degraded of the three. In the cpDNA enriched libraries, the read length of grains 11292a and 11295a was shifted towards longer sequences, which is expected as larger DNA fragments form more stable complexes with enrichment probes and are thus recovered more efficiently in hybridization capture procedures. In contrast, the enriched library for 11294a experienced a downward shift in read length, which likely stems from the stochastic effect of having only 72 unique mapped reads in the shotgun library for this sample. In mapDamage analysis of the cpDNA enriched libraries, the ‘5 ends of reads exhibited the increased levels of C $\to$ T misincorporations caused by deamination of cytosine typical of aDNA (Fig. 3b and Fig S2). MapDamage results are not presented for shotgun data because the library from grain 11294a contained an insufficient number of reads for this analysis. Since direct $^{14}$C dating of the Areni-1 millet used for
sequencing was not possible due to the small mass of the grains, the fragmentation patterns and miscoding profiles are important data that support the medieval dating of the samples.
### Table 2
Mapping statistics of shotgun and cpDNA enriched libraries

| Sample     | # Collapsed Reads | # Mapped Reads | # Unique Mapped Reads | % Reference covered by least 1 read (sans IRs) |
|------------|------------------|----------------|-----------------------|---------------------------------------------|
| **SHOTGUN LIBRARIES** |                  |                |                       |                                             |
| 19885      | 6702745          | 22             | 14                    |                                             |
| (extraction blank) |                  |                |                       |                                             |
| 11292a     | 31367680         | 155496         | 92560                 |                                             |
| 11293a     | 150049           | 2              | 2                     |                                             |
| 11294b     | 6404935          | 0              | 0                     |                                             |
| 18335      | 6493567          | 15             | 9                     |                                             |
| (extraction blank) |                  |                |                       |                                             |
| 11294a     | 4635555          | 72             | 72                    |                                             |
| 18324      | 6771883          | 7              | 7                     |                                             |
| (extraction blank) |                  |                |                       |                                             |
| 11295a     | 42796618         | 444954         | 137838                |                                             |
| **cpDNA ENRICHED LIBRARIES** |                  |                |                       |                                             |
| 19885      | 9896218          | 1451964        | 67                    |                                             |
| (extraction blank) |                  |                |                       |                                             |
| 11292a     | 36092254         | 21953919       | 163494                | 99.77                                       |
| 11293a     | 35345608         | 88165          | 232                   |                                             |
| 11294b     | 32245568         | 269895         | 1532                  |                                             |
| 18335      | 77604223         | 270461         | 36                    |                                             |
| (extraction blank) |                  |                |                       |                                             |
| 11294a     | 26076965         | 7472857        | 19420                 | 91.28                                       |
### Mapping statistics for Areni-1 enriched cpDNA data

In the mapping of cpDNA enriched libraries, grains 11292a and 11295a produced similar numbers of unique reads (≈ 170,000) that mapped to the chloroplast genome reference, while sample 11294a generated approximately a tenth of the mapped reads observed in the other two grains. The maximum read depth was greater with grains 11292a and 11295a at 229 and 331 reads respectively, compared to 52 reads for 11294a (Fig. 3). Despite the differences in coverage of the chloroplast reference genome and read depth, all three samples produced chloroplast data that allowed for phylogenetic and haplotype network analysis.

### Generation of chloroplast genomes from 50 modern accessions of broomcorn millet

For comparative purposes chloroplast genomes were generated from 50 accessions of modern broomcorn millet (Table S2). These accessions had a mainly Eurasian distribution and were weighted towards Chinese accessions with 29 of the samples originating from various locations within China. Included in these accessions were two wild (undomesticated) broomcorn millet (SampleID 31 and 167) and two accessions that have previously been used to produce nuclear reference genomes (SampleID 69 and T296) \(^33,34\). Paired-end Illumina sequencing data from these modern accessions generated average read depths of ≥ 1356 reads across the reference chloroplast genome.

### Paniceae phylogenetic analysis

To confirm the visual identification and the taxonomic placement of the Areni-1 millet, the chloroplast genomes from these medieval grains were compared to similar sequences from 11 members of the Paniceae tribe, including broomcorn millet (Supplementary Table S1). A phylogeny was generated of the Paniceae and Areni-1 chloroplast genomes using RAxML (1,000 bootstraps and the GTRGAMMA model) and visualized using the Figtree software with rooting of the tree on maize (Fig. 4). In the tree, the Areni-1 millet and broomcorn millet were placed together on a branch with a bootstrap support of 100, which indicated that the medieval grains were closely related to broomcorn millet and confirmed the taxonomic placement and visual identification of the medieval grains. The placement of the *Panicum* and *Setaria* species in the tree suggested the topography produced in this analysis was correct. All the *Panicum* spp
were placed on the same branch within the tree and domesticated foxtail millet (S. *italica*) formed a clade with the wild antecedent of this cereal, *Setaria viridis*.

**Broomcorn millet phylogenetic and haplotype network analysis**

At the time the current study was started little was known about broomcorn millet chloroplast diversity or what information the chloroplast genome could provide on the demographic history of broomcorn millet. To address these deficiencies, we compared the Areni-1 chloroplast genomes to similar organelle sequences from 50 modern varieties of broomcorn millet including wild accessions. Between the modern accessions and the Areni-1 millet a total of 63 SNPs were identified in the chloroplast genomes and these genotypes were used to generate a phylogeny using the RAxML and FigTree methods described above (Fig. 5a). The tree produced in this analysis had no geographic structure and very low bootstrap support values at many of the nodes, so no inference on broomcorn millet demographics could be made. However, in the tree, the broomcorn accessions were split into two clades with a bootstrap value of 100 with all the Areni-1 millet residing in the same clade.

The relationship between the modern and Areni-1 chloroplast genomes was further investigated in a haplotype network using the PopArt software with the median spanning network algorithm (Fig. 5b). In this network, all the domesticated accessions, including the Areni-1 samples, were separated by only 1 to 5 SNPs. A core haplogroup, represented by the large node at the centre of the network, contained 19 accessions with a wide geographic range across Eurasia. The Areni-1 millet were separated from the core haplogroup by two mutations, one of which was common to all three medieval genomes. The largest number of variations was observed in the wild millet, which were separated from the domesticated varieties by 11 to 14 SNPs.

**Discussion**

In this study, we used DNA extraction and hybridization capture enrichment techniques to recover aDNA from medieval millet grains found in Areni-1, a cave site located in the Southern Caucasus a region that has an important history in the domestication and dispersal of many crops. Despite variable quality of aDNA, we were able to recover three chloroplast genomes from five Areni-1 grains assayed.

To date, excavations at Areni-1 have only produced millet grains from the medieval occupation. The Chalcolithic occupation of the cave occurred prior to the introduction of millet into the Southern Caucasus in the 2nd millennium BCE, so grains would not be expected to be present in excavated material from this period. As for the Bronze and Iron Ages, the occupation of Areni-1 during these times was brief and did not yield substantial plant deposits. The medieval occupants of Areni-1 utilized a wide range of plants as food including various fruits (apricots, plums, and grapes) and cereals (wheat, barley, and sorghum). Inclusion of millet in the suite of plants grown may have been a crop rotation strategy that the medieval occupants of Areni-1 adopted to maximize the productivity of agricultural land. Drought
resistant millet may have been grown in the arid summers experienced in the Southern Caucasus, while wheat and barley could have been grown in the cooler/wetter seasons of the year. Millet may have been grown as food and fodder, but also used to produce alcoholic beverages. In his travel log of the 4th century BCE, the Greek mercenary and historian, Xenophon of Athens, described a wine made by the local inhabitants of the Armenian highlands which was strengthened with barley and may have also contained millet.

Despite a similar well-preserved physical appearance, the endogenous DNA content varied among the Areni-1 grains used in this study as did the quality of the aDNA. For example, in the millet enriched for cpDNA, grain 11295a contained the highest level of endogenous DNA and the longest read length, while grain 11294a contained the least amount of endogenous DNA with the shortest read length (Table 2, Fig. 3a). The environment of the cave appears to be suited to prevent decomposition of biological material and macromolecules as excavations at Areni-1 have unearthed vertebrate soft tissues including a mummified goat and brain tissue in a human skull, demonstrating that the conditions in the cave can preserve ephemeral biological material. It is not clear why the Areni-1 millet exhibited different levels of DNA preservation. Grains with poor preservation may have undergone different harvesting, threshing, or other anthropogenic treatments that influenced DNA stability. Post excavation factors, such as storage or handling conditions, may have also influenced the different levels of DNA preservation. Further, the micro-environment has been shown to impact the survival of DNA in mammalian bones and similar processes may have influenced the decay of aDNA in the Areni-1 millet grains. Local variations in pH, ion concentration, moisture, or other abiotic factors could have affected DNA stability in the Areni-1 millet. Variable preservation of aDNA from grains is not unique to the Areni-1 millet as a study of 6,000 year old barley found large differences in recoverable endogenous DNA in grains excavated from the same occupation layer of a cave. Regardless of the varying integrity of the aDNA, the recovery of chloroplast genomes from the Areni-1 samples demonstrates that broomcorn millet grains can, under the right conditions, preserve DNA for at least a 1,000 years. Even with a poorly preserved sample (11294a), hybridization capture enrichment was able to recover sufficient cpDNA to perform a chloroplast phylogenetic and haplotype analysis.

In the Paniceae grass chloroplast phylogeny, the Areni-1 and broomcorn millet reference clustered with several species native to North America (Panicum virgatum, switchgrass; Panicum capillare, witchgrass) and Whiteochloa capillipes, which is from Australia, New Guinea, and Indonesia. Panicum sumatrense (little millet), a cereal crop originally from India also clustered with the Areni-1/broomcorn millet. Panicum is the largest genera of the grass family Paniceae and has a cosmopolitan distribution so the clustering of the Areni-1/broomcorn millet with widely distributed species is not unusual. While it is unlikely that the specimens from Areni-1 are directly related to these species, it is possible that some of them share a common ancestor from Southeast Asia, further supporting the east to west introduction of P. miliaceum in Eurasia. It should be noted that the chloroplast phylogeny indicates a close relationship between broomcorn millet and P. capillare (bootstrap support = 100) an observation that supports a previous study.
of *Panicum* nuclear DNA, which found that diploid *P. capillare* or a close relative is the likely maternal genome donor to tetraploid broomcorn millet.

In an attempt to investigate the demographic history of broomcorn millet, chloroplast phylogenetic and haplotype analyses were performed using the Areni-1 samples with modern accessions of broomcorn millet. In these analyses only 63 SNPs were identified in a group that included 48 modern domesticated accessions, two wild accessions, and the three medieval millet samples from Areni-1 cave. Accordingly, there was little differences between these millet genomes which was reflected in the phylogenetic and haplotype analyses. In the phylogeny there were low bootstrap support values through much of the tree. In the haplotype network, the vast majority of the haplotypes were separated by 1 to 5 SNPs, which did not allow for any conclusions to be drawn on the demographics of broomcorn millet (Fig. 5a and 5b). The inability of chloroplast genomes to resolve intraspecies relationships is not unique to the current study. Research on the genera *Amaranthus* and *Euphrasia* found that whole chloroplast genomes were highly conserved and lacked the power to discriminate between some intraspecies accessions.

These chloroplast analyses did allow other observations to be made about the phylogenetics of broomcorn millet. In the chloroplast phylogeny, the domesticated accessions were split into two clades with no geographical structure and each clade containing accessions from locations throughout Eurasia. These results are similar to a study by Hunt et. al. 2011 which used 16 microsatellite loci across 98 broomcorn landraces and divided modern broomcorn millet into two clades, one composed solely of Chinese accessions and the other clade contained accessions from throughout Eurasia. It is not clear why the chloroplast phylogeny in the current study lacked a clade of solely Chinese accessions as observed by Hunt et. al. 2011. Previous studies have reported incongruent chloroplast and nuclear phylogenies and have attributed these differences to incomplete lineage sorting and introgression/hybridization. Similar processes may be involved with the discrepancies between the broomcorn nuclear and chloroplast genomes, however the low levels of diversity in the chloroplast genomes are likely the main contributing factor for these differences. The broomcorn chloroplast genomes contained sufficient information for gross scale separation of the different accessions but was insufficient to accurately resolve fine scale relationships.

In the haplotype analysis, the wild broomcorn millet contained a relatively high number of variants, 11 SNPs, that were not present in the domesticated accessions. Several scenarios that are not mutually exclusive could explain this observation. One possibility is that the wild accessions sampled may not be directly related to the broomcorn lineages that were domesticated. Another possible scenario is the SNPs observed in the wild accessions may represent diversity lost in the chloroplast genome during a domestication bottle neck. Such a loss of diversity has been reported in a study by Leigh et. al. 2013, which described a reduction in both alleles and haplotypes in the chloroplast genomes of different wheat species after domestication. The current study would not be able to capture pre-domestication diversity because the domestication process can happen relatively quickly as exemplified by barley. Although barley was domesticated approximately 10,000 years ago, the nuclear genomes from 6,000 year old
barley grains from Israel show striking similarity to the genomes of modern local landraces, indicating the domestication process for barley was largely finished after 4,000 years. If following a similar timeline, the domestication of broomcorn millet would have been completed well before the medieval era and it would not be expected for the Areni-1 millet to contain significant levels of pre-domestication diversity. Resolving the reasons for the higher diversity in the wild accessions of the current study will require additional sampling of undomesticated and ancient chloroplast genomes.

It is possible that mapping of promiscuous DNA that has broken off the chloroplast genome and become incorporated in either the nuclear or mitochondrial genome has influenced the Areni-1 chloroplast results. Effort was made to minimize the impact of promiscuous cpDNA on the Areni-1 results by requiring a relatively high read depth (≥ 10) and at least 95% of the reads to call a mutation as an alternative allele. However, as it will be difficult to distinguish between promiscuous cpDNA and DNA from the chloroplast, proving that promiscuous cpDNA has not influenced the current results is extremely challenging. Regardless, the agreement of the visual identification of the Areni-1 millet as broomcorn millet and the placement of the Areni-1 millet with broomcorn millet in the Paniceae chloroplast phylogeny (Fig. 4) indicate that promiscuous cpDNA did not have a significant impact on the interspecies analysis.

In conclusion, here we generated 53 broomcorn millet chloroplast genomes, which includes three dating to the medieval era that can be used as resources to study this cereal crop and other C4 plants. A phylogeny generated from these chloroplast genomes divided the domesticated forms of broomcorn millet into two clades, an observation that parallels reports for using nuclear DNA. The wild accessions sampled in this study contained SNPs that were outside the diversity of the domesticated forms of broomcorn millet and the significance of these differences will require further sampling of chloroplast genomes to resolve. Lastly, this study demonstrates that broomcorn millet grains can preserve significant levels of DNA for hundreds if not thousands of years and can serve as an important genetic resource to investigate the evolution and domestication of this cereal.

Materials And Methods

Areni-1 archaeology of botanical material

To identify the plant material at Areni-1, a comprehensive archaeobotanical sampling strategy was adopted for the cave sediment. Sediment samples with a volume of 5 L were collected throughout the excavation and separated using dry sieving instead of flotation. The majority of plant remains found at Areni-1 were desiccated and contact with water was likely to damage this material. After passing each sediment sample through a 1 mm sieve, material >1 mm was kept for study and < 1mm was discarded, which allowed for some small weed seeds and plant parts to be lost. Similar approaches to sieving have been used at other sites that contained copious amounts of desiccated plant material. Sieved material was hand sorted into categories of plant remains, bones, and pottery fragments. Millet grains were among the plant material recovered from Areni-1 excavations, which also included other domesticated
cereals including wheat and barley. To date, all millet found at Areni-1 came from Trench 1 and 3 within the cave site and in stratigraphic horizons attributed to medieval occupations.

Areni-1 millet grain morphology and identification

Archaeobotanical remains from Areni-1 were identified by comparing morphological features on ancient specimens to those on securely identified modern comparative materials. This approach accounts for both inter- and intra-species variability and assumes that the overall morphology of seeds of any given species is consistent through time. Since the remains were preserved via desiccation rather than charring, colour and husk texture could also be used to help identify the specimens. The archaeological seeds were photographed and measured using a Nikon AZ microscope with associated NIS Elements software and were visually compared with modern seeds representing a variety of species within both the Panicum and Setaria genera held within the botanical reference collection curated within the Archaeobotany Laboratory at the University of Connecticut.

Carbon dating of botanical material from Areni-1

A single Vitis sp. pedicel from Trench 1 and a pool of three millet grains from Trench 3 (Fig. 1) were sent to the NSF Arizona AMS Laboratory (University of Arizona) for radiocarbon dating. Fraction Modern Carbon and Radiocarbon Age were calculated as weighted averages of combined machine runs to reduce overall error.

Areni-1 millet grains for molecular studies

For genetic studies, five millet grains were chosen at random for the extraction and enrichment of cpDNA (Table 1). Two of the millet grains (11294a and 11294b) were found in soil excavated from Trench 1 located in the first gallery of the cave, while the remaining grains (11292a, 11293a, and 11295a) were excavated near a round medieval structure located under the overhang of the cave in Trench 3 (Fig. 1a and Fig S1).

Areni-1 millet library construction and cpDNA enrichment

Extraction of aDNA and construction of Illumina libraries from the Areni-1 millet was performed in a dedicated low-DNA laboratory were no work had previously been done with millet. Isolation of aDNA from the Areni-1 grains and negative blank controls was performed with a two-step protocol, an initial extraction using a guanidine thiocyanate/detergent lysing solution (plant DNAzol) followed by a purification using silica matrix spin columns, which has been previously described. Extracted aDNA was initially converted into truncated Illumina sequencing libraries that included a unique set of internal
barcodes for each sample. The truncated libraries were treated with the enzyme cocktail USER to partially remove deaminated cytosines and then amplified by PCR to produce material for shotgun sequencing or enrichment of cpDNA. Enrichment of cpDNA was performed using RNA capture baits purchased from Arbor Biosciences (https://arborbiosci.com/products/targeted-sequencing-kits/). All Areni-1 libraries were sequenced on various Illumina platforms using 2 x 150 bp (300 cycle) chemistry. A detailed description of these protocols is given in the Supplemental Methods.

Mapping of Areni-1 millet sequencing data

Raw FASTQ files demultiplexed on the i7 adapter index were obtained directly from the sequencer. Run-specific FASTQ files were demultiplexed by the internal barcodes using Sabre 1.0 (https://github.com/najoshi/sabre). AdapterRemoval v2.2.1 was used to trim adapters, collapse reads, discard reads < 25 bp, and remove reads of low quality (< 4). Collapsed reads were mapped to a *P. miliaceum* chloroplast reference genome (Genbank #: KU343177.1) using BWA aln (v0.5.11-foss-2016b) with parameters recommended for aDNA (-l 1024 -n 0.01 -o 2) and a mapping quality (MAPQ) ≥ 30. Duplicate reads were removed from the mapped data using Picard Tools (v 2.2.4: https://broadinstitute.github.io/picard/index.html). Levels of nucleotide misincorporation caused by deaminated cytosine were assayed using mapDamage 2.0 in cpDNA enriched libraries. Read depths were extracted from cpDNA enriched mapped data using SamTools (v1.3.1-foss-2016a) and plotted using R (v3.4.2). Read length distributions for the shotgun and cpDNA enriched libraries were produced using SeqKit (v 0.7.2) and SamToools and plotted with R using the ggplot2 package. For all libraries, except the shotgun library for 11294a, read length analysis was performed using 19,000 reads, a number based on the smallest number of mapped reads produced by a library in this dataset. The shotgun library for 11294a produced 72 mapped reads and this number was used for the read length analysis.

Sequencing of modern broomcorn millet accessions

DNA was extracted from seedling leaves of 50 modern broomcorn millet accessions (Table S1) using the DNAeasy Plant Mini Kit (QIAGEN) and the provided instructions. Shotgun libraries were constructed from the isolated DNA using a NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs) using the protocol provided by the manufacture. Sequencing of the modern libraries was performed at Core Facility for Genomics at Shanghai Center for Plant Stress Biology (Shanghai, China) on an Illumina HiSeq 2500 platform using 2 x 125 bp (250 cycle) chemistry.

Mapping of modern broomcorn millet sequencing data

Fastq reads from the modern broomcorn millet accessions were firstly mapped to a broomcorn millet nuclear reference genome (NCBI Assembly: GCA_003046395.2) using BWA mem (v0.5.11-foss-2016b:
https://arxiv.org/abs/1303.3997v2). From the resulting bam file, reads that did not map to the nuclear reference were extracted using SAMtools (v1.12) and remapped to the broomcorn millet chloroplast genome (Accession: KU343177.1) using BWA mem. Duplicates were marked with Picard Tools to prevent processing of these reads in downstream analysis. Mapping to the chloroplast reference genome produced average read depths of between 1356 and 4548 across the chloroplast genome (Table S2).

Variant calling in the Areni-1 and modern millet sequencing data

Variants were called in parallel in the Areni-1 and modern millet using freebayes (v 1.0.2-GCC-4.9.3-binutils-2.25: Garrison and Marth, 2012, arXiv:1207.3907) with the following parameters: --min-base-quality 30 --min-coverage 10 --hwe-priors-off --ploidy 1 --pooled-continuous --genotype-qualities --min-alternate-fraction 0.95 to generate vcf files. The value for the --min-alternate-fraction was chosen to mask polymorphic sites caused by promiscuous cpDNA (chloroplast DNA that has migrated to the nuclear or mitochondrial genomes) 61. Freebayes automatically disregards reads marked as duplicate, so only unique reads from the modern millet were analysed. Chloroplast genome consensus sequences were generated for the Areni-1 millet using the FastaAlternateReferenceMaker function of GATK (v 3.5-Java-1.8.0_71) 62 using the freebayes vcf files and the KU343177.1 reference. Two sections of the Areni-1 consensus sequences, bp 81851 to 104489 and 117101 to 139826, were removed using Geneious Prime using 11295a as a guide (Build 2019-04-26) 63, because these regions are the IRs of the chloroplast genome were ancient reads could not be mapped (Fig. 3).

Paniceae phylogenetic analysis

Chloroplast reference genomes for 10 members of the Paniceae tribe of grasses (which includes members of the Panicum genus) were downloaded from the NCBI website (Table S1). The reference sequence for the Zea mays (maize) chloroplast genome was also downloaded for use as an out-group. Although the Grass Phylogeny Working Group II, which used three genetic markers, placed Whiteochloa capillipes in the Cenchrinae subtribe 64, this species was included in the analysis because a more recent study of complete chloroplast genomes placed W. capillipes in the Paniceae subtribe 65. The Areni-1 consensus sequences and the grass references were aligned using MAFFT (v 7.130-foss-2016b-with-extensions) and the default settings 66.

Millet variety 69 from the modern accessions (Table S2) was included the MAFFT alignment to serve as the broomcorn millet chloroplast genome. After alignment, the Paniceae genomes were trimmed of the IRs with Geneious Prime using sample 11295a as before. Poorly aligned regions were removed from the alignment using the trimAl program with the default parameters on the Phylemon 2.0 website (http://phylemon2.bioinfo.cipf.es) 67. A maximum likelihood tree was built for the alignment with RAxML 68 (v. 8.2.10) with 1,000 rapid bootstraps and the “GTRGAMMA” model. Tree visualization was
performed with FigTree (v 1.4.3; http://tree.bio.ed.ac.uk/software/figtree/) with rooting on the branch containing the maize chloroplast genome.

**Broomcorn millet phylogenetic and haplotype analysis**

The translated genotypes of the Areni-1 and modern millet were extracted from the freebayes vcf file and exported as a multiple sequence file using bcftools (v1.12) \(^{59}\). An MAFFT alignment with the IRs and poorly aligned regions removed was produced as before. A RAxML tree was generated using the methods described above with rooting on the branch containing the wild broomcorn millet varieties (SampleID: 31 and 167, Table S2). A PHYLIP format file generated from the MAFFT alignment using Geneious Prime was used to produce a haplotype network in the PopART program (http://popart.otago.ac.nz/index.shtml) with the median joining network algorithm \(^{69}\).

**Declarations**

**Data Availability**

Sequencing data for the Areni-1 samples can be found at the NCBI website under BioProject: PRJNA575768 and the chloroplast sequencing data for the 50 modern broomcorn millet accessions are deposited at figshare: https://doi.org/10.6084/m9.figshare.14914605.

**Acknowledgements**

The excavations at Areni-1 were directed by Boris Gasparyan (Institute of Archaeology and Ethnography, National Academy of Sciences, Armenia and Gfoeller Renaissance Foundation, USA) and co-directed by Ron Pinhasi (School of Archaeology, University College Dublin, Ireland) and Gregory Areshian (Cotsen Institute of Archaeology at UCLA, USA). The sorting of biological material was primarily supervised by Tamara Bagoyan (Institute of Archaeology and Ethnography, National Academy of Sciences, Armenia).

**Funding**

Laboratory work for the Areni-1 millet was funded by the following grant: LP130100648: Identifying the diversity and evolution of loci associated with adaptation to aridity/heat and salinity in ancient cereal crops awarded to MG and AC from the Australian Research Council. Archaeobotany work was partly funded by a Curtiss T. and Mary G. Brennan Award to AS. Sequencing of the 50 modern broomcorn millet accessions was funded by a grant from the National Natural Science Foundation of China (31900189) awarded to LL and grants from Shanghai Science and Technology Committee (17391900200), National Natural Science Foundation of China (31922008), and CAS Youth Innovation Promotion Association (Y201844) awarded to HZ.
Competing Interests

The authors declare no competing interests.

Author Contribution

SMR and LL initiated and designed the study. NH, BG, and AS provided the ancient broomcorn millet. HZ and LL provided the modern accessions of broomcorn millet. SMR and LL generated the data, whilst all authors contributed to analysis and interpretation of data. SMR wrote the initial draft of the manuscript and Supplementary Information. All authors reviewed and edited the manuscript and Supplementary Information.

References

1. Habiyaremye, C. et al. Proso Millet (Panicum miliaceum L.) and Its Potential for Cultivation in the Pacific Northwest, U.S.: A Review. *Frontiers in Plant Science* 7, doi:10.3389/fpls.2016.01961 (2017).
2. Miller, N. F., Spengler, R. N. & Frachetti, M. Millet cultivation across Eurasia: Origins, spread, and the influence of seasonal climate. *The Holocene* 26, 1566–1575 https://doi.org/10.1177/0959683616641742 (2016).
3. Cannarozzi, G. et al. Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (Eragrostis tef). *BMC Genomics* 15, 581 https://doi.org/10.1186/1471-2164-15-581 (2014).
4. Motuzaite-Matuzeviciute, G., Staff, R. A., Hunt, H. V., Liu, X. & Jones, M. K. The early chronology of broomcorn millet (Panicum miliaceum) in Europe. *Antiquity* 87, 1073–1085 https://doi.org/10.1017/S0003598X00049875 (2013).
5. Lu, H. et al. Earliest domestication of common millet (Panicum miliaceum) in East Asia extended to 10,000 years ago. *Proceedings of the National Academy of Sciences* 106, 7367-7372, doi:10.1073/pnas.0900158106 (2009).
6. Hunt, H. V. et al. Genetic evidence for a western Chinese origin of broomcorn millet (Panicum miliaceum). *The Holocene*, 0, 0959683618798116 https://doi.org/10.1177/0959683618798116 (2018).
7. Leipe, C., Long, T., Sergusheva, E. A., Wagner, M. & Tarasov, P. E. Discontinuous spread of millet agriculture in eastern Asia and prehistoric population dynamics. *Science advances*, 5, eaax6225–eaax6225 https://doi.org/10.1126/sciadv.aax6225 (2019).
8. Stevens, C. J., Shelach-Lavi, G., Zhang, H., Teng, M. & Fuller, D. Q. A model for the domestication of Panicum miliaceum (common, proso or broomcorn millet) in China. *Vegetation History and Archaeobotany*, https://doi.org/10.1007/s00334-020-00804-z (2020).
9. Filipović, D. et al. New AMS 14C dates track the arrival and spread of broomcorn millet cultivation and agricultural change in prehistoric Europe. *Sci. Rep.*, 10, 13698 https://doi.org/10.1038/s41598-020-70495-z (2020).
10. Spengler, R. N. Agriculture in the Central Asian Bronze Age. *J. World Prehist*, **28**, 215–253 (2015).

11. Martin, L. *et al.* The place of millet in food globalization during Late Prehistory as evidenced by new bioarchaeological data from the Caucasus. *Sci. Rep*, **11**, 13124 https://doi.org/10.1038/s41598-021-92392-9 (2021).

12. Hunt, H. V. *et al.* Millets across Eurasia: chronology and context of early records of the genera Panicum and Setaria from archaeological sites in the Old World. *Vegetation History and Archaeobotany*, **17**(5), https://doi.org/10.1007/s00334-008-0187-1 (2008).

13. Zhang, J. *et al.* Early Mixed Farming of Millet and Rice 7800 Years Ago in the Middle Yellow River Region, China. *PLOS ONE*, **7**, e52146 https://doi.org/10.1371/journal.pone.0052146 (2012).

14. Xu, Y. *et al.* Domestication and Spread of Broomcorn Millet (Panicum miliaceum L.) Revealed by Phylogeography of Cultivated and Weedy Populations. *Agronomy*, **9**, 1 (2019).

15. Yang, X. *et al.* Early millet use in northern China. *Proceedings of the National Academy of Sciences*, **109**, 3726-3730, doi:10.1073/pnas.1115430109 (2012).

16. Wang, J. *et al.* Revealing a 5,000-y-old beer recipe in China. *Proceedings of the National Academy of Sciences*, **113**, 6444-6448, doi:10.1073/pnas.1601465113 (2016).

17. Spengler, R. *et al.* Early agriculture and crop transmission among Bronze Age mobile pastoralists of Central Eurasia. Proceedings of the Royal Society B: Biological Sciences 281(2014).

18. Zeder, M. A. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences*, **105**, 11597, doi:10.1073/pnas.0801317105 (2008).

19. Christian, D. Silk Roads or Steppe Roads? The Silk Roads in World History. *Journal of World History*, **11**, 1–26 (2000).

20. Areshian, G. E. *et al.* The chalcolithic of the Near East and south-eastern Europe: discoveries and new perspectives from the cave complex Areni-1, Armenia. *Antiquity*, **86**, 115–130 (2012).

21. Pinhasi, R. *et al.* First Direct Evidence of Chalcolithic Footwear from the Near Eastern Highlands. *PLoS ONE*, **5**, e10984 https://doi.org/10.1371/journal.pone.0010984 (2010).

22. Wilkinson, K. N. *et al.* Areni-1 Cave, Armenia: A Chalcolithic–Early Bronze Age settlement and ritual site in the southern Caucasus. *Journal of Field Archaeology*, **37**, 20–33 https://doi.org/10.1179/0093469011Z.0000000002 (2012).

23. Smith, A., Bagoyan, T., Gabrielyan, I., Pinhasi, R. & Gasparyan, B. in Stone Age of Armenia, A Guidebook to the Stone Age Archaeology in the Republic of Armenia (eds Boris Gasparyan & Makoto Arimura) 233-260(Center for Cultural Resource Studies, Kanazawa University, 2014).

24. Barnard, H., Dooley, A. N., Areshian, G., Gasparyan, B. & Faull, K. F. Chemical evidence for wine production around 4000 BCE in the Late Chalcolithic Near Eastern highlands. *Journal of Archaeological Science*, **38**, 977–984 https://doi.org/10.1016/j.jas.2010.11.012 (2011).

25. Zohary, D., Hopf, M. & Weiss, E. Domestication of plants in the Old World: the origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin. Fourth edition edn,
26. Khan, A. et al. First complete chloroplast genomics and comparative phylogenetic analysis of Commiphora gileadensis and C. foliacea: Myrrh producing trees. *PLOS ONE, 14*, e0208511 https://doi.org/10.1371/journal.pone.0208511 (2019).

27. Brotherton, P. et al. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nature communications, 4*, 1764 https://doi.org/10.1038/ncomms2656 (2013).

28. Richards, S. M. et al. Low-cost cross-taxon enrichment of mitochondrial DNA using in-house synthesised RNA probes. *PLOS ONE, 14*, e0209499 https://doi.org/10.1371/journal.pone.0209499 (2019).

29. Pääbo, S. & Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America, 86*, 1939–1943 (1989).

30. Dabney, J., Meyer, M. & Pääbo, S. Ancient DNA Damage. *Cold Spring Harbor perspectives in biology, 5*, 1–6 (2013).

31. Brotherton, P. et al. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res, 35*, 5717–5728 (2007).

32. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters., *29*, 1682–1684 https://doi.org/10.1093/bioinformatics/btt193 (2013).

33. Shi, J. et al. Chromosome conformation capture resolved near complete genome assembly of broomcorn millet. *Nature communications, 10*, 464 https://doi.org/10.1038/s41467-018-07876-6 (2019).

34. Zou, C. et al. The genome of broomcorn millet. *Nature communications, 10*, 436 https://doi.org/10.1038/s41467-019-08409-5 (2019).

35. Diao, X. & Jia, G. in Plant Genetics and Genomics: Crops and Models Vol. Volume 19 (ed Richard A Jorgensen) Ch. 4, 61-72(Springer International Publishing, 2017).

36. Nesbitt, M. & Summers, G. D. Some Recent Discoveries of Millet (Panicum miliaceum L. and Setaria italica (L.) P. Beauv.) at Excavations in Turkey and Iran. *Anatolian Studies, 38*, 85–97 https://doi.org/10.2307/3642844 (1988).

37. Max, N. & DID ANCIENT GREEKS DRINK BEER? *Phoenix, 68*, 27–46 https://doi.org/10.7834/phoenix.68.1-2.0027 (2014).

38. Zarikian, N. & Gasparyan, B. Micromammal remains from Areni-1 Cave, Armenia. *IOSR Journal of Humanities and Social Science, 21*, 20–25 (2016).

39. Campana, M. G., McGovern, T. & Disotell, T. Evidence for Differential Ancient DNA Survival in Human and Pig Bones from the Norse North Atlantic. *International Journal of Osteoarchaeology, 24*, 704–708 https://doi.org/10.1002/oa.2255 (2014).
40. Pääbo, S. et al. Genetic analyses from ancient DNA. *Annual Review of Genetics, 38*, 645–679 (2004).

41. Mascher, M. et al. Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nature Genet*, 48, 1089–1093 https://doi.org/10.1038/ng.3611 (2016). http://www.nature.com/ng/journal/v48/n9/abs/ng.3611.html#supplementary-information

42. Aliscioni, S. S., Giussani, L. M., Zuloaga, F. O. & Kellogg, E. A. A molecular phylogeny of Panicum (Poaceae: Paniceae): tests of monophyly and phylogenetic placement within the Panicoideae. *American Journal of Botany, 90*, 796–821 https://doi.org/10.3732/ajb.90.5.796 (2003).

43. Hunt, H. V. et al. Reticulate evolution in Panicum (Poaceae): the origin of tetraploid broomcorn millet, *P. miliaceum*. *Journal of Experimental Botany, 65*, 3165–3175 https://doi.org/10.1093/jxb/eru161 (2014).

44. Viljoen, E., Odeny, D. A., Coetzee, M. P. A., Berger, D. K. & Rees, D. J. G. Application of Chloroplast Phylogenomics to Resolve Species Relationships Within the Plant Genus *Amaranthus*. *Journal of Molecular Evolution, 86*, 216–239 https://doi.org/10.1007/s00239-018-9837-9 (2018).

45. Zhou, T. et al. The Complete Chloroplast Genome of *Euphrasia regelii*, Pseudogenization of ndh Genes and the Phylogenetic Relationships Within Orobanchaceae. *Frontiers in Genetics, 10*, https://doi.org/10.3389/fgen.2019.00444 (2019).

46. Hunt, H. V. et al. Genetic diversity and phylogeography of broomcorn millet (*Panicum miliaceum L.*) across Eurasia. *Mol. Ecol, 20*, 4756–4771 https://doi.org/10.1111/j.1365-294X.2011.05318.x (2011).

47. Yu, W. B., Huang, P. H., Li, D. Z. & Wang, H. Incongruence between Nuclear and Chloroplast DNA Phylogenies in *Pedicularis Section Cyathophora* (Orobanchaceae). *PLOS ONE, 8*, e74828 https://doi.org/10.1371/journal.pone.0074828 (2013).

48. Leigh, F. J. et al. Using diversity of the chloroplast genome to examine evolutionary history of wheat species. *Genet. Resour. Crop Evol, 60*, 1831–1842 https://doi.org/10.1007/s10722-013-9957-4 (2013).

49. Stegemann, S., Hartmann, S., Ruf, S. & Bock, R. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proceedings of the National Academy of Sciences 100*, 8828, doi:10.1073/pnas.1430924100 (2003).

50. Nakazono, M. & Hira, A. Identification of the entire set of transferred chloroplast DNA sequences in the mitochondrial genome of rice. *Molecular and General Genetics MGG, 236*, 341–346 https://doi.org/10.1007/BF00277131 (1993).

51. Van der Veen, M. *Consumption, trade and innovation. Exploring the botanical remains from the Roman and Islamic Ports at Quseir a-Qadim, Egypt*20–33(Africa Magna Verlag, 2011).

52. Chomczynski, P., Mackey, K., Drews, R., Wilfinger, W. & DNAzol® A Reagent for the Rapid Isolation of Genomic DNA. *BioTechniques, 22*, 550–553 https://doi.org/10.2144/97223pf01 (1997).

53. Rohland, N., Glocke, I., Aximu-Petri, A. & Meyer, M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nature Protocols, 13*, 2447–2461 https://doi.org/10.1038/s41596-018-0050-5 (2018).
54. Knapp, M., Stiller, M. & Meyer, M. Generating barcoded libraries for multiplex high-throughput sequencing. *Methods in Molecular Biology, 840*, 155–170 (2012).

55. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 370*, https://doi.org/10.1098/rstb.2013.0624 (2015).

56. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes, 9*, 88 https://doi.org/10.1186/s13104-016-1900-2 (2016).

57. Schubert, M. *et al.* Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics, 13*, 178 https://doi.org/10.1186/1471-2164-13-178 (2012).

58. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Genome research, 25*, 1754–1760 https://doi.org/10.1101/biostats10324 (2009).

59. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Nature Methods, 25*, 2078–2079 https://doi.org/10.1093/bioinformatics/btp352 (2009).

60. Shen, W., Le, S., Li, Y., Hu, F. & SeqKit A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE, 11*, e0163962 https://doi.org/10.1371/journal.pone.0163962 (2016).

61. Scarcelli, N. *et al.* Intra-individual polymorphism in chloroplasts from NGS data: where does it come from and how to handle it? *Mol. Ecol. Resour, 16*, 434–445 https://doi.org/10.1111/1755-0998.12462 (2016).

62. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research, 20*, 1297–1303 https://doi.org/10.1101/gr.107524.110 (2010).

63. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England), 28*, 1647–1649 https://doi.org/10.1093/bioinformatics/bts199 (2012).

64. Edwards, E. J. New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytol, 193*, 304–312 https://doi.org/10.1111/j.1469-8137.2011.03972.x (2012).

65. Burke, S. V. *et al.* Evolutionary relationships in Panicoid grasses based on plastome phylogenomics (Panicoidae; Poaceae). *BMC Plant Biology, 16*, 140 https://doi.org/10.1186/s12870-016-0823-3 (2016).

66. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution, 30*, 772–780 https://doi.org/10.1093/molbev/mst010 (2013).

67. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England), 25*, 1972–1973 https://doi.org/10.1093/bioinformatics/btp348 (2009).

68. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*, 1312–1313 https://doi.org/10.1093/bioinformatics/btu033 (2014).
69. Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution, 16*, 37–48 https://doi.org/10.1093/oxfordjournals.molbev.a026036 (1999).

**Figures**

**Figure 1**

Topographic Plan of Areni-1 Cave and examples of excavated millet

a: Topographic plan of Areni-1 Cave with excavation trenches highlighted in red. The millet grains used in the present study were excavated from Trenches 1 and 3. b: A - Areni-1 millet dorsal, B - Areni-1 millet ventral, C - modern comparative P. miliaceum subsp. miliaceum grain (NPGS accession: 170588, Turkey), and D - modern comparative Setaria italica subsp. grain (NPGS accession: 173103, Turkey). NPGS: U.S. National Plant Germplasm System (https://www.ars-grin.gov/npgs/index.html).
Figure 2

Read depth of mapped cpDNA. Read depth of cpDNA enriched libraries from the three Areni-1 millet grains mapped to the chloroplast reference genome: Genbank# KU343177.1. The inverted repeat regions (IRs) are shown in the large tracks of the graph which lack mapped reads.

![Graph showing read depth of cpDNA](image)

Figure 3

Read length distribution and mapDamage analysis. a: Read length distributions of the shotgun and cpDNA enriched libraries made from three Areni-1 millet grains. All libraries were mapped to the Panicum miliaceum chloroplast reference genome: Genbank# KU343177.1. The shape of the violin displays the frequency distribution of the read lengths in each library. Inside each violin, the grey rectangle represents the interquartile range of the read lengths, while the white circle is the median 37. b: The nucleotide damage pattern of the cpDNA enriched library from the 11295a grain was assayed using mapDamage 2. 36. The enriched library exhibited the C → T nucleotide misincorporations pattern typical of authentic aDNA. The cpDNA libraries from grains 11292a and 112924a gave similar results and are presented in Fig S2.
Figure 4

Paniceae chloroplast phylogeny A phylogeny generated with chloroplast genomes (sans inverted repeat regions) from 11 modern Paniceae species and 3 Areni-1 medieval millet grains using the RAxML program with the GTRGAMMA model 41. The tree was rooted on the outgroup Zea mays (maize). The scale bar at the bottom of the phylogeny represents the substitutions per site.
Figure 5

Broomcorn millet phylogeny and haplotype network analysis a: A RaxML chloroplast phylogeny of Areni-1 millet and 50 modern accessions of broomcorn millet. The tree was rooted on the branch containing the two wild broomcorn millet (31,167). The scale bar at the bottom of the cladogram represents the substitutions per site. b: A haplotype network of 3 medieval Areni-1 millet chloroplast genome and 50 modern accessions of broomcorn millet was generated using the PopArt program (http://popart.otago.ac.nz/index.shtml) with the median joining network algorithm 42. The hash marks on the lines connecting the circles indicate the number of mutations separating the haplotypes. The accessions contained in the core haplotype, the large node at centre of network, are listed in the upper right of the figure.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf