A European Mitochondrial Haplotype Identified in Ancient Phoenician Remains from Carthage, North Africa

Elizabeth A. Matisoo-Smith[1]*, Anna L. Gosling[1,2], James Boocock[2], Olga Kardailsky[1], Yara Kurumilian[3], Sihem Roudesli-Chebbi[4], Leila Badre[5], Jean-Paul Morel[6], Leïla Ladjimi Sebaï[4], Pierre A. Zalloua[3]*

1 Department of Anatomy and Allan Wilson Centre, University of Otago, Dunedin, New Zealand,
2 Department of Biochemistry, University of Otago, Dunedin, New Zealand, 3 School of Medicine, Lebanese American University, Byblos, Lebanon, 4 National Heritage Institute, Tunis, Tunisia, 5 Archaeological Museum, American University of Beirut, Beirut, Lebanon, 6 Université d'Aix-Marseille, Centre Camille Jullian, Aix-en-Provence, France

* matisoo-smith@otago.ac.nz (EAM-S); pierre.zalloua@lau.edu.lb (PAZ)

Abstract

While Phoenician culture and trade networks had a significant impact on Western civilizations, we know little about the Phoenicians themselves. In 1994, a Punic burial crypt was discovered on Byrsa Hill, near the entry to the National Museum of Carthage in Tunisia. Inside this crypt were the remains of a young man along with a range of burial goods, all dating to the late 6th century BCE. Here we describe the complete mitochondrial genome recovered from the Young Man of Byrsa and identify that he carried a rare European haplogroup, likely linking his maternal ancestry to Phoenician influenced locations somewhere on the North Mediterranean coast, the islands of the Mediterranean or the Iberian Peninsula. This result not only provides the first direct ancient DNA evidence of a Phoenician individual but the earliest evidence of a European mitochondrial haplogroup, U5b2c1, in North Africa.

Introduction

The Phoenicians are recognized as one of the great early civilizations of the Mediterranean. First recorded as the descendants of the Canaanites, they inhabited the shores of the eastern Mediterranean and dominated the maritime trade routes of both the eastern and, later, the western Mediterranean during the second and first millennium BCE. The creators of the first alphabet, the Phoenicians documented their own records on papyrus and parchment which, unfortunately, disintegrate over time leaving behind limited historical information. The main Phoenician coastal cities, Tyre, Sidon, Byblos and Arwad, located in what is now Lebanon and southern Syria, have been continuously occupied, so rarely subjected to major archaeological excavations. As a result, we actually know little about the Phoenicians other than what was written about them by the Greeks and Egyptians [1].

At the advent of the mass migration and destruction caused by the Sea-peoples throughout the eastern Mediterranean in the later part of the 2nd millennium, the Phoenician city-states

Competing Interests: The authors have declared that no competing interests exist.

were the only ones to escape destruction, with the notable exception of Arwad, which had been under Hittite occupation [2, 3]. It was the clearing off of neighboring powers by the Sea-peoples that allowed the Phoenicians to re-assert their independence and in fact explore and move farther westward establishing a vast array of colonies throughout the Mediterranean region.

Known for their voyaging and trading skills, craftsmanship, and their access to the highly prized purple dye, from which the Greeks derived the name for these maritime traders, the Phoenicians had an important influence across the Mediterranean and beyond. While there have been claims that they had circumnavigated the African continent, there is clear evidence that they reached as far westward as the Atlantic coasts of Spain and Morocco [4]. It has also been suggested that Phoenician contact via the tin route reached as far as the south of Britain [4], though no archaeological evidence to date has been found to support this claim.

The primary strategy for Phoenician expansion and influence was the establishment of trading posts in cities across the Mediterranean where mobile Phoenician traders would interact with locals. However, they also established a few true and permanent Phoenician colonies. The most well known Phoenician settlement was the city of Carthage in Tunisia, founded by Queen Elissa (also known as Dido) and her followers who arrived from the Phoenician city of Tyre around 813 BCE. Carthage grew from a small Phoenician trading port to one of the most powerful cities in the Mediterranean. Phoenician domination was to be eventually eroded and ultimately replaced by the Greek and later the Roman might in the Mediterranean. Carthage was destroyed by the Romans in the Third Punic War (149–146 BCE), and by the end of the first millennium, the world only knew of the Phoenicians from biased Roman and Greek history books [1].

Byrsa Hill was the highest point in the city of Carthage and was the site of a Phoenician acropolis. Today it is the location of the National Museum of Carthage. In 1994, gardeners planting a tree at the front of the Museum discovered a 4 meter deep shaft which lead to a tomb holding two carved sandstone sarcophagi. One of the sarcophagi was empty, but the other contained the undisturbed skeletal remains of a young man. Funerary goods placed in the sarcophagus and above the lid of the tomb include two commercial Punic amphorae, a lamp, an ivory box of ointments, food offerings and several amulets which all date to the late sixth century BCE [5].

An osteological analysis of the young man from Byrsa, or Ariche, as he has become known, determined that he was approximately 1.7 m tall and aged between 19 and 24 years, and a craniometric analysis indicated likely Mediterranean/European ancestry as opposed to African or Asian [5, 6]. A full reconstruction of the young man was undertaken and this was part of an international exhibition organized by the National Heritage Institute of Tunis and the Carthage Museum. In 2014 the skeletal remains of the young man from Byrsa and accompanying exhibition were taken to the Archaeological Museum of the American University of Beirut where we were allowed to remove small bone samples for ancient DNA analyses to identify the origins of the young man and shed light on the genetic ancestry of the ancient Phoenicians of Carthage.

Our previous research on the genetic ancestry of the Phoenicians focused on analyses of modern Y chromosome variation in historically Phoenician and neighboring non-Phoenician sites across the Mediterranean [7]. Given that the Phoenician traders were primarily men, it was recognised that there should be remnants of the genetic "footprint" of Phoenician Y chromosomes spread across the region. Weak but systematic signatures were recognised which linked the Y chromosome STR haplotypes across Phoenician sites, with six Y-STR markers of likely Phoenician introduction identified. Here we present the first ancient DNA study of Phoenician remains and the first complete ancient Phoenician mitochondrial genome.

Materials and Methods

Ethics Statement

All modern Lebanese samples were collected from donors after they had given their written informed consent to the project and to the data analysis. This project was approved by the IRB of the Lebanese American University.

Ancient DNA sampling

Two small pieces of rib were removed from the skeleton of the young man from Byrsa, under sterile conditions, and were sent to the University of Otago aDNA facility for analyses.

DNA extraction

All DNA extraction and Illumina sequencing library preparation before PCR amplification were conducted in the purpose-built University of Otago aDNA facility [8].

The bone was treated with UVC light ($\lambda = 254$nm) for 45 minutes on each side prior to grinding with a sterile pestle and mortar. The interior cancellous portion of the rib was targeted for extraction which also reduces the chances of including surface contamination in the extraction process. Around 250 mg of bone was used for DNA extraction following the silica based extraction protocol [9]. An extraction blank was run alongside the sample.

Preparation of Illumina Sequencing Libraries

An Illumina sequencing library was produced directly from the aDNA extract by using custom Illumina shotgun adapters [10]. The extraction blank was treated as a normal sample. To the samples, a P7 and P5 adapter were ligated to produce Illumina sequencing libraries. The number of PCR cycles to reach plateau were determined using quantitative PCR (qPCR) on a Strategene MxPro 3000P system using SYBR Green dye (ABI). Products from the qPCR were run out on 2% agarose gels and checked for the presence of adapter-ligated libraries. The libraries were then indexed and "immortalised" by PCR amplification. To reduce PCR artefacts, the amplification was stopped once plateau was reached. Immortalised libraries were purified using Qiagen MinElute purification columns following the manufacturer's protocol.

In-Solution Hybridisation Capture

One PCR of 1 μL immortalised library in a 50 μL reaction was performed to obtain 2 μg of Illumina sequencing library. Only 10 cycles of PCR were run at this point to avoid reaching the PCR plateau point. The library was purified and concentrated using Qiagen MinElute purification columns following the manufacturer's protocol. Human mtDNA was enriched in the Illumina library by in-solution hybridization capture [11] with some modifications. Modern human DNA of a known European haplogroup (T2b) was used to produce the hybridization capture probe. In contrast to the published protocol [11], the library was capture-enriched individually. The enriched library was then amplified via PCR for 20 cycles, stopping well before the plateau was reached. The post-capture library was then purified using Qiagen MinElute purification columns following the manufacturer's protocol. The resulting library was quantified using a Qubit, prior to sequencing on an Illumina MiSeq sequencing platform in a 2 x 75 base paired-end run.

Modern DNA extraction and sequencing

The complete mitochondrial genomes for 47 modern Lebanese samples, all previously identified as belonging to mitochondrial haplogroup U, based on HVR analysis [12], were amplified in two long range PCR products, produced using primers HUM-LR1 and HUM-LR2 described previously [13]. The PCR products were pooled and 1 μg of the products, in equal molar concentrations, were mechanically sheared using sonication to produce fragments of approximately 500 bp in length. Blunt end repair, ligation of sequencing adaptors, sample barcoding and pooling were carried out following Meyer and Kircher [14] and Kircher and Kelso [15] with modifications for Illumina sequencing adaptors. Pooled samples were sequenced on one lane of the Illumina MiSeq in a 2 x 250 base paired-end run with version 2 chemistry. All members of the research team who handled the ancient Phoenician sample, both in the collection process and in the laboratory, were also sequenced to determine their mitochondrial haplotypes to assess the possibility of contamination of the ancient sample.

Ancient DNA processing and analysis

Raw reads from the ancient Phoenician were processed to remove sequencing adaptors and merge paired end fragments, which overlapped by at least 11 base-pairs, using Adapter-Removal (v1.5.4) [16]. Reads were also processed to remove stretches of Ns, bases that had a low quality score (<30), and short reads (<25). To reduce the impact of contamination due to laboratory reagents, all read mapping was done onto a composite reference genome consisting of the Cambridge Reference Sequence for humans (*Homo sapiens*, GenBank NC_012920), in addition to mitochondrial reference genomes from cow (*Bos taurus*, GenBank NC_006853), pig (*Sus scrofa*, NC_0012095.1), and chicken (*Gallus gallus*, GenBank NC_001323.1). The contamination ratio was determined by calculating the ratio of all reads with mapping quality > = 20 for each reference, compared with those for the human mitochondrial reference sequence [17].

Ancient DNA is usually characterised by C→T transitions at the 5' ends of the molecule, and following double stranded library preparation G→A transitions at the 3' ends of the molecules [18]. Reads were aligned to the composite reference genome using BWA (0.7.10) [19] with recommended ancient DNA settings [20]; specifically, seeding was disabled (-l 1024), the number of gap opens was set to 2 (-o 2), and the maximum edit distance was set to 0.03 (-n 0.03). PCR duplicates were removed from the merged reads using a python script originally developed for the Neanderthal genome project [21]. These PCR duplicates were also removed from the unmerged reads using Picard's MarkDuplicates tool (http://broadinstitute.github.io/picard/). To confirm the authenticity of our ancient DNA, the program mapDamage (v2.0.2–9) [22] which identifies characteristic aDNA damage patterns, was used with the '-rescale' option to lower the quality score of likely damaged sites. The plots of characteristic damage patterns were individually created for merged and unmerged reads. Finally, Ts (thymines) found at the 5' end of a read and Gs (guanines) at the 3' end of the read, within the first two bases, had their quality scores rescaled to zero. The presence of modern human DNA in the sequencing libraries presents a challenge for ancient DNA, as it can introduce major biases into downstream population genetic and phylogenetic [23, 24]. Modern human mtDNA contamination levels were estimated using the ContamMix package [25].

Reads that mapped to the human reference genome were extracted from the PMD filtered BAM file and a variant call file (VCFs) was generated using the GATK Haplotype Caller (v3.3) [26] with settings specific for haploid genomes, such as the mitochondrial genome. This file was filtered for mapping quality (<20), and bases supported by fewer than three reads [27, 28]. This approach was justified under the assumption that the above quality control (QC)

procedure preferentially reduces the quality of the damaged positions in the sequencing reads, which would lead to a reduction in the probability that the damaged position would be called downstream. A coverage plot and a fragment length plot were created for the Phoenician samples using the ggplot2 package for the R programming language [29]. The fragment length plot was created only using the merged reads, this was done because estimation of the fragment length is error prone using unmerged reads. A consensus sequence containing indels (insertions and/or deletions) was created, and has been deposited in GenBank (KT760574). In addition to the aforementioned QC measures, non-variant sites that were supported by fewer than three reads were changed to 'Ns'. To assign a haplogroup to the Phoenician, we used HaploGrep [30] (http://haplogrep.uibk.ac.at/) for Phylotree 16 [31]. All Illumina reads generated for the ancient Phoenician have been submitted to the NCBI short read archive as sample MS10148, accession number PRJNA295854.

Modern sequence processing and analysis

The raw reads from the modern Lebanese samples were mapped to the human mitochondrial reference genome (*Homo sapiens*, GenBank NC_012920) using the BWA [19] (v 0.7.10) maximal exact matches (MEM) command [32]. This was followed by PCR duplicate removal using Picard's MarkDuplicates command. Haplogroups were assigned by first generating a joint-called VCF using the GATK HaplotypeCaller, converting the VCF into the HaploGrep input format, and finally assigning a haplogroup to each Lebanese sample with HaploGrep [30] (http://haplogrep.uibk.ac.at/). All 47 complete modern Lebanese mitochondrial genome sequences have been submitted to GenBank (accession numbers KT779157-KT779203).

Phylogenetic Analysis

To investigate the phylogenetic relationship of the ancient Phoenician (KT760574) to other U5b2c sequences, we obtained the publicly available U5b2c sequences from GenBank and the supplementary information of a recent publication investigating present day mitochondrial diversity in Portugal [33]. A maximum likelihood (ML) tree was generated with MEGA using the HKY + I model. Preliminary analyses revealed that the sequence EF758625 and KU587510 had many missing positions reducing the phylogenetic resolution, thus these samples were removed from the final analysis.

Results

The complete ancient mtDNA genome

A total of 11,129,232 paired-end reads were returned for the ancient Phoenician. After bioinformatics processing of these raw reads, which included adaptor removal, PCR duplicate removal, read merging, and quality filtering, 4,851 reads remained which mapped uniquely (MAPQ > 20) to the human mitochondrial reference genome. This represents an endogenous mitochondrial DNA fraction of 0.0004%. The final read-depth for the ancient Phoenician was 33.1x (Fig 1). Analysis of the variable sites with the online tool Haplogrep revealed that the ancient Phoenician had the mitochondrial haplotype U5b2c1.

Sequence authenticity

For human ancient DNA studies two potential sources of contamination exist, these are the reagents [34] and modern human DNA [35]. To reduce the impact and to estimate the effect of reagent contamination, we performed all read mapping to a composite mitochondrial reference genome that included the major reagent contaminators cow, chicken, and pig [34]. Only 4
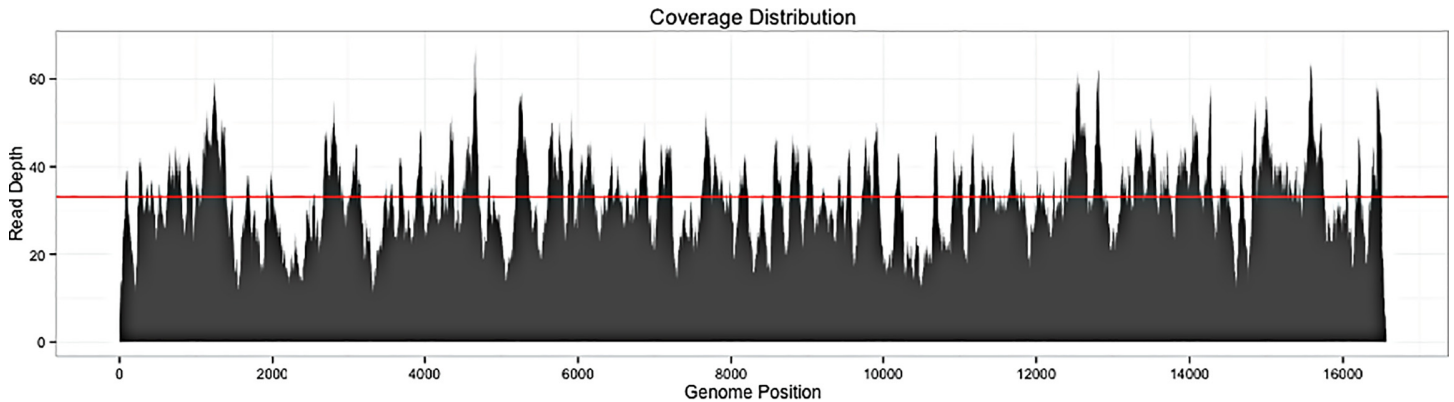
Fig 1. Final read depth and coverage for the mitochondrial genome of the ancient Phoenician from Carthage. The red line shows the mean read-depth of 33.15x.

reads mapped confidently to these genomes. To assess the totality of modern human DNA contamination we used the program ContamMix with a set of 311 representative modern human mtDNA sequences. This estimated that the ancient Phoenician had been contaminated with 1.1% (95% C.I. 0.004%-2.34%) of modern human DNA. The deamination patterns and fragment lengths (Figs 2 and 3) were consistent with those expected from ancient DNA [36]. For example, the observed C to T misincorporation rate for the first base of the 5' end was 0.16. This is outside the maximum rate of 0.05 for modern samples (<117 years old) and within the range of what is to be expected for a sample that dates to the 6th Century BCE (see Fig 2 in [36]).

Haplogroup assignment and phylogenetic analysis

The complete mitochondrial genome of the young man from Byrsa was identified as belonging to haplogroup U5b2c1 possessing all of the defining mutations with the exception of 16192T, one of the two mutations defining U5, which has shown to be unstable with frequent reversions [37]. He also had five additional, coding region mutations (Fig 4). None of the researchers who collected or processed the Phoenician sample have mitochondrial haplotypes belonging to Haplogroup U. All methods to avoid contamination were applied and the possibility of modern contamination in the sequence obtained was assessed using standard aDNA authenticity criteria and shown to be unlikely, thus we are confident that the haplotype identified is indeed that of the young man from Byrsa.

A phylogenetic analysis of all U5b2c sequences available in GenBank or from published supplementary data (Z2478) is shown in Fig 5.

Modern mitochondrial genome sequencing

Given the reputed Lebanese origins of the founders of Carthage, we undertook full mitochondrial genome sequencing of 47 modern Lebanese samples that had previously been typed to Haplogroup U through analysis of the HVR-1 [12]. Haplogroups identified and haplogroup frequency are shown in Table 1. Only seven of the modern Lebanese samples belonged to Haplogroup U5 and of those, two were U5b, but neither belonged to U5b2c or derived haplotypes.

Discussion and Conclusion

Haplogroup U5 is considered to be one of the most ancient haplogroups in Europe and is believed to have arisen there [43]. The coalescence time estimate for U5 is 29.6 kya (22.7–37.2
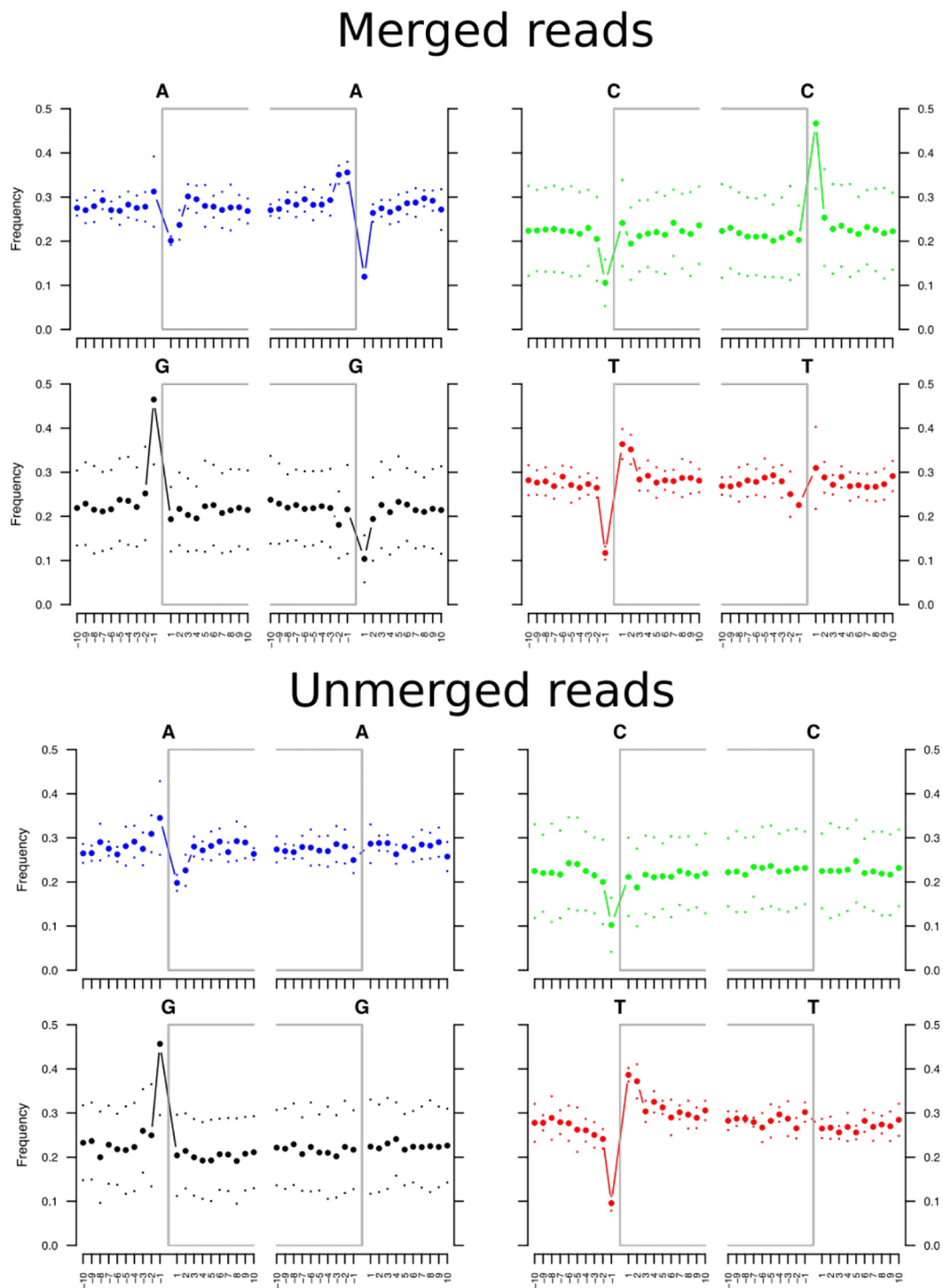
Fig 2. Base frequency 5' and 3' of strand breaks. The gray brackets indicate the start and end of molecules (strand breaks). Frequencies are displayed for A, G, C, and T for the 10 bases 5' and 3' of the breaking site. The top panel shows that merged reads have an elevated frequency of purines (A and G) before strand breaks; this is consistent with the DNA molecules being ancient [36]. The bottom panel shows that unmerged reads have an elevated frequency of purines (A and G) before 5' strand breaks but not 3' strand breaks, this has occurred because mapDamage utilizes only the first read of a pair to calculate these frequencies. As the end of the first-read does not represent the end of the molecule, we would not expect to see an elevated purine frequency at the 3' breaking site. Therefore, these results are consistent with the longer unmerged reads being ancient.

doi:10.1371/journal.pone.0155046.g002

Merged reads
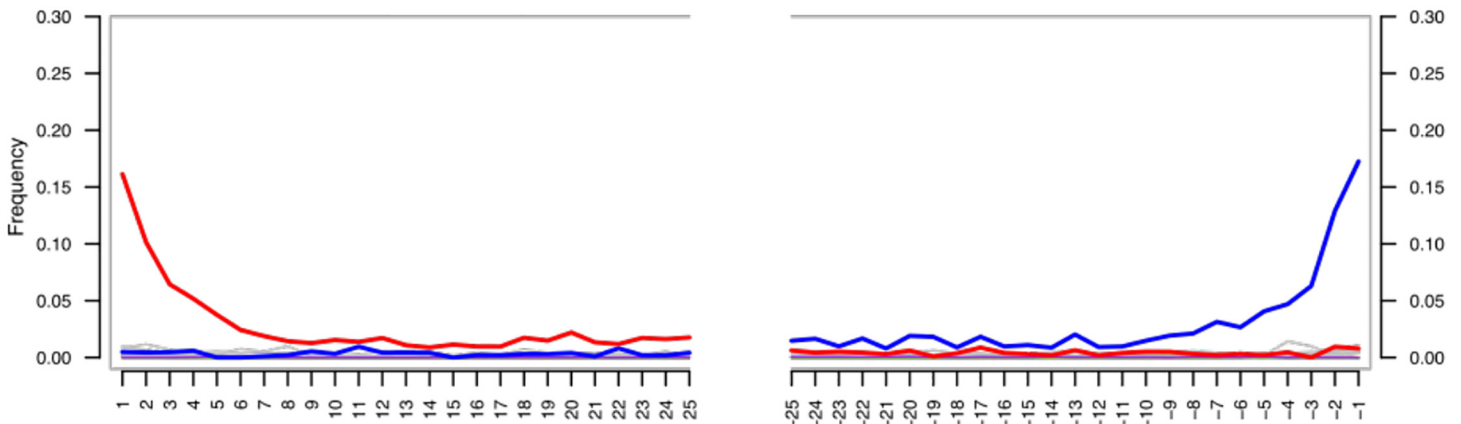
Unmerged reads
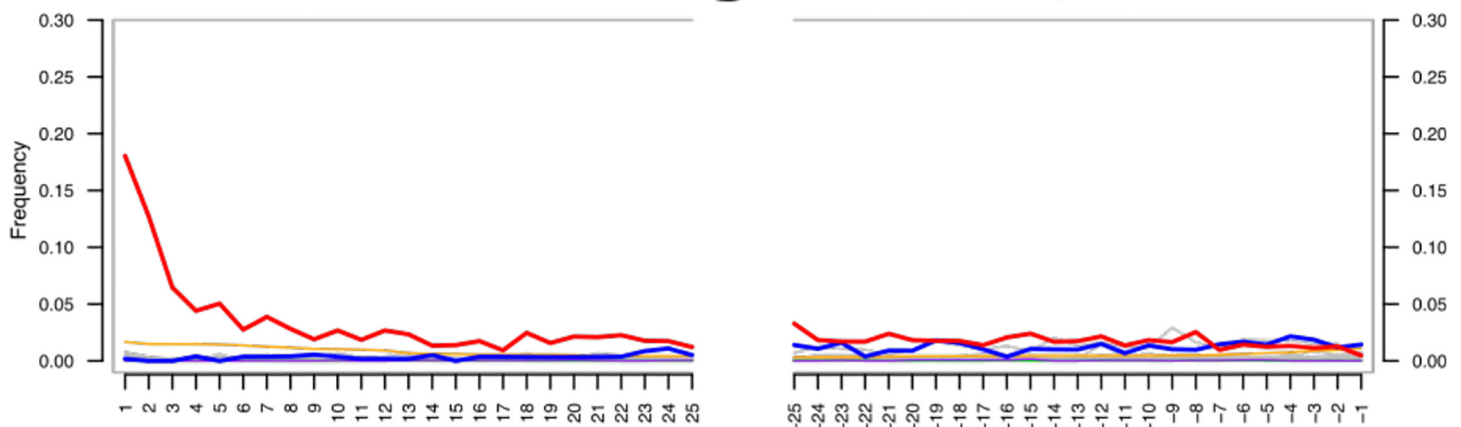
Fig 3. C to T nucleotide misincorporations for the first and last 25 bases of endogenous mtDNA fragments. Red T; green C; blue A; purple G. The top panel shows that merged reads have an increased frequency of T at the 5' end and A at the 3' end (G to A misincorporation on the opposite strand of C to T misincorporation), a typical pattern of ancient DNA damage [35, 36]. The bottom panel shows that unmerged paired-end reads have an elevated frequency of T at the 5' end, but in contrast with the above panel no increase in A at the 3' end. This has occurred because mapDamage utilizes only the first read of a pair to calculate these frequencies. As the end of the first-read does not contain the sequence of the entire molecule, we would not expect to see an elevated A frequency at the end of the read. Therefore, these results are consistent with the unmerged reads originating from larger DNA fragments being ancient.

doi:10.1371/journal.pone.0155046.g003

kya) [25] and 20–24 kya for U5b [44]. It is not uncommon in Mesolithic European populations, particularly those from Central and Eastern Europe [45]. Haplogroup U5b2c1 has been identified in both La Braña 1 and 2, the 7000 year-old remains recovered from the La Braña-

11467G, 12308G, 12372A ➔ **U** ➔16192T, 16270T ➔ **U5** ➔ 3197C, 9477A, 13617C ➔ **U5a'b** ➔ 150T, 7768G, 14182C ➔ **U5b** ➔ 1721T, 13637G ➔ **U5b2** ➔ 723G, 960.1C, 13017G ➔ **U5b2c** ➔ 6920A, 13434G ➔ **U5b2c1** ➔ 3882A, 5351G, 5773A, 6023A, 9869T

Fig 4. Key variable sites from rCRS found in the young man from Byrsa, with the marker path of mutations defining Haplogroup U to U5b2c1. Note, the mutation 16192T was not present in our sample.

doi:10.1371/journal.pone.0155046.g004
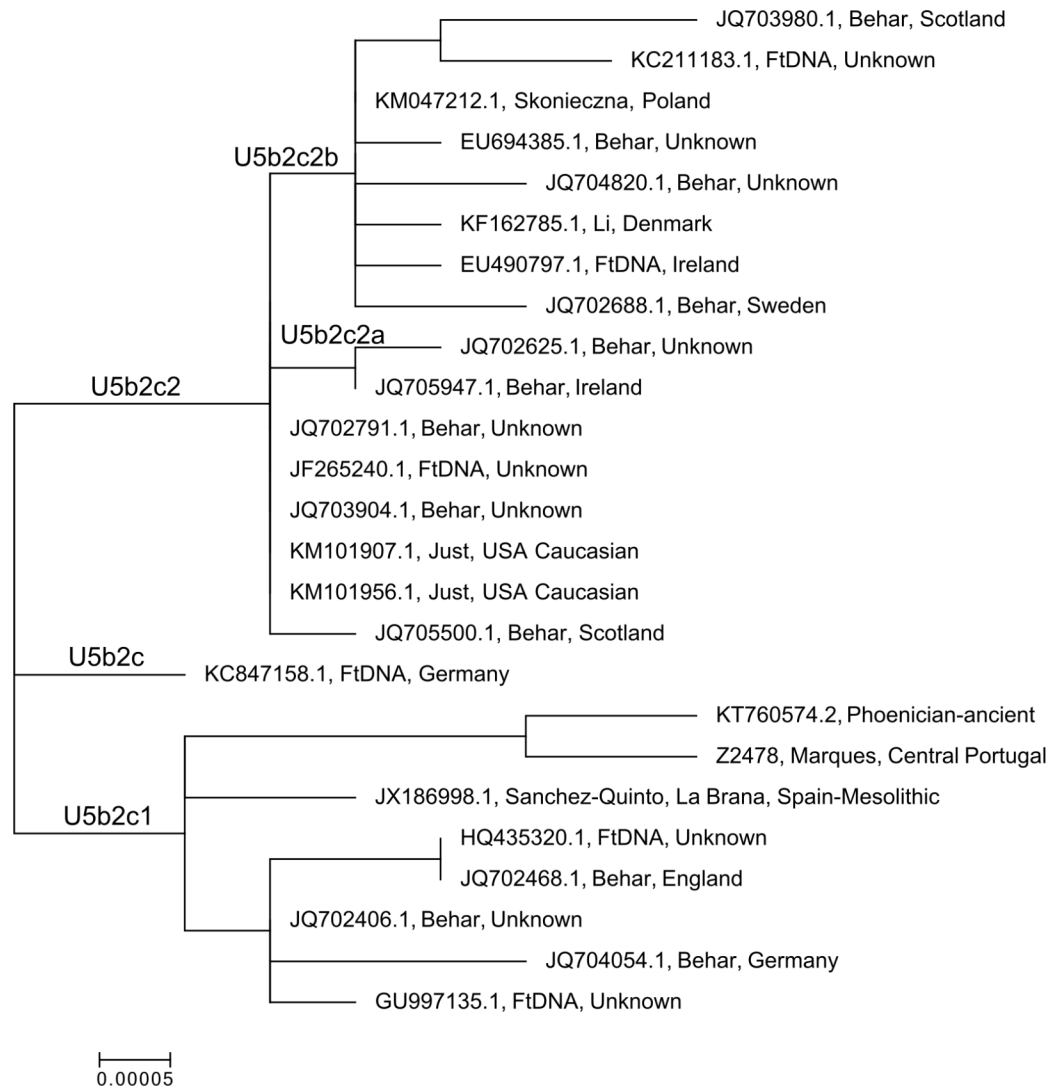
Fig 5. A maximum likelihood (ML) tree for the ancient Phoenician (KT760574) and other publicly available U5b2c sequences. All samples other than our Phoenician and the La Braña sample are from modern populations. Each node is annotated with the GenBank accession number or sample identification (e.g. Z2478 from [33]), source or author if published, and the origin of the sample if recorded [33, 38–42].

doi:10.1371/journal.pone.0155046.g005

Arintero site in León in Northwestern Spain [42]. Our Phoenician differed from the La Braña 1 complete mitochondrial genome at eight sites (positions 3882, 5351, 5773, 6023, 9869, 16069, 16126, and 16192). The mutations at sites 16069 and 16126 appear to be private mutations for La Braña.

It appears that the U5b2c1 haplogroup is rare in modern populations, with only a few modern sequences published [33, 38, 46] or available in public databases (Family Tree DNA). All of the reported U5b2c1 carriers are of presumably (if not specifically stated) European ancestry, from Spain, Portugal, England, Ireland, Scotland, the United States and Germany. Three of the additional non-defining mutations found in our Phoenician, 5351G, 6023A, and 9869T, are shared with one "European" sample [46] and an individual from central Portugal [33]. Interestingly, Fig 5 shows that our Phoenician sample is most closely related to the modern sample from central Portugal. The unidentified "European" sequence (EF758625) which was deleted

Table 1. Mitochondrial haplogroups identified via complete mitochondrial genome sequencing in 47 modern Lebanese previously identified (based on HVR-1) as belonging to Haplogroup U.

Haplogroup	N
U1a1	1
U1a1a	3
U1b	4
U2d2	2
U2e1'2'3	1
U3a	4
U3a1	1
U3b	7
U3b1	2
U3b2a	4
U3b2a1	2
U3b3	2
U3c	1
U4b3	2
U5a1a1g	2
U5a1g	2
U5a2b	1
U5b2b	1
U5b3a2	1
U6a7a1b	1
U6b1a	1
U7b	1
U8a1a	1
Total	**47**

doi:10.1371/journal.pone.0155046.t001

from the analysis due to missing data, also carries the three mutations which define the branch on which we find the Phoenician and the Portuguese sample. A separate branch within U5b2c1 contains five samples, from England and Germany or otherwise unidentified as to location of origin, with the La Braña Mesolithic sample (JX186998) on its own branch. Given the limited numbers of published full mitochondrial genomes, it is difficult to identify a likely origin for the mutations defining the Phoenician and Portuguese branch, but it is currently not inconsistent with a Southwest European origin. While the U5b2c2 sequences all originate in Central- and North-Western European or likely derived populations (e.g. US Caucasian), and the U5b2c sequences in GenBank (KC847158 and KU587510, which was excluded from the analysis due to missing data) originate in Germany and England respectively (both samples from Family Tree DNA), it is likely that U5b2c originated in Central/Northwestern Europe. Future complete mitochondrial genome sequencing of European populations, both ancient and modern, will undoubtedly provide more information regarding the distribution and frequency of the U5b2c1 haplogroup across the continent and thus will help us to reconstruct its evolutionary history.

It is generally argued that Mesolithic populations carrying high frequencies of U5 were replaced by Neolithic populations expanding into Western Europe during the Neolithic transition [47]. Recent analyses of ancient DNA suggest that there were two waves of Neolithic expansions that dramatically influenced the genetic makeup of Western Europe [48–51]. An early Neolithic expansion out of Anatolia, occurring some 7000–9000 years ago, moved into

Western Europe and replaced many Mesolithic Hunter-Gatherer populations, resulting in an increase of Near Eastern mitochondrial lineages. After this initial invasion of Neolithic farmers, there appears to have been a resurgence in the Western Hunter-Gatherer genetic signatures, followed by a second migration and replacement in many locations from the Eurasian Steppes, associated with the spread of the Yamnaya culture, beginning some 4500 years ago [49]. Several other more regionally focused aDNA studies, however, have shown that there are different patterns of population replacements in various regions of Europe [52, 53]. Unlike the early Neolithic arrival in the Iberian peninsula, which likely arrived in coastal regions [54, 55], it is possible that the second, Eurasian Steppe derived, replacement occurred later or had less of an impact on the southern and western Iberian coast and other Mediterranean coastal regions than in more northern or eastern inland locations [49, 53] and thus the Western Hunter-Gatherer mtDNA lineages, including U5b2c1, may have been more common there at the time of Phoenician contact and settlement.

While the U5b frequency in modern populations in Western Europe is less than 2% it is much rarer in today's Near Eastern populations [12]. In our full mitogenome sequencing results of 47 modern Lebanese U samples, we found only two individuals who carried haplotypes belonging to U5b2 and neither of these belonged to the U5b2c1 subgroup. It has been demonstrated, however, that modern Levantine populations do not reliably represent the haplogroups present in the Neolithic period, though haplogroup U5 was only found at low frequencies in pre-pottery Neolithic (PPNB) samples from the Levant [54]. U5b is found in only 0.4% of the modern populations from the Iberian peninsula, and only 0.18% Europe-wide [42], which provides us with further confidence of the authenticity of our ancient DNA result. Only the Saami, in northern Scandinavia, retain high levels of U5, and U5b1b in particular, with frequencies over 50% in some Saami populations [56].

Interestingly, haplogroup U5 and U5b have been identified in modern populations from North and Northwest Africa [57–59]. It must be noted, however, that there are few North African complete mitogenomes publicly available and HVR sequencing alone cannot identify beyond the U5b haplogroup (with the C150T HVR mutation) since the seven derived defining mutations (C1721T and A13637G, which define U5b2; A723G, 960XC and A13017G, which define U5b2c; and C6920a and A13484G, which define U5b2c1) are all in the coding regions. Achilli, et al. [59], using full mitochondrial genome sequencing identified a U5b1b1 cluster that grouped Amazigh (North African Berbers) and Saami populations. This cluster is based on the control region motif (16270–150) which is present at low frequencies in Amazigh, North African and nearly all European populations with the exception of the Scandinavian Saami where it is at about 48%. The divergence time of this cluster is around 8600 years ago (+/- 2400) consistent with an expansion from Franco-Cantabrian refuge which is believed to have been a major refuge for the European hunter-gatherers prior to their post LGM expansion [44]. It is suggested that European haplogroup U5 and the more prevalent U6 "Berber cluster" diverged from a common ancestor in the Near East and spread along the north and south coasts, respectively, of the Mediterranean, as far as Iberia to the north and Cyrenaica to the south [60]. It is very plausible that descendants of the Mesolithic hunter-gatherers carried U5b1b1 and sister lineages across the Straits of Gibraltar into North Africa [59], but there is no indication of when this migration may have happened. While it is possible that U5b2 haplogroups were also carried across the Straits of Gibraltar prior to Phoenician arrival in North Africa, our result now provides a minimum date of arrival. We can now say that U5b2c1 was present by the late 6th century BCE.

Though it is believed that Carthage was established by colonisers from Tyre, in what is today Lebanon, it is unlikely that our Phoenician young man had maternal ancestry that traced back to this founding population as the haplogroup U5b2c has not been identified in our

modern Lebanese samples or in ancient (early Neolithic, PPNB) remains from the Levant [54]. Given that haplogroup U5b2c1 has not been previously reported in North Africa, we suggest that the ancestry of our young man is likely traced to some population across the Mediterranean, which is consistent with his proposed European cranial traits [6].

So how did a young Phoenician man, with a European mitochondrial lineage end up in Carthage, North Africa? The earliest Phoenician site in Iberia is believed to be Gadir or Cadiz as it is known today, established in 1110 BCE [1]. Phoenician colonies were also established in Ibiza, southern Sardinia, western Sicily and along the southern shores of the Iberian Peninsula [1, 61, 62] and, later, these were also major Carthagenian trade ports. It has been suggested that the Phoenician east-west trade route from Tyre and Sidon across the Mediterranean and through the Straits of Gibraltar departed following a northerly route, travelling with the prevailing winds and tides westwards, stopping in Cyprus, Sicily, Ibiza and several locations along the southern coast of Spain, reaching Cadiz and beyond. Return trips back to the Levant took a southerly Mediterranean route, again, with the prevailing currents, sailing along the North African coast [1]. While early Phoenician occupation of and trade in the western Mediterranean would have been from the Levant, Carthage, once established, came to dominate the western Mediterranean trade networks, but would most likely have continued with this well travelled, counterclockwise circular route.

To date, as far as we can determine, U5b2c1 has not been identified in the Balearic Islands, Sardinia, or Sicily [62–64], though it may have been on these islands in the past and lost due to drift. While some of the reported modern individuals carrying the U5b2c1 haplogroup today are found in the Iberian peninsula, they are also found in Britain, Germany, the United States, and undefined "Europe", so we cannot say with certainty that the haplogroup is more than European. However, the highest frequency of the haplogroup U5b today is in the Iberian peninsula and U5b2c1 was also present there in Mesolithic hunter gatherers [42]. While many of the Mesolithic populations carrying these ancient U5b European lineages were replaced by later Neolithic peoples, perhaps some of these lineages persisted longer in the extreme southern part of the Iberian peninsula, on the offshore islands, or other island or coastal locations of Phoenician influence and individuals carrying these ancient European lineages were transported to Carthage via the Phoenician/Punic trade networks.

We know of no other U5b2c haplotypes recorded in North Africa [58]. Our Phoenician young man significantly predates the Moorish expulsion from Iberia (1492 AD) which resulted in the Andalusian communities that were established in Tunisia. Thus we now have a minimum date for the introduction of the European derived U5b2c1 haplotype to North Africa and its direct association with a Phoenician individual. Further work on aDNA from Phoenician remains from throughout the Mediterranean, including the homeland region in Lebanon is on-going. This research will help us better understand the origins and impact of Phoenician peoples and their culture throughout the Mediterranean region and beyond and better reconstruct ancient population migrations and trade and exchange networks and the degree to which these influenced genetic variation seen in the Mediterranean region today.

Author Contributions

References

1. Aubet ME. The Phoenicians and the West. Cambridge: Cambridge University Press; 1997.

2. Gilboa A. Sea Peoples and Phoenicians along the Southern Phoenician Coast—A Reconciliation: An Interpretation of Šikila (SKL) Material Culture. Bulletin of the American Schools of Oriental Research. 2005; 337:47–78.

3. McMahon G. Reflections of a Late Bronze Age Empire: The Hittites. The Biblical Archaeologist. 1989; 52(2/3):62–77.

4. Carpenter R. Phoenicians in the West. American Journal of Archaeology. 1958; 62(1):35–53.

5. Morel J-P. Les Fouilles de Byrsa (Secteur B) a Carthage. Comptes Rendus des Seances 2011; 1:325–63.

6. Institute INdPNH. Exposition (octobre 2010—mars 2011):: «Le jeune homme de Byrsa» au Musée de Carthage [updated Wednesday, 02 January 2013 14:07 cited 2015]. Available: http://www.inp.rnrt.tn/index.php?option=com_content&view=article&id=75:jeune-homme-de-byrsa&catid=3:activites&Itemid=7&lang=en.

7. Zalloua PA, Platt DE, El Sibai M, Khalife J, Makhoul N, Haber M, et al. Identifying Genetic Traces of Historical Expansions: Phoenician Footprints in the Mediterranean. American Journal of Human Genetics. 2008; 83(5):633–42. doi: 10.1016/j.ajhg.2008.10.012 PMID: ISI:000261006900010.

8. Knapp M, Clarke AC, Horsburgh KA, Matisoo-Smith EA. Setting the stage–Building and working in an ancient DNA laboratory. Annals of Anatomy—Anatomischer Anzeiger. 2012; 194(1):3–6. http://dx.doi.org/10.1016/j.aanat.2011.03.008.

9. Rohland N, Hofreiter M. Ancient DNA extraction from bones and teeth. Nature Protocols. 2007; 2(7):1756–62. doi: 10.1038/nprot.2007.247 PMID: ISI:000253139200021.

10. Knapp M, Stiller M, Meyer M. Generating Barcoded Libraries for Multiplex High-Throughput Sequencing. In: Shapiro B, Hofreiter M, editors. Ancient DNA: Methods and Protocols. Methods in Molecular Biology. New York: Humana Press, Springer; 2012. p. 155–70.

11. Maricic T, Whitten M, Pääbo S. Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. PLoS One. 2010; 5(11). e14004. doi: 10.1371/journal.pone.0014004 PMID: WOS:000284257400008.

12. Badro DA, Douaihy B, Haber M, Youhanna SC, Salloum A, Ghassibe-Sabbagh M, et al. Y-Chromosome and mtDNA Genetics Reveal Significant Contrasts in Affinities of Modern Middle Eastern Populations with European and African Populations. PLoS One. 2013; 8(1):e54616. doi: 10.1371/journal.pone.0054616 PMID: 23382925

13. Clarke A, Prost S, Stanton J-A, White WT, Kaplan M, Matisoo-Smith E, et al. From cheek swabs to consensus sequences: an A to Z protocol for high-throughput DNA sequencing of complete human mitochondrial genomes. BMC Genomics. 2014; 15(1):68. doi: 10.1186/1471-2164-15-68

14. Meyer M, Kircher M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harbor Protocols. 2010; 2010(6):pdb.prot5448. doi: 10.1101/pdb.prot5448

15. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 2012; 40(1):e3. Epub 2011/10/25. doi: 10.1093/nar/gkr771 PMID: 22021376; PubMed Central PMCID: PMCPmc3245947.

16. Lindgreen S. AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Research Notes. 2012; 5:337. doi: 10.1186/1756-0500-5-337 PMID: 22748135

17. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 1999; 23(2):147–. PMID: 10508508

18. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A Draft Sequence of the Neandertal Genome. Science. 2010; 328(5979):710–22. doi: 10.1126/science.1188021 PMID: 20448178

19. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754–60. doi: 10.1093/bioinformatics/btp324 PMID: WOS:000267665900006.

20. Schubert M, Ginolhac A, Lindgreen S, Thompson JF, AL-Rasheid KAS, Willerslev E, et al. Improving ancient DNA read mapping against modern reference genomes. BMC Genomics. 2012; 13(178). doi: 10.1186/1471-2164-13-178

21. Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature. 2014; 514(7523):445–9. doi: 10.1038/nature13810 Available: http://www.nature.com/nature/journal/v514/n7523/abs/nature13810.html—supplementary-information. PMID: 25341783

22. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics. 2013; 29(13):1682–4. doi: 10.1093/bioinformatics/btt193 PMID: PMC3694634.

23. Green RE, Briggs AW, Krause J, Prufer K, Burbano HA, Siebauer M, et al. The Neandertal genome and ancient DNA authenticity. EMBO J. 2009; 28(17):2494–502. Epub 2009/08/08. emboj2009222 [pii] doi: 10.1038/emboj.2009.222 PMID: 19661919; PubMed Central PMCID: PMC2725275.

24. Wall JD, Kim SK. Inconsistencies in Neanderthal genomic DNA sequences. PLoS Genetics. 2007; 3 (10):1862–6. doi: 10.1371/journal.pgen.0030175 PMID: ISI:000251072900007.

25. Fu Q, Mittnik A, Johnson Philip LF, Bos K, Lari M, Bollongino R, et al. A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. Current Biology. 2013; 23(7):553–9. doi: http://dx.doi.org/10.1016/j.cub.2013.02.044 PMID: 23523248

26. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–303. doi: 10.1101/gr.107524.110 PMID: 20644199

27. Thalmann O, Shapiro B, Cui P, Schuenemann VJ, Sawyer SK, Greenfield DL, et al. Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. Science (New York, NY). 2013; 342(6160):871–4. doi: 10.1126/science.1243650 PMID: MEDLINE:24233726.

28. Knapp M, Horsburgh KA, Prost S, Stanton J- A, Buckley HR, Walter RK, et al. Complete mitochondrial DNA genome sequences from the first New Zealanders. Proceedings of the National Academy of Sciences. 2012; 109(45):18350–4. doi: 10.1073/pnas.1209896109

29. Team RDC. R: A Language and Environment for Statistical Computing https://www.r-project.org/. In: Computing RFfS, editor. Vienna, Austria 2015.

30. Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, et al. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Human Mutation. 2011; 32(1):25–32. doi: 10.1002/humu.21382 PMID: 20960467

31. van Oven M, Kayser M. Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation. Human Mutation. 2009; 30:E386–E94. doi: 10.1002/humu.20921 PMID: 18853457

32. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM 2013 [cited 2015]. Available: http://arxiv.org/abs/1303.3997v2.

33. Marques SL, Goios A, Rocha AM, Prata MJ, Amorim A, Gusmao L, et al. Portuguese mitochondrial DNA genetic diversity-An update and a phylogenetic revision. Forensic Sci Int Genet. 2015; 15:27–32. Epub 2014/12/03. doi: 10.1016/j.fsigen.2014.10.004 PMID: 25457629.

34. Leonard JA, Shanks O, Hofreiter M, Kreuz E, Hodges L, Ream W, et al. Animal DNA in PCR reagents plagues ancient DNA research. Journal of Archaeological Science. 2007; 34(9):1361–6. doi: 10.1016/j.jas.2006.10.023 PMID: ISI:000248408600003.

35. Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, et al. Analysis of one million base pairs of Neanderthal DNA. Nature. 2006; 444(7117):330–6. doi: 10.1038/Nature05336 PMID: ISI:000242018300040.

36. Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. PLoS One. 2012; 7(3):e34131. doi: 10.1371/journal.pone.0034131 PMID: 22479540

37. Howell N, Smejkal CB. Persistent heteroplasmy of a mutation in the human mtDNA control region: hypermutation as an apparent consequence of simple-repeat expansion/contraction. Am J Hum Genet. 2000; 66(5):1589–98. Epub 2000/04/14. doi: 10.1086/302910 PMID: 10762545; PubMed Central PMCID: PMCPMC1378018.

38. Behar DM, van Oven M, Rosset S, Metspalu M, Loogvali E-L, Silva NM, et al. A Copernican Reassessment of the Human Mitochondrial DNA Tree from its Root. American Journal of Human Genetics. 2012; 90(4):675–84. doi: 10.1016/j.ajhg.2012.03.002 PMID: 22482806

39. Skonieczna K, Malyarchuk B, Jawień A, Marszałek A, Banaszkiewicz Z, Jarmocik P, et al. Heteroplasmic substitutions in the entire mitochondrial genomes of human colon cells detected by ultra-deep 454 sequencing. Forensic Science International: Genetics. 2015; 15:16–20. http://dx.doi.org/10.1016/j.fsigen.2014.10.021

40. Li S, Schierup MH. Variation and association to diabetes in 2000 full mtDNA sequences mined from an exome study in a Danish population. unpublished2013.

41. Just RS, Scheible MK, Fast SA, Sturk-Andreaggi K, Röck AW, Bush JM, et al. Full mtGenome reference data: Development and characterization of 588 forensic-quality haplotypes representing three U.S. populations. Forensic Science International: Genetics. 2015; 14:141–55. http://dx.doi.org/10.1016/j.fsigen.2014.09.021.

42. Sánchez-Quinto F, Schroeder H, Ramirez O, Ávila-Arcos María C, Pybus M, Olalde I, et al. Genomic Affinities of Two 7,000-Year-Old Iberian Hunter-Gatherers. Current Biology. 2012; 22(16):1494–9. doi: 10.1016/j.cub.2012.06.005 PMID: 22748318

43. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, et al. Tracing European Founder Lineages in the Near Eastern mtDNA Pool. The American Journal of Human Genetics. 2000; 67(5):1251–76. PMID: 11032788

44. Malyarchuk B, Derenko M, Grzybowski T, Perkova M, Rogalla U, Vanecek T, et al. The Peopling of Europe from the Mitochondrial Haplogroup U5 Perspective. PLoS One. 2010; 5(4). e10285. doi: 10.1371/journal.pone.0010285 PMID: WOS:000276952500024.

45. Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, et al. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. Science. 2009; 326(5949):137–40. Epub 2009/09/05. 1176869 [pii]. doi: 10.1126/science.1176869 PMID: 19729620.

46. Herrnstadt C, Elson J, Fahy E, Preston G, Turnbull D, Anderson C, et al. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet. 2002; 70(5):1152–71. doi: 10.1086/339933 PMID: 11938495

47. Fu Q, Rudan P, Pääbo S, Krause J. Complete Mitochondrial Genomes Reveal Neolithic Expansion into Europe. PLoS ONE. 2012; 7(3):e32473. doi: 10.1371/journal.pone.0032473 PMID: 22427842

48. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, et al. Genome-wide patterns of selection in 230 ancient Eurasians. Nature. 2015; 528(7583):499–503. doi: 10.1038/nature16152 Available: http://www.nature.com/nature/journal/v528/n7583/abs/nature16152.html—supplementary-information. PMID: 26595274

49. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015; 522(7555):207–11. doi: 10.1038/nature14317 Available: http://www.nature.com/nature/journal/v522/n7555/abs/nature14317.html—supplementary-information. WOS:000356016700037. PMID: 25731166

50. Allentoft ME, Sikora M, Sjogren K-G, Rasmussen S, Rasmussen M, Stenderup J, et al. Population genomics of Bronze Age Eurasia. Nature. 2015; 522(7555):167–72. doi: 10.1038/nature14507 Available: http://www.nature.com/nature/journal/v522/n7555/abs/nature14507.html—supplementary-information. PMID: 26062507

51. Günther T, Valdiosera C, Malmström H, Ureña I, Rodriguez-Varela R, Sverrisdóttir ÓO, et al. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. Proceedings of the National Academy of Sciences. 2015. doi: 10.1073/pnas.1509851112

52. Hervella M, Izagirre N, Alonso S, Fregel R, Alonso A, Cabrera VM, et al. Ancient DNA from Hunter-Gatherer and Farmer Groups from Northern Spain Supports a Random Dispersion Model for the Neolithic Expansion into Europe. PLoS One. 2012; 7(4):e34417. doi: 10.1371/journal.pone.0034417 PMID: 22563371

53. Hervella M, Rotea M, Izagirre N, Constantinescu M, Alonso S, Ioana M, et al. Ancient DNA from South-East Europe Reveals Different Events during Early and Middle Neolithic Influencing the European Genetic Heritage. PLoS ONE. 2015; 10(6):e0128810. doi: 10.1371/journal.pone.0128810 PMID: 26053041

54. Fernández E, Pérez-Pérez A, Gamba C, Prats E, Cuesta P, Anfruns J, et al. Ancient DNA Analysis of 8000 B.C. Near Eastern Farmers Supports an Early Neolithic Pioneer Maritime Colonization of Mainland Europe through Cyprus and the Aegean Islands. PLoS Genet. 2014; 10(6):e1004401. doi: 10.1371/journal.pgen.1004401 PMID: 24901650

55. de Lagrán ÍG- M. Recent Data and Approaches on the Neolithization of the Iberian Peninsula. European Journal of Archaeology. 2015; 18(3):429–53. doi: 10.1179/1461957114Y.0000000084

56. Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogväli E- L, et al. The Western and Eastern Roots of the Saami—the Story of Genetic "Outliers" Told by Mitochondrial DNA and Y Chromosomes. The American Journal of Human Genetics. 2004; 74(4):661–82. PMID: 15024688

57. Rosa A, Brehm A. African human mtDNA phylogeography at-a-glance. Journal of Anthropological Sciences. 2011; 89:25–58. doi: 10.4436/jass.89006 PMID: WOS:000296201800004.

58. Bekada A, Fregel R, Cabrera VM, Larruga JM, Pestano J, Benhamamouch S, et al. Introducing the Algerian Mitochondrial DNA and Y-Chromosome Profiles into the North African Landscape. PLoS ONE. 2013; 8(2):e56775. doi: 10.1371/journal.pone.0056775 PMID: 23431392

59. Achilli A, Rengo C, Battaglia V, Pala M, Olivieri A, Fornarino S, et al. Saami and Berbers—An Unexpected Mitochondrial DNA Link. The American Journal of Human Genetics. 2005; 76(5):883–6. PMID: 15791543

60. Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, et al. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet. 1999; 64(1):232–49. doi: 10.1086/302204 PMID: 9915963

61. Picornell A, Gómez-Barbeito L, Tomàs C, Castro JA, Ramon MM. Mitochondrial DNA HVRI variation in Balearic populations. American Journal of Physical Anthropology. 2005; 128(1):119–30. doi: 10.1002/ajpa.10423 PMID: 15761883

62. Ramon M, Picornell A, J. C. Human population of the Balearic Island: the case of Chuetas and Ibizans. Contributions to Science. 2008; 4(1):85–91.

63. Boattini A, Martinez-Cruz B, Sarno S, Harmant C, Useli A, Sanz P, et al. Uniparental Markers in Italy Reveal a Sex-Biased Genetic Structure and Different Historical Strata. PLoS One. 2013; 8(5). doi: 10.1371/journal.pone.0065441 PMID: WOS:000319725500011.

64. Sarno S, Boattini A, Carta M, Ferri G, Alù M, Yao DY, et al. An Ancient Mediterranean Melting Pot: Investigating the Uniparental Genetic Structure and Population History of Sicily and Southern Italy. PLoS ONE. 2014; 9(4):e96074. doi: 10.1371/journal.pone.0096074 PMID: 24788788