




Article

Y-chromosome and Surname Analyses for Reconstructing Past Population Structures: The Sardinian Population as a Test Case

Viola Grugni ¹, Alessandro Raveane ¹, Giulia Colombo ¹, Carmen Nici ¹, Francesca Crobu ^{1,2}, Linda Ongaro ^{1,3,4}, Vincenza Battaglia ¹, Daria Sanna ^{1,5}, Nadia Al-Zahery ¹, Ornella Fiorani ⁶, Antonella Lisa ⁶, Luca Ferretti ¹ , Alessandro Achilli ¹, Anna Olivieri ¹, Paolo Francalacci ⁷, Alberto Piazza ⁸, Antonio Torroni ¹ and Ornella Semino ^{1,*}

¹ Dipartimento di Biologia e Biotecnologie “L. Spallanzani”, Università di Pavia, 27100 Pavia, Italy; viola.grugni@unipv.it (V.G.); alessandro.raveane01@universitadipavia.it (A.R.); giulia.colombo01@universitadipavia.it (G.C.); nici.carmen@gmail.com (C.N.); f.crobu@irgb.cnr.it (F.C.); linda.ongaro@ut.ee (L.O.); vincenza.battaglia@unipv.it (V.B.); darsanna@uniss.it (D.S.); nadia.alzahery@unipv.it (N.A.-Z.); luca.ferretti@unipv.it (L.F.); alessandro.achilli@unipv.it (A.A.); anna.olivieri@unipv.it (A.O.); antonio.torroni@unipv.it (A.T.)

² Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), 09042 Monserrato, Italy

³ Estonian Biocentre, Institute of Genomics, Riia 23, 51010 Tartu, Estonia

⁴ Department of Evolutionary Biology, Institute of Molecular and Cell Biology, Riia 23, 51010 Tartu, Estonia

⁵ Dipartimento di Scienze Biomediche, Università di Sassari, 07100 Sassari, Italy

⁶ Istituto di Genetica Molecolare “L.L. Cavalli-Sforza”, Consiglio Nazionale delle Ricerche (CNR), 27100 Pavia, Italy; fiorani@igm.cnr.it (O.F.); antonella.lisa@igm.cnr.it (A.L.)

⁷ Dipartimento di Scienza della Vita e dell’Ambiente, Università di Cagliari, 09123 Cagliari, Italy; paolo.francalacci@unica.it

⁸ Dipartimento di Scienze Mediche, Scuola di Medicina, Università di Torino, 10124 Torino, Italy; alberto.piazza@unito.it

* Correspondence: ornella.semino@unipv.it; Tel.: +39-0382-985543

Received: 16 October 2019; Accepted: 14 November 2019; Published: 16 November 2019



Abstract: Many anthropological, linguistic, genetic and genomic analyses have been carried out to evaluate the potential impact that evolutionary forces had in shaping the present-day Sardinian gene pool, the main outlier in the genetic landscape of Europe. However, due to the homogenizing effect of internal movements, which have intensified over the past fifty years, only partial information has been obtained about the main demographic events. To overcome this limitation, we analyzed the male-specific region of the Y chromosome in three population samples obtained by reallocating a large number of Sardinian subjects to the place of origin of their monophyletic surnames, which are paternally transmitted through generations in most of the populations, much like the Y chromosome. Three Y-chromosome founding lineages, G2-L91, I2-M26 and R1b-V88, were identified as strongly contributing to the definition of the outlying position of Sardinians in the European genetic context and marking a significant differentiation within the island. The present distribution of these lineages does not always mirror that detected in ancient DNAs. Our results show that the analysis of the Y-chromosome gene pool coupled with a sampling method based on the origin of the family name, is an efficient approach to unravelling past heterogeneity, often hidden by recent movements, in the gene pool of modern populations. Furthermore, the reconstruction and comparison of past genetic isolates represent a starting point to better assess the genetic information deriving from the increasing number of available ancient DNA samples.

Keywords: Human Y-chromosome variation; haplogroups; family name origin; migrations; peopling of Sardinia; phylogenetics

1. Introduction

Sardinians, albeit clearly Europeans, represent the main outlying gene pool in the European genetic landscape [1–5]. To understand the origin and the evolutionary forces at the basis of their differentiation, Sardinians have been the subject of numerous genetic, linguistic and anthropological analyses. According to linguistic studies, the origin of Sardinians pre-dates the settlement of the Indo-Europeans in western Europe [6–8]. The earliest archaeological evidence of modern humans in the island goes back to the Paleolithic (between 20 to 14 kya), when Sardinia and Corsica were a single land, yet separated from the mainland [9–12]. Initially, the island population was small; it gradually increased in the Neolithic period and later, especially in the Bronze Age with the development of the advanced civilization characterized by the *nuraghi*, megalithic edifices of a circular shape very similar to buildings observed in other islands of the Mediterranean Basin. Afterwards, the population size remained approximately constant until the last three centuries when it underwent a significant growth [11]. Following the first settlement, the most important external contributions were provided by the Phoenicians (9th century BCE) and the Carthaginians (5th century BCE), who controlled the entire island, with the only exception of the region of Olbia, ruled by the Greeks [11]. At the end of the First Punic war (238 BCE), Sardinia passed under the control of Rome. Yet, many archaeological remains prove that the influence of the Romans and previous conquerors was limited to the coastal regions, whereas the mountainous central district of the island, the so-called “archaic zone”, became the refuge of the indigenous non-Indo-European inhabitants of Sardinia (*Nuragians*). Likewise, subsequent invasions by Vandals (456 CE), Byzantines (534 CE), Saracens (7–10 century CE) and Pisans (1052–1295) had a limited impact [13]. With regards to the Spanish, who ruled the island until 1713, they did have an important cultural impact in the North-West of the island, as attested by the language spoken there [14]. Eventually, from 1718 to 1820 Sardinia was annexed by the Savoy.

In spite of this continuous chain of invasions, the gene flow into the Sardinian population has been relevant only in some coastal areas where well-known foreign settlements took place [15]. As a result, the ancient origin of native Sardinians and their long-standing isolation might provide an explanation for their genetic peculiarity [16], characterized by high frequencies of uniparental haplotypes that are rare elsewhere in Europe [2,17–23], extensive linkage disequilibrium of autosomal markers [24], as well as high degrees of homozygosity at the genomic level [25].

However, a main settlement in Neolithic times followed by long-lasting isolation can also explain the extreme similarity at the nuclear genomic level with early European Neolithic farmers [26,27] and with the Late Neolithic/Chalcolithic Tyrolean Iceman [28,29], but not the similarity observed with the Near Eastern Neolithic farmers including those from Anatolia [30]. The small population size (from pre-history to 1700 CE the Sardinian population never exceeded 300 thousand inhabitants, and in around 1348 CE, the Black Plague reduced the population by half), the presence in the island of natural barriers such as mountains, but also the endemic malaria in the lower lands, which kept certain areas very isolated [31], contributed to creating different genetic isolates and consequently heterogeneity within regions. On the whole, three large areas of Sardinia reflecting its ancient history and geography were identified. The northern zone is delimited by the mountain chain crossing Sardinia from the Central-West to the North-East and is linguistically different from the rest of the island. The south-western zone is delineated by the presence of many Phoenician and Carthaginian archeological sites [13]. The central-eastern zone is the asylum land of the ancient Sardinian population during invasions and is a domain of pastoral culture. This zone includes the more conservative or “archaic” area, defined by archaeological, linguistic [32], geo-linguistic and genetic factors [8] (for a more detailed subdivision of Sardinia on the basis of genes, languages and surnames, see [14]).

Although genetic differences are still detectable between communities [33–42], the increasing internal migration toward the main villages and towns of the last 150 years has weakened and sometimes erased the boundaries of these isolates, partially blurring the ancient genetic structure of the island [2,4,20] and making it difficult to reconstruct its past demographic history. However, the results of a previous study indicate that the use of a sampling method based on the geographic origin

of family names (territorial monophyletic family names), in comparison with the usual grandparent's birthplace sample collection strategy, allows, for the Y-specific gene pool at least, the reconstruction of ancient isolates, bypassing the effect of recent migrations [19].

Thus, this study exploited a sampling strategy based on the origin of the family name [19] and a detailed Sardinian Y-chromosome phylogeny [22,43] to reconstruct ancient genetic isolates of the Sardinian male component and to address the following questions: can we detect the ancient heterogeneity in the actual Sardinian gene pool? And, if so, what information does it provide about the early peopling of the island and subsequent migrations?

To answer these questions, Y-chromosome high-resolution analyses were performed on 603 Sardinian males representative of the different zones of the island, after having also assessed the linguistic and geographic origins of their family names. Our results provide new clues for understanding the fine genetic structure of the Sardinian population, an essential piece of information not only in an evolutionary context, but also for reducing confounding effects caused by population structure in association studies.

2. Results

2.1. Classification and Distribution of Y-chromosome Haplogroups in Ancient Isolates of Sardinia

The analysis of 603 subjects with monophyletic surnames allowed the identification of 62 Y-chromosome lineages belonging to 14 main Y-chromosome haplogroups (Hgs). The relative frequencies of the haplogroups observed in the global sample and in the three main areas of Sardinia are listed per haplogroup in the table of Figure 1 and are summarized in Figure 2.

No significant difference in the haplogroup profiles was observed in comparison with previous datasets [18–20,44–46], thus showing that our “monophyletic” sample well represents Sardinian variability.

The most frequent haplogroups are I-M170 (41.7%) (almost exclusively represented by its sub-clade I2-M26 (38.9%)) and R1-M207 (21.1%) (represented mainly by its branch R1b-M269: 18.6%), followed by G-M201 (14.3%) (with its most frequent sub-haplogroup G2-L91: 6.5%) and J-M304 (11.06%) (with its most frequent sub-haplogroup J2-M410: 7.8%). Haplogroups observed in southern Europe such as the Balkan E-V13 (3.0%), the Arab J1-M267 (2.7%), the African E-M33, E-M81, E-V12, E-V22, E-V65 (2.2%) and peculiar haplogroups such as A-M13 (0.5%) common in East Africa, R1b-V88* (0.8%) and its derivative R1b-M18 (0.5%), as well as R2-M124 (1.0%) observed mainly in South West Asia, were also detected. In addition, the rare lineage H2-M282 characterizes 0.8% of our Sardinian sample.

After the re-distribution of the subjects in the different regions of the island according to the origin of their family name (Figure S1), a general heterogeneity in the allocation of many haplogroups emerged ($\chi^2_{[df12]} = 164.96; p < 0.001$) with only 21 out of the 62 defined lineages shared among the three main geographic regions (Table of Figure 1). In particular, the χ^2 per cell analysis (Table S1) showed significantly higher than expected frequencies of haplogroups G2-L91 and R1b-L2 in the northern areas compared to frequencies in the central region, which, in turn, is characterized by a significantly higher incidence (50%) of the lineage I2-M26, represented almost exclusively by its sub-clade I2-L160.

2.2. Sardinian Populations in the Mediterranean Context

In order to evaluate the position of Sardinians in a wider European and Mediterranean population context and visualize the relationships between Sardinians and other Italian populations, a Principal Component (PC) analysis was carried out on haplogroup frequencies, exploiting available literature data normalized to the highest possible level of phylogenetic resolution (Table S2). The plot of the two PCs is shown in Figure 3 together with a plot displaying the contribution of each haplogroup to the first and second PC.

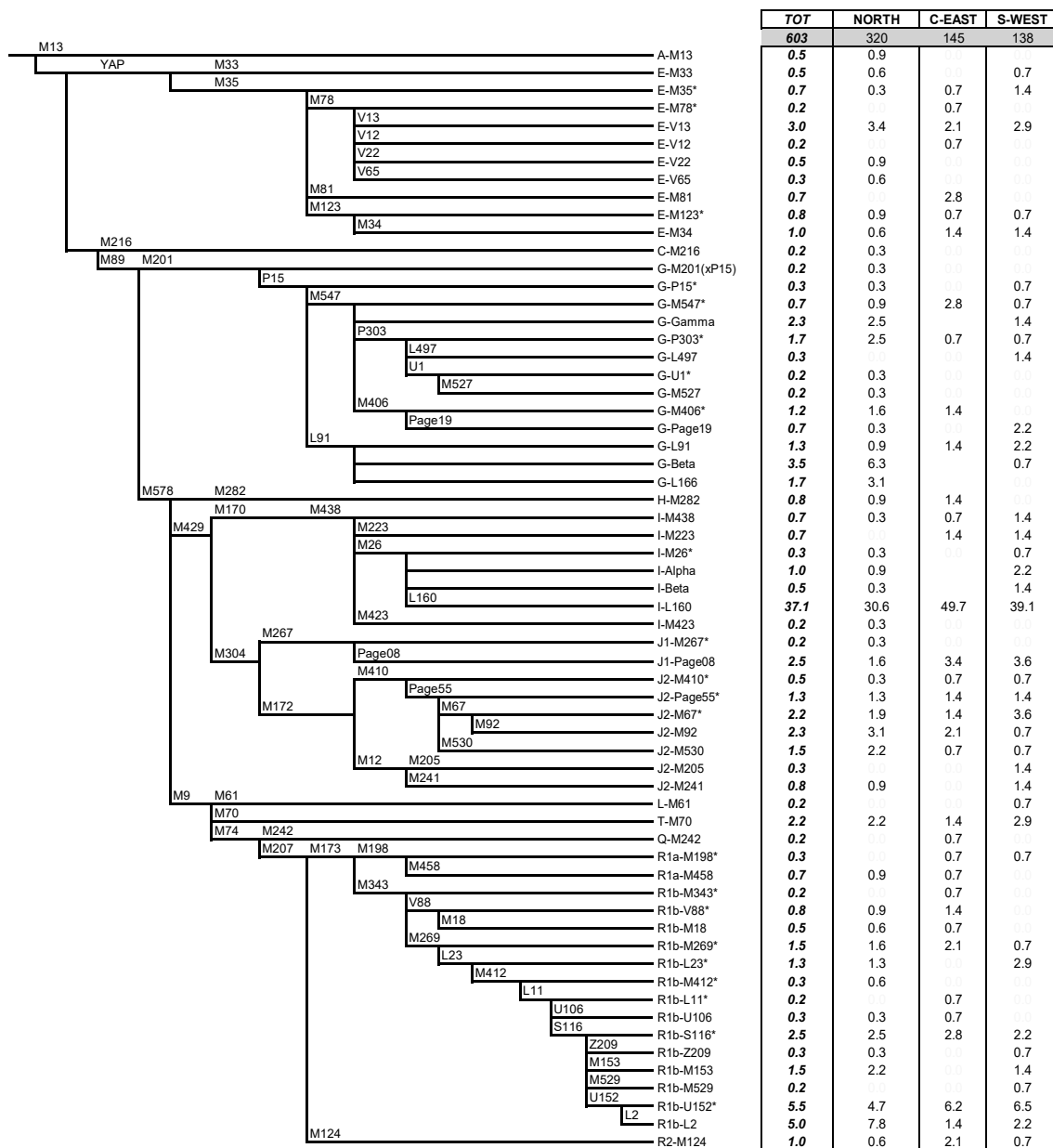


Figure 1. Phylogenetic tree of Y-chromosome haplogroups and their frequencies (as percentage) in the whole Sardinian sample and in the three main geo-cultural regions of the island. Names of markers are indicated above the lines; the lengths of branches are not drawn to scale for better readability. Asterisk (*) indicates a paragroup: group of Y chromosomes defined by the derived state of the main haplogroup but by any downstream mutation.

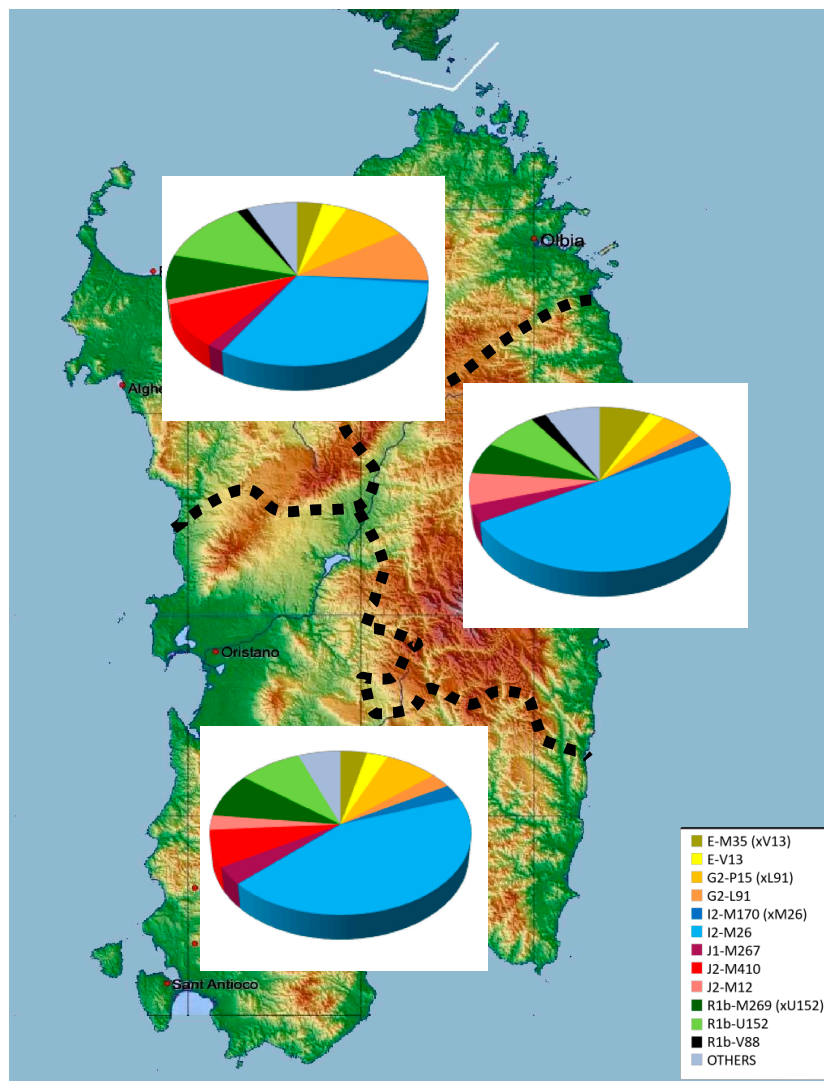


Figure 2. Proportion of the observed Y-chromosome haplogroups in the three main cultural and geographical areas of Sardinia. Dashed lines indicate boundaries of the three main areas.

The distribution, which is based on 27.84% of the total variance, shows an overall general agreement with geography: while the first component separates the populations according to longitude, the second component discriminates the populations approximately according to latitude. The continental populations from lands lying around the Mediterranean Sea are clearly separated at the periphery of the plot in five clusters (North Africa, Spanish, French and Basque populations, Balkans, Caucasus-Anatolia and Middle East). The populations of the Italian Peninsula and the Mediterranean islands follow a North-South cline with the northern-central Italians (together with Corsicans) closer to Spanish and French groups while the southern central Italians (and Sicilians) are closer to Caucasus-Anatolian and middle eastern populations. A similar pattern is also detected when looking at the entire genome, with Corsicans closer to the Central-North Italian groups [25], and Sicilians to Central-South Italian populations [5]. Conversely Sardinians, who strongly behave as outliers at the genomic level [5], appear located at a fringe of the Western European distribution of Y-chromosome variation. Such a position is explained by a markedly higher incidence of haplogroups I2-M26 and G2-L91 and by the presence of the R1b-V88 clade, virtually absent in other European populations [47].

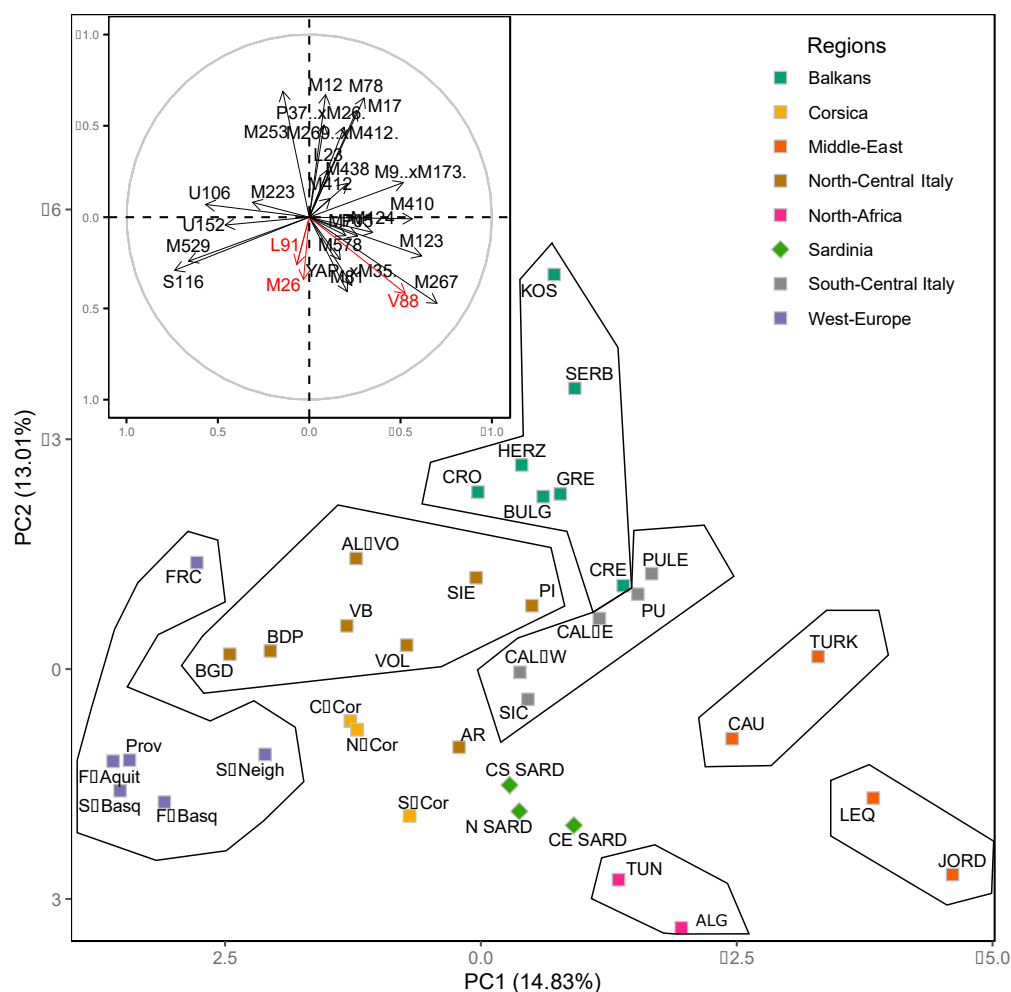


Figure 3. Principal component analysis performed using haplogroup frequencies in the Sardinian samples of the present study compared with those of relevant populations from the literature at the highest possible level of resolution (Table S2). Of the total variance, 27.84% is represented: 14.83% by the first PC and 13.01% by the second PC. The inset illustrates the contribution of each haplogroup. S-Neigh: Spanish from Western Bizkaia, Cantabria, Burgos, La Rioja, North Aragon; S-Basq: Spanish Basques from Roncal, Naffarroa, Gipuzcoa, Araba, Bizkaia; F-Basq: French Basques from Soule, Aquitani; F-Aquit: Franch from Bigorre, Bearn, Chalosse; FRC: France; PROV: Provence; N-COR: Corsica-North; C-COR: Corsica-Central; S-COR: Corsica-South; N SARD: Sardinia-North; CE SARD: Sardinia-Central East; CS SARD: Sardinia-Central South; AL-VO: Alessandria-Voghera, Piedmont; VB: Borbera Valley, Piedmont; BDP: Bergamo Plain, Lombardy; BGD: Bergamo Valley, Lombardy; VOL: Volterra, Tuscany; PI: Pisa, Tuscany; AR: Arezzo, Tuscany; SIE: Siena, Tuscany; PU: Apulia; PULE: Apulia-Grecanica; CAL-E: Calabria Ionian; CAL-W: Calabria Tyrrhenian; SIC: Sicily; CRO: Croatia Mainland; HERZ: Herzegovina; SERB: Serbia; BULG: Bulgaria; GRE: Greece Mainland; CRE: Crete; TURK: Turkey; CAU: Caucasus; LEQ: Lebanon+Iraq; TUN: Tunisia; ALG: Algeria.

2.3. The Founder Paternal Lineages G2-L91, I2-M26 and R1b-V88

Three haplogroups were previously identified as founding lineages: G2-L91, I2-M26 and R1b-V88.

Haplogroup G2-L91 (Table S3) reaches its highest frequency in southern Corsica (22.1%) and North Sardinia (10.5%). It appears at low frequencies elsewhere without any apparent pattern: in Iberia (0.3–1.0%), South France (0.8%), Continental Italy (1–1.2%), Austria (0.4%), Germany (0.3%), Czech Republic (2.9%), Armenia (0.9%), Iran (1.0%), Israel (0.3–0.6%), Egypt (4.1%) and in Moroccan Berbers (0.8%) [48,49]. This haplogroup characterizes also Ötzi, the Tyrolean Copper Age man [28,49].

The Network analysis of G2-L91 haplotypes (Table S4; Figure S2) reveals a composite pattern of evolution characterized by haplotypes widely shared between Corsican and Sardinian samples, suggesting a major demographic expansion in Corsica and by complex reticulations connecting the most extreme outlier haplotypes from the middle eastern, Egyptian but also Tyrolean samples. On the other hand, the Multi-Dimensional Scaling (MDS) analysis of the Short Tandem Repeat (STR) haplotype length variation (Figure S3) shows that Sardinian (and Corsican) G2-L91 chromosomes are much closer to the middle eastern chromosomes than to the European ones. Although the available data do not provide any indications about the dispersal route of G2-L91 in Europe and Africa, nor concerning its arrival in Corsica and Sardinia, it is interesting to note that the diffusion and microsatellite dating (Table S5) of this lineage along the coasts of the Mediterranean overlap that of Cardial Ware pottery [50]. This suggests at least two different dispersal routes of G2-L91 from the Middle East, where the haplotype diversity is the highest (Table S5): one through the Balkans up over and beyond the Alps, which would explain the presence of ancient and modern Tyrolean G2-L91 Y chromosomes; the second, likely by sea, which would explain the dispersal along the coasts of North African and southern Italy and the Mediterranean islands.

Although the highest frequency of G2-L91 is observed in Corsica, the highest STR variation of the lineage is detected in the northern area of Sardinia (Table S5), suggesting that the spread was from Sardinia to Corsica and not vice versa, as previously proposed [22,28,48]. On the other hand, the high frequency of G2-L91 in southern Corsica (Table S3) is characterized by a low diversity (Table S5), which could be explained by the flourish of the Torrean civilization, a Nuragic culture that spread to southern Corsica from North Sardinia in the Bronze Age [11].

Haplogroup I2-M26, likely of south-western European origin [51], accounts for more than one third of Sardinian Y-chromosomes, while it is rare in most other modern European populations, including the neighboring Corsicans (Table S6). Frequencies above 5% are observed only in Basque groups.

The dissection of this haplogroup into its main subclades (I2-M26alpha-Z27361, I2-M26beta-Z27401, I2-L160 and its sub-lineages) previously identified in Sardinia [43] highlights that the majority of the I2-M26 chromosomes belong to I2-L160, both in Sardinia and outside the island (Table S6, [52]). Three Volterra samples, two I2-M26*(xM26alpha-Z27361, M26beta-Z27401) and one I2-M26alpha-Z27361, and a Basque subject, I2-M26beta-Z27401 [22], are the only exceptions.

Network analyses reveal a massive expansion of sub-clade I2-L160delta in Sardinia (Figure S4; Table S7) and a local expansion in the British Isles of branches I2-M26(xL160) (Figure S5; Table S8) that do not involve haplotypes observed in Sardinia and south-western Europe. In both networks, Sardinian samples share haplotypes with, or are directly connected to, French/Spanish/Italian subjects. Interestingly, one of the two 5 ky old males buried in a Neolithic French necropolis and classified as I-P37 [53] shares its STR haplotype with Y chromosomes I2-M26(xL160) of modern French, Irish and Norwegian samples and I2-L160 of modern Spaniards, Sardinians and continental Italians (Table S8). Although the lack of a sub-classification of this ancient specimen and the low frequency of haplogroup I2-M26 outside Sardinia do not allow any inference about the source area of the Sardinian I2-M26 Y chromosomes, MDS analysis of the haplotype STR length variation highlights a closer relation between Sardinian and French samples (Figure S6). Based on microsatellite variability, [51] south-western Europe (Iberian Peninsula/southern France) is a likely area of origin for this haplogroup and a starting point for the first colonization of Sardinia. However, the observation of chromosomes I2-M26 (xM26alpha, M26beta) and I2-M26alpha (I-Z27361) in Tuscany highlights a possible alternative route for the arrival of haplogroup I2-M26 in Sardinia: from Tuscany through the islands of Elba and Corsica.

The sub-classification of the I2-L160 chromosomes into the main subgroups [I2-L160gamma (Z27138), I2-L160delta A (Y20194), I2-L160delta B (Z26452), I2-L160delta C (Z26534), I2-L160delta D (PF4225), I2-L160delta E (PF4265), I2-L160delta F (PF4301), I2-L160delta G (Z26723), I2-L160delta H (PF4364), I2-L160delta I (Z26773) and I2-L160delta L (PF4421)] (Table S9, Figure S4) revealed that some of them have comparable frequencies in the three areas of the island, while others are more frequent or unique to only one area (I2-L160delta G observed only in the South). Although the high level of

resolution achieved in the sub-classification has strongly reduced the sample size of the identified sub-clades, preventing a statistically significant distribution being obtained of I2-L160 sub-lineages in the Sardinian isolates ($X_{26df} = 28.4$, p -value = 0.3), higher frequencies of I2-L160delta D (PF4225) in North Sardinia and I2-L160delta I (Z26773) in the central-eastern area of the island are apparent. It is of note that with the only exceptions of a I2-L160delta L (PF4421) Y chromosome in Calabria and a I2-L160delta A (Y20194) Y chromosome in Andalusia (easily explained as a recent Sardinian contribution), the I2-L160 sub-lines are virtually only observed in the island, supporting a Sardinian origin of I2-L160delta clades. This interpretation seems to be confirmed by the results obtained by the YFull Tree results [54].

Haplogroup R1b-V88 is a scarcely represented early branch of haplogroup R1b, mainly observed in sub-Saharan Africa. Its highest frequency is reported in Central Sahel (northern Cameroon, northern Nigeria, Chad and Sudan) where R1b-V88 sub-lineages underwent an expansion in Chadic-speaking groups [47,55,56]. Outside Africa, R1b-V88 lineages have been sporadically observed in the Middle East [57] and Europe, particularly in Sardinia [17,20,22,47,56,58]. Different hypotheses have been proposed concerning its ancestral homeland: a western Asian/middle eastern origin [47,59] and a sub-Saharan African origin [55] were hypothesized to explain its diffusion and variation in Africa, while no explanation was advanced for the presence of its sub-lineage R1b-M18 in Sardinia. The recent and detailed reconstruction of the phylogeny of this haplogroup [56] has revealed that the rare European R1b-V88 lineages (R1b-M18 and R1b-V35) originated from the root of the phylogeny much earlier (about 12.34 kya) than the separation of the African lineages (7.85 ± 0.90 kya), thus supporting an origin of R1b-V88 outside Africa and a subsequent diffusion in sub-Saharan Africa through the Last Green Sahara period during the Middle-Holocene [56]. Interestingly, recent studies on ancient DNA [60–62] identified the most ancient R1b-V88 samples (dated 11 and 9 ky) in East Europe (Serbia and Ukraine, respectively) and more recent R1b-V88 samples (dated 7 and 6 ky) in Spain (I0410) and Germany (I1593, I0559) thus supporting a European origin and opening new grounds for discussion concerning the routes towards Africa and Sardinia, where R1b-V88 characterizes a considerable number of ancient specimens [62,63].

2.4. Haplogroup Distribution in Pre-Historic Sardinian Samples

Ancient genomes of a number of Sardinian specimens, 44 of which males, derived from various caves located in different areas of the island and belonging to different archaeological phases, have been recently analyzed [62,63]. On the whole, nine Y-chromosome haplogroups have been identified, all of them observed in modern samples (Figure 1): E-L618 (derivative of E-M78, precursor of E-V13), G2-L166 (derivative of G-L91), G2-F872 (derivative of G2-L30, equivalent to G2-M547), I2-M223, I2-M26, I-M423, J1-L862 (derivative of J1-Page08), J2-M241 or J2-L283 (derivative of J2-M241), R1b-V88 and R1b-M269. In addition, chromosomes I2-M436(xM223) and R1b-L754(xM269) were also reported.

G2-L91, G2-F872 characterized the oldest specimens (from Middle to Late Neolithic), R1b-V88 and I2-M223 were observed in samples from the Early Copper Age and especially from the Early Bronze Age and the Nuragic period. I2-M26 appeared starting from the Early Bronze Age, J2-L283 from the Nuragic period and J1-L862 and R1b-M269 from the Punic period. Finally, haplogroups E-L618 and I2-M423 appeared only in Punic and Medieval specimens (Figure S7).

Haplogroup R1b-V88, observed in pre-Neolithic times in Balkan subjects [64], is the most represented haplogroup characterizing central-eastern and south-western Sardinian samples from the Early Copper Age (one subject in the South West), to the Nuragic period (four Early Bronze Age subjects all located in the central-eastern area; five subjects of Nuragic period-three in the central-eastern area and two in the south-western area). Haplogroup I2, which has been observed in pre-Neolithic times in Europe, is considered a hunter-gatherer signature. Together with haplogroup G2, it was common in Copper Age Iberia and appeared in Sardinia in Early Bronze Age as I2-M223, mainly in the North, and as I2-M26 in the Central-East area. Conversely, I2-M223 is observed in modern samples at low frequency (1.4%) only in the Central-East and South-West areas, while I2-M26 is

present at high frequency on the entire island, especially in the Central-East area (49.7%). Haplogroups G2-F872 (M547) and J2-M241 emerge in the ancient genetic Sardinian landscape only in the Nuragic period. Haplogroup G2-F872 (M547), which characterizes three ancient DNAs (aDNAs) [62,63], one from Central-East, and two from South-West Sardinia and 7.3% of modern samples (with higher frequency in the North (8.4%) and lower in the South-West (6.4%) and Central-East (4.9%)), was described in an Anatolian sample older than 7 ky [65]. G2-L166 (L91) was observed in four aDNAs from northern Sardinia and in 6.5% of present-day Sardinians at a higher frequency in the North (10.3%) than in the Central-East and South-West areas. As previously mentioned, this haplogroup characterizes the 5.3 ky old Tyrolean Iceman while its precursor (G2a2a1) was commonly found in Anatolia and eastern European Neolithic specimens as well as in Chalcolithic Iberians [60,61,66,67]. Haplogroup R1b is represented in ancient samples mainly by derivatives of R1b-V88. This haplogroup, which likely originated in eastern Europe, where the most ancient samples (dated 11–9 ky) have been reported [61,62], characterizes Sardinian samples older than 5 ky. Taking into account that the European branches of R1b-V88 are different and phylogenetically older than the African ones [56], it is likely that R1b-V88 chromosomes reached Sardinia through western Europe. Interestingly, ancient R1b chromosomes have been described in Italy (Villabruna, dated 14 ky, [68]) and Iberia (dated 7 ky, [27]). On the other hand, haplogroup R1b-M269, common in the Iberian Peninsula since 4.5 kya, where it almost completely replaced the pre-existing haplogroups I2, G2 and R1b(xM269) [67] and is frequent (21.3%) in modern samples from North Sardinia, was observed in ancient DNAs from Punic and Medieval sites. The observation in Punic sites (South-West area) of haplogroups J1-L862 and J2-L283, described within Levantine Bronze Age individuals [69] and very common in modern North African [70] and Balkan populations [71], respectively, may represent traces of migrations from Levant/North African following the conquest of the island by the growing Carthaginians. Accordingly, these haplogroups in modern samples show their highest frequencies (3.6% and 1.4%, respectively) in the south-western area. Finally, haplogroup E-L618, described in a 15,000-year-old modern human from eastern Morocco attributed to the Iberomaurusian culture [72], may testify a further link with North Africa, although modern samples belonging to the equivalent haplogroup E-M78*(xV13) have been described in Egypt [73] and in the Balkans [71].

3. Discussion

Genome-wide analyses of modern DNA place Sardinian samples in an outlier position in the European genomic landscape [2,5,25,74–76] strongly mirroring the outcome of genomic analyses of early European Neolithic farmers [26,27]. The peculiarity of modern Sardinians is confirmed by the analysis of Y-chromosome haplogroup frequencies. Genetic drift and long isolation can explain the presence of haplogroups that are very rare in other European populations such as R1b-V88 and G2-L91 as well as the increase of frequency of haplogroup I2-M26. Indeed, as highlighted by the Principal Component Analysis plot (PCA, Figure 3), Sardinian groups are located at the boundary of the European distribution towards North Africa with a relative closeness to Corsican samples as also detected by genome-wide based haplotypes [5]. While the proximity with Corsican groups, especially of southern Corsicans and northern Sardinians, is due to the high incidence of haplogroup G2-L91 in both populations, the closeness to North Africa, mainly due to the sharing of R1b-V88, is overestimated since Sardinian and African R1b-V88 Y chromosomes belong to different sub-lineages that phylogenetically diverged more than 7 kya [56].

Although archaeological data indicate that Sardinia has been inhabited since Paleolithic times, this early human presence might have been very limited, and the time of the first peopling is still a matter of debate. Modern and ancient mitogenome analyses [23] have shown that the majority of Sardinian-specific mtDNA haplogroups coalesce in post-Nuragic, Nuragic and Neolithic-Copper Age periods, although some rare maternal lineages (K1a2d and U5b1i1) might have been on the island already in pre-Neolithic times. The recent analysis of ancient Sardinians belonging to different archaeological periods revealed a high level of genetic continuity with the Nuragic period for the male

counterpart, also. Three Y-chromosome haplogroups (G2-L91, I2-M26, R1b-V88), which on the basis of ancient and modern samples had been previously proposed as founder lineages [22,28], were indeed all observed [62,63].

Although the demographic and genetic features associated with insularity offer the opportunity to better evaluate the impact of evolutionary forces such as founder events, gene flow and genetic drift that acted on the present Sardinian population, the homogenizing effect caused by internal movements during the last 150 years is a major confounding element. Here, through the “monophyletic surname sampling” approach, we obtained, at least for the Y-chromosome variation, novel clues concerning the past genetic isolates of the island, thus bypassing the confounding effect of recent migrations. With this approach we obtained a picture of the island population at the time of the introduction of the use of surnames, an event that in Italy occurred in the Middle Age. The comparison of the Y-chromosome haplogroups of ancient samples dated from the Early Neolithic [62] with those present in modern samples revealed a markedly different haplogroup distribution in the three large areas reflecting the ancient historic and geographic subdivision of the island (Figure S7). For some haplogroups the differences are likely the legacy of the ancient distribution showing local continuity albeit with varying frequencies. Thus, for example, G2-L91 shows its highest frequency (10.3%) in the North where it was also observed in aDNAs (four subjects), while it shows low frequencies in the Central-East (1.4%) and South-West (2.9%); I2-M26 characterizing two aDNAs, one from central-eastern and one from south-western areas, is found across the entire island at high frequency but especially in the Central-East. Differently, R1b-V88 is not observed in the South where it was observed in aDNAs; in contrast, I2-M223 was observed in different aDNAs in the North where it was not detected in modern samples.

On the whole, the Sardinian sample that we analyzed is characterized by a haplogroup heterogeneity of 0.8466 with high values in the northern (0.8814) and southern (0.8363) regions. The lowest heterogeneity in the central-eastern area (0.7469) reflects a reduction of the heterogeneity of all main haplogroups and can be explained by genetic drift due to long-term isolation in this region identified as the “archaic zone”. Indeed, it is reported that indigenous populations retreated to this mountainous region when Phoenicians and, later, Carthaginians colonized the southern part of the island [13]. The extraordinary low value of heterogeneity registered in the entire island for haplogroup I (HGI = 0.1898: 0.1766 in the North, 0.1264 in the Centre and 0.2857 in the South), which is almost completely represented by its sub-haplogroup I2-M26, which is rare elsewhere, can be explained only by a founder effect and genetic drift associated with the early peopling of Sardinia. In addition, taking into account that the three sub-clades stemming from the root of I2-M26 (I2-M26alpha (Z27361)-Table S6, I2-M26-beta-Z27401 and I2gamma-Z27138 [54]) are also present in other populations of Europe, haplogroup I-M26 must have been carried to Sardinia when it was already differentiated. Conversely, I2-M26delta, virtually only observed in Sardinia, likely originated and differentiated in situ on the island. Thus, the significantly higher frequency of some I2-M26delta (I2-L160 xZ27138) clades and the lack of chromosomes belonging to the rare clades I2-M26alpha (Z27361) and I2-M26beta (Z27401) in the “archaic zone” (Table S9) may reflect a second founder effect during the initial peopling of this area and/or genetic drift during the following long period of isolation (Figure 4).

For those haplogroups equally diversified in and outside Sardinia, it is difficult to discriminate between founder lineages and lineages that arrived later.

In brief, this study shows that the analysis of the Y-chromosome gene pool coupled with a sampling method based on the origin of the family name provides a greatly improved picture of the genetic structure of past population isolates. The application of this approach to a sample of Sardinian subjects carrying territorial monophyletic surnames (whose origin can be assigned to a specific location) allowed us to confirm the peculiarity of the Sardinian population in the European context and to detect ancient heterogeneities among the three main geographical areas of the island. We observed differences in the distribution not only of founding lineages and of lineages acquired through subsequent migrations, but also of sub-lineages that are Sardinian-specific. In particular,

the high homogeneity displayed by the central-eastern sample is in agreement with the long-lasting isolation and founder and genetic drift effects of the “archaic zone”. In addition, the comparison of the haplotype distribution in past isolates with that displayed by recent data on ancient DNA from different geographical areas and archaeological periods offered the unique opportunity to better understand the genetic history and demography of this peculiar insular population.

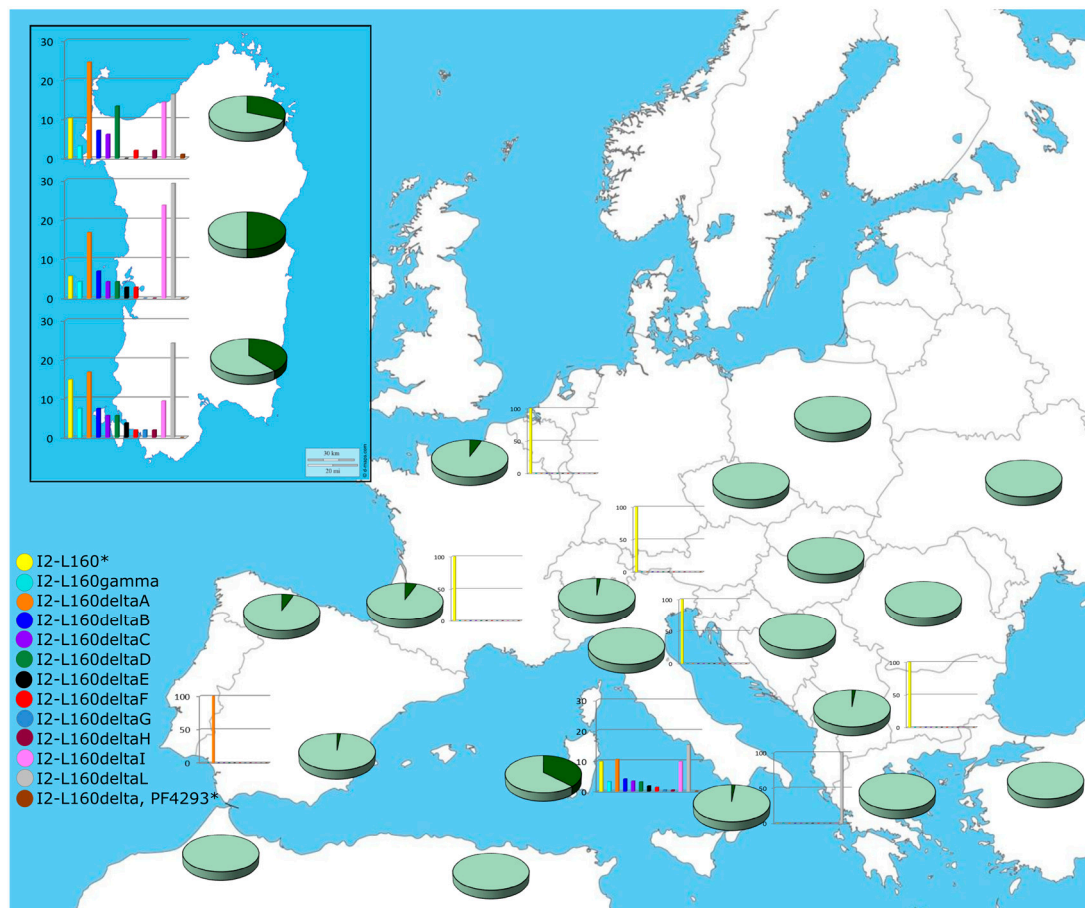


Figure 4. Geographical distribution of the Y-chromosome haplogroup I2-L160 and its sub-clades in Europe. Each pie represents a population and the dark green sector the percentage of I2-L160 (Tables S6 and S9); the histograms represent the percentages of the I2-L160 sub-clades according to the color codes shown on the figure. The inset shows the distribution of haplogroups and sub-haplogroups in the three geographical areas into which the Sardinian samples were subdivided according to the origin of their surnames.

4. Materials and Methods

4.1. The Sample

The sample consisted of 603 Sardinian males carrying Sardinian territorial monophyletic surnames [19] whose place of origin could be referenced to a specific Sardinian linguistic area (Figure S1) according to linguistic analyses [77,78].

These subjects were selected from a larger collection ($n > 1000$) of apparently unrelated healthy males, whose three-generation Sardinian origin was ascertained by interview, gathered thanks to the collaboration of many laboratories on the Sardinian territory, during different campaigns over more than 20 years. The selected samples were then assigned according to the place of origin of their surnames to the three large macro-areas of the island (North, Central-East and South-West Sardinia).

The “surname sampling approach”, previously applied in several studies [79–82], has been proved to be able to detect information earlier than the beginning of surnames history, which for Italy goes back to the Middle Age [19]. Indeed, in most societies, surnames are transmitted from father to child (just like Y chromosomes) and, being a linguistic attribution, reflect the identity and living area of the people. Thus, individuals bearing “monophyletic” surnames derived from words of the same local dialect likely share the same place of origin. In the case of Sardinia, the use of a sampling method based on the origin of monophyletic surnames, compared with the standard method based on the birthplace of the subjects, suggested a significant heterogeneity in the distribution of the Y-chromosome haplogroups [19].

4.2. DNA Analysis

Y-chromosome haplogroups were defined by hierarchical order analysis of 65 biallelic markers of the Male-Specific region of Y chromosome (MSY) (Figure 1), following the latest Y-chromosome phylogeny [54,83,84] and according to Battaglia et al. [71], Grugni et al. [3] and Karachanak et al. [85]. Seven markers, not included in previous papers, were analyzed as follows: L160 [86], M13 [87], M153 [44] and Z209 [86] using Restriction Fragment Length Polymorphism (RFLP) analysis, L2 [86] and V65 [73] using Denaturing High Performance Liquid Chromatography (DHPLC) and M282 [84] by sequencing.

All samples were also analyzed at ten STRs loci (DYS19, DYS388, DYS389I, DYS389B, DYS390, DYS391, DYS392, DYS393, YCAIIa and YCAIIb) using two multiplex reactions and a 3730 Applied Biosystems sequencer, as previously described [71]. Nomenclature details are available at the STRBase web site [88].

4.3. Statistical Analysis

Considering the results previously obtained [19], three main areas of Sardinia were considered (Figure S1): North (linguistic zones 1–5 and 23), Central-East (linguistic zones 6–9, 11, 14 and 15) and South-West (linguistic zones 10, 12, 13 and 16–22).

Haplogroup distributions were compared through the Chi Square Test of independence using the Xlstat add-on for Excel. Haplogroup heterogeneity (H) was computed using Nei’s [89] standard method. PC analysis was performed on Y-chromosome haplogroup frequencies disregarding those lower than 5% through `prcomp()` function on R (R Core Team 2017) and setting as TRUE the center and scale function.

MDS was performed using the Xlstat add-on for Excel and R_{ST} values [90] calculated on STRs haplotypes associated to haplogroups G2-L91 and I2-M26(xL160). Genetic structure was examined using analysis of molecular variance (AMOVA, [91]) using Arlequin software Ver 3.5 (Laurent Excoffier and Heidi Lischer, Bern, Switzerland).

Within specific haplogroups, median-joining (MJ) networks [92] were constructed using Network 4.6.0.0 [93], after processing the data with the reduced-median method [94] and weighting the STR loci proportionally to the inverse of the repeat variance. Time estimates were calculated only when more than five observations per population/region were available. Coalescent times were defined using the methodology of Zhivotovsky et al. [95] as modified according to Sengupta et al. [96] by using a microsatellite evolutionary effective mutation rate of 6.9×10^{-4} per generation (25 years).

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/22/5763/s1>. Figure S1. Boundaries of the three main areas of Sardinia that reflect the ancient history and geography of the island and their relative linguistic zones [77,78]. Table S1. Frequencies, observed and expected, of the Y-chromosome haplogroups in the North, Central-East and South-West areas of the island, together with significance per cell. Table S2. Frequencies, in percent, of Y-chromosome haplogroups and sub-haplogroups in the 36 populations included in the PCA (Figure 3). Table S3. Frequencies, in percent, of haplogroup G2-L91 and its sub-clades. Table S4. STR haplotypes associated to haplogroup G2-L91. Figure S2. Network of six-loci Y-STR haplotypes (DYS389I, DYS389B, DYS390, DYS391, DYS392 and DYS393) belonging to haplogroup G2-L91. Network analysis was performed on Sardinian samples in comparison with all the samples available in the literature analyzed for the same set of STRs (Table S4). Circles and colored sectors are proportional to the number

of subjects, with the smallest circle equal to one Y chromosome; connecting lines are proportional to the number of mutations. Figure S3. Two-dimensional plot obtained from MDS analysis based on pairwise R_{ST} values for six-loci Y-STR haplotypes (DYS389I, DYS389B, DYS390, DYS391, DYS392 and DYS393) associated to haplogroup G2-L91 observed in Sardinians and other populations of the Mediterranean Basin. Table S5. Ages of microsatellite variation and mean variance of microsatellite loci within haplogroup G2-L91. Table S6. Frequencies, in percent, of haplogroup I2-M26 and its main sub-clades. Figure S4. Network of Y-STR haplotypes belonging to haplogroup I2-L160. Network analysis was performed on Sardinian samples in comparison with all the samples available in the literature for this haplogroup (Table S7). Circles and colored sectors are proportional to the number of subjects, with the smallest circle equal to one Y chromosome; connecting lines are proportional to the number of mutations. Table S7. STR haplotypes associated to haplogroup I2-L160. Figure S5. Network of Y-STR haplotypes belonging to I2-M26(xL160) Y chromosomes. Network analysis was performed on Sardinian samples in comparison with all the I2-M26(xL160) samples available in the literature (Table S8). Circles and colored sectors are proportional to the number of subjects, with the smallest circle equal to one Y chromosome; connecting lines are proportional to the number of mutations. Table S8. STR haplotypes associated to haplogroup I2-M26(xL160). Figure S6. Two-dimensional plot obtained from MDS analysis based on pairwise R_{ST} values for seven-loci Y-STR haplotypes associated to haplogroup I2-M26 observed in Sardinians and other populations of the Mediterranean Basin. Table S9. Frequencies of haplogroup I2-M26 and its I2-L160 sub-clades. Figure S7. Y-chromosome haplogroup distribution per macro-area of ancient Sardinian samples. Data from Marcus et al. [62] and Fernandes et al. [63]. For each bar plot: the name of haplogroups on the left side, the temporal range of the ancient samples, on the right side.

Author Contributions: conceptualization, A.T., O.S.; data curation, V.G., A.R., G.C., L.F., A.O. and A.A.; formal analysis, F.C.; funding acquisition, L.F., A.O., A.P., A.T. and O.S.; investigation, V.G., A.R., G.C., C.N., F.C., L.O., V.B., D.S. and N.A.-Z.; methodology, O.F. and A.L.; project administration, A.T. and O.S.; resources, A.O., P.F., A.A., A.T. and O.S.; writing—original draft, V.G., A.R. and O.S.; writing—review & editing, V.G., L.F., A.T. and O.S.

Funding: This research was supported by Compagnia di San Paolo (to A.T. and O.S.) and received support from the University of Pavia strategic theme “Towards a governance model for international migration: an interdisciplinary and diachronic perspective” (MIGRAT-IN-G) (to L.F., A.A., A.O., O.S., A.T.) and from the Italian Ministry of Education, University and Research (MIUR): Dipartimenti di Eccellenza Program (2018–2022), Dept. of Biology and Biotechnology “L. Spallanzani”, University of Pavia (to L.F., A.A., A.O., O.S., A.T.).

Acknowledgments: We are grateful to all the donors for providing the blood samples and to all the people and institutions that contributed to their collection. These include the “Centri trasfusionali” of Cagliari (c/o Brotzu), Alghero, Oristano, Ozieri, Olbia, Lanusei, Iglesias and Sassari. We also thank the editor and two anonymous reviewers for their useful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

BCE	Before common era
CE	Common era
DHPLC	Denaturing high performance liquid chromatography
ky	Kilo years
kya	Kilo years ago
MDS	Multi-dimensional scaling analysis
MSY	Male specific region of the Y chromosome
PCA	Principal component analysis
RFLP	Restriction fragment length polymorphism
STR	Short tandem repeat

References

1. Cavalli-Sforza, L.L.; Piazza, A. Human genomic diversity in Europe: A summary of recent research and prospects for the future. *Eur. J. Hum. Genet.* **1993**, *1*, 3–18. [[CrossRef](#)]
2. Di Gaetano, C.; Voglino, F.; Guarrera, S.; Fiorito, G.; Rosa, F.; Di Blasio, A.M.; Manzini, P.; Dianzani, I.; Betti, M.; Cusi, D. An overview of the genetic structure within the Italian population from genome-wide data. *PLoS ONE* **2012**, *7*, e43759. [[CrossRef](#)]
3. Grugni, V.; Battaglia, V.; Hooshiar Kashani, B.; Parolo, S.; Al-Zahery, N.; Achilli, A.; Olivieri, A.; Gandini, F.; Houshmand, M.; Sanati, M.H. Ancient migratory events in the Middle East: New clues from the Y-chromosome variation of modern Iranians. *PLoS ONE* **2012**, *7*, e41252. [[CrossRef](#)]

4. Pardo, L.M.; Piras, G.; Asproni, R.; Van Der Gaag, K.J.; Gabbas, A.; Ruiz-Linares, A.; De Knijff, P.; Monne, M.; Rizzu, P.; Heutink, P. Dissecting the genetic make-up of North-East Sardinia using a large set of haploid and autosomal markers. *Eur. J. Hum. Genet.* **2012**, *20*, 956–964. [[CrossRef](#)]
5. Raveane, A.; Aneli, S.; Montinaro, F.; Athanasiadis, G.; Barlera, S.; Birolo, G.; Boncoraglio, G.; Di Blasio, A.M.; Di Gaetano, C.; Pagani, L. Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe. *Sci. Adv.* **2019**, *5*, eaaw3492. [[CrossRef](#)]
6. Wagner, M.L. *Historische Lautlehre des Sardischen*, 1st ed.; Halle: Niemeyer, Germany, 1941; pp. 1–344.
7. Spoor, C.; Sondaar, P. Human fossils from the endemic island fauna of Sardinia. *J. Hum. Evol.* **1986**, *15*, 399–408. [[CrossRef](#)]
8. Contini, M.; Cappello, N.; Griffo, R.; Rendine, S.; Piazza, A. Géolinguistique et géogénétique: Une démarche interdisciplinaire. *Géolinguistique* **1989**, *4*, 129–197.
9. Hofmeijer, G.K.; Alderliesten, C.; Van Der Borg, K.; Houston, C.; De Jong, A.; Martini, F.; Sanges, M.; Sondaar, P.; De Visser, J. Dating of the upper Pleistocene lithic industry of Sardinia. *Radiocarbon* **1989**, *31*, 986–991. [[CrossRef](#)]
10. Sondaar, P.; Elhurg, R.; Holmeijer, G.K.; Martini, F.; Sanges, M.; Spaan, A.; de Visser, H. The human colonization of Sardinia: A Late-Pleistocene human fossil from Corbeddu cave. *CR Acad. Sci. Paris* **1995**, *320*, 145–150.
11. Dyson, S.L.; Rowland, R.J., Jr. *Archaeology and History in Sardinia from the Stone Age to the Middle Ages: Shepherds, Sailors, and Conquerors*, 1st ed.; University of Pennsylvania Press: Pennsylvania, NJ, USA, 2007; pp. 1–248.
12. Broodbank, C. “Minding the gap”: Thinking about change in Early Cycladic island societies from a comparative perspective. *Am. J. Archaeol.* **2013**, *117*, 535–543. [[CrossRef](#)]
13. Barreca, F. *La Sardegna Fenicia e Punica*, 1st ed.; Sassari: Chiarella Editore, Italy, 1979; pp. 1–303.
14. Cavalli-Sforza, L.L.; Menozzi, P.; Piazza, A. *The History and Geography of Human Genes*; Princeton University Press: Pennsylvania, NJ, USA, 1994; pp. 273–275.
15. Chiang, C.W.K.; Marcus, J.H.; Sidore, C.; Biddanda, A.; Al-Asadi, H.; Zoledziewska, M.; Pitzalis, M.; Busonero, F.; Maschio, A.; Pistis, G. Genomic history of the Sardinian population. *Nat. Genet.* **2018**, *50*, 1426–1434. [[CrossRef](#)] [[PubMed](#)]
16. Calò, C.; Melis, A.; Vona, G.; Piras, I. Review synthetic article: Sardinian population (Italy): A genetic review. *AJOL* **2008**, *1*, 39–64. [[CrossRef](#)]
17. Morelli, L.; Grosso, M.; Vona, G.; Varesi, L.; Torroni, A.; Francalacci, P. Frequency distribution of mitochondrial DNA haplogroups in Corsica and Sardinia. *Hum. Biol.* **2000**, *72*, 585–595. [[PubMed](#)]
18. Semino, O.; Passarino, G.; Oefner, P.J.; Lin, A.A.; Arbuzova, S.; Beckman, L.E.; De Benedictis, G.; Francalacci, P.; Kouvatsi, A.; Limborska, S. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: A Y chromosome perspective. *Science* **2000**, *290*, 1155–1159. [[CrossRef](#)] [[PubMed](#)]
19. Zei, G.; Lisa, A.; Fiorani, O.; Magri, C.; Quintana-Murci, L.; Semino, O.; Santachiara-Benerecetti, A.S. From surnames to the history of Y chromosomes: The Sardinian population as a paradigm. *Eur. J. Hum. Genet.* **2003**, *11*, 802–807. [[CrossRef](#)] [[PubMed](#)]
20. Contu, D.; Morelli, L.; Santoni, F.; Foster, J.W.; Francalacci, P.; Cucca, F. Y-chromosome based evidence for pre-neolithic origin of the genetically homogeneous but diverse Sardinian population: Inference for association scans. *PLoS ONE* **2008**, *3*, e1430. [[CrossRef](#)]
21. Pala, M.; Achilli, A.; Olivieri, A.; Hooshiar Kashani, B.; Perego, U.A.; Sanna, D.; Metspalu, E.; Tambets, K.; Tamm, E.; Accetturo, M. Mitochondrial haplogroup U5b3: A distant echo of the Epipaleolithic in Italy and the legacy of the early Sardinians. *Am. J. Hum. Genet.* **2009**, *84*, 814–821. [[CrossRef](#)]
22. Francalacci, P.; Morelli, L.; Angius, A.; Berutti, R.; Reinier, F.; Atzeni, R.; Pilu, R.; Busonero, F.; Maschio, A.; Zara, I. Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. *Science* **2013**, *341*, 565–569. [[CrossRef](#)]
23. Olivieri, A.; Sidore, C.; Achilli, A.; Angius, A.; Posth, C.; Furtwängler, A.; Brandini, S.; Capodiferro, M.R.; Gandini, F.; Zoledziewska, M. Mitogenome diversity in Sardinians: A genetic window onto an island’s past. *Mol. Biol. Evol.* **2017**, *34*, 1230–1239. [[CrossRef](#)]
24. Tenesa, A.; Wright, A.; Knott, S.; Carothers, A.; Hayward, C.; Angius, A.; Persico, I.; Maestrale, G.; Hastie, N.; Pirastu, M. Extent of linkage disequilibrium in a Sardinian sub-isolate: Sampling and methodological considerations. *Hum. Mol. Genet.* **2004**, *13*, 25–33. [[CrossRef](#)]

25. Tamm, E.; Cristofaro, J.D.; Mazières, S.; Pennarun, E.; Kushniarevich, A.; Raveane, A.; Semino, O.; Chiaroni, J.; Pereira, L.; Metspalu, M. Genome-wide analysis of Corsican population reveals a close affinity with Northern and Central Italy. *Sci. Rep.* **2019**, *9*, 13581. [[CrossRef](#)] [[PubMed](#)]
26. Lazaridis, I.; Patterson, N.; Mittnik, A.; Renaud, G.; Mallick, S.; Kirsanow, K.; Sudmant, P.H.; Schraiber, J.G.; Castellano, S.; Lipson, M. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **2014**, *513*, 409–413. [[CrossRef](#)] [[PubMed](#)]
27. Haak, W.; Lazaridis, I.; Patterson, N.; Rohland, N.; Mallick, S.; Llamas, B.; Brandt, G.; Nordenfelt, S.; Harney, E.; Stewardson, K. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **2015**, *522*, 207–211. [[CrossRef](#)] [[PubMed](#)]
28. Keller, A.; Graefen, A.; Ball, M.; Matzas, M.; Boisguerin, V.; Maixner, F.; Leidinger, P.; Backes, C.; Khairat, R.; Forster, M. New insights into the Tyrolean Iceman’s origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* **2012**, *3*, 698. [[CrossRef](#)] [[PubMed](#)]
29. Sikora, M.; Carpenter, M.L.; Moreno-Estrada, A.; Henn, B.M.; Underhill, P.A.; Sánchez-Quinto, F.; Zara, I.; Pitzalis, M.; Sidore, C.; Busonero, F. Population genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe. *PLoS Genet.* **2014**, *10*, e1004353. [[CrossRef](#)] [[PubMed](#)]
30. Lazaridis, I.; Nadel, D.; Rollefson, G.; Merrett, D.C.; Rohland, N.; Mallick, S.; Fernandes, D.; Novak, M.; Gamarra, B.; Sirak, K. Genomic insights into the origin of farming in the ancient Near East. *Nature* **2016**, *536*, 419–424. [[CrossRef](#)]
31. Workman, P.; Lucarelli, P.; Agostino, R.; Scarabino, R.; Scacchi, R.; Carapella, E.; Palmarino, R.; Bottini, E. Genetic differentiation among Sardinian villages. *Am. J. Phys. Anthropol.* **1975**, *43*, 165–176. [[CrossRef](#)]
32. Wagner, M.L. La lingua sarda. Storia spirito e forma. In *Max Leopold Wagner (Book Review)*; Brepols Publishers: Chicago, IL, USA, 1950; pp. 1–438.
33. Barbujani, G.; Sokal, R.R. Genetic population structure of Italy. I. Geographic patterns of gene frequencies. *Hum. Biol.* **1991**, *63*, 253–272.
34. Cappello, N.; Rendine, S.; Griffo, R.; Mameli, G.; Succa, V.; Vona, G.; Piazza, A. Genetic analysis of Sardinia: I. data on 12 polymorphisms in 21 linguistic domains. *Ann. Hum. Genet.* **1996**, *60*, 125–141. [[CrossRef](#)]
35. Zavattari, P.; Deidda, E.; Whalen, M.; Lampis, R.; Mulargia, A.; Loddo, M.; Eaves, I.; Mastio, G.; Todd, J.A.; Cucca, F. Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: Demography, chromosome recombination frequency and selection. *Hum. Mol. Genet.* **2000**, *9*, 2947–2957. [[CrossRef](#)]
36. Angius, A.; Melis, P.M.; Morelli, L.; Petretto, E.; Casu, G.; Maestrale, G.; Fraumene, C.; Bebbere, D.; Forabosco, P.; Pirastu, M. Archival, demographic and genetic studies define a Sardinian sub-isolate as a suitable model for mapping complex traits. *Hum. Genet.* **2001**, *109*, 198–209. [[CrossRef](#)] [[PubMed](#)]
37. Scozzari, R.; Cruciani, F.; Pangrazio, A.; Santolamazza, P.; Vona, G.; Moral, P.; Latini, V.; Varesi, L.; Memmi, M.M.; Romano, V. Human Y-chromosome variation in the western Mediterranean area: Implications for the peopling of the region. *Hum. Immunol.* **2001**, *62*, 871–884. [[CrossRef](#)]
38. Fraumene, C.; Petretto, E.; Angius, A.; Pirastu, M. Striking differentiation of sub-populations within a genetically homogeneous isolate (Ogliastra) in Sardinia as revealed by mtDNA analysis. *Hum. Genet.* **2003**, *114*, 1–10. [[CrossRef](#)] [[PubMed](#)]
39. Destro-Bisol, G.; Anagnostou, P.; Batini, C.; Battaglia, C.; Bertoncini, S.; Boattini, A.; Caciagli, L.; Caló, M.; Capelli, C.; Capocasa, M. Italian isolates today: Geographic and linguistic factors shaping human biodiversity. *J. Anthropol. Sci. JASS* **2008**, *86*, 179–188. [[PubMed](#)]
40. Pistis, G.; Piras, I.; Pirastu, N.; Persico, I.; Sassu, A.; Picciau, A.; Prodi, D.; Fraumene, C.; Mocci, E.; Manias, M.T. High differentiation among eight villages in a secluded area of Sardinia revealed by genome-wide high density SNPs analysis. *PLoS ONE* **2009**, *4*, e4654. [[CrossRef](#)]
41. Montinaro, F.; Boschi, I.; Trombetta, F.; Merigioli, S.; Anagnostou, P.; Battaglia, C.; Capocasa, M.; Crivellaro, F.; Destro-Bisol, G.; Coia, V. Using forensic microsatellites to decipher the genetic structure of linguistic and geographic isolates: A survey in the eastern Italian Alps. *Forensic Sci. Int. Gen.* **2012**, *6*, 827–833. [[CrossRef](#)]
42. Capocasa, M.; Anagnostou, P.; Bachis, V.; Battaglia, C.; Bertoncini, S.; Biondi, G.; Boattini, A.; Boschi, I.; Brisighelli, F.; Calò, C.M. Linguistic, geographic and genetic isolation: A collaborative study of Italian populations. *J. Anthropol. Sci. JASS* **2014**, *92*, 201–231.

43. Francalacci, P.; Sanna, D.; Useli, A.; Berutti, R.; Barbato, M.; Whalen, M.B.; Angius, A.; Sidore, C.; Alonso, S.; Tofanelli, S. Detection of phylogenetically informative polymorphisms in the entire euchromatic portion of human Y chromosome from a Sardinian sample. *BMC Res. Notes* **2015**, *8*, 174. [[CrossRef](#)]
44. Underhill, P.A.; Shen, P.; Lin, A.A.; Jin, L.; Passarino, G.; Yang, W.H.; Kauffman, E.; Bonn -Tamir, B.; Bertranpetit, J.; Francalacci, P. Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **2000**, *26*, 358–361. [[CrossRef](#)]
45. Passarino, G.; Underhill, P.A.; Cavalli-Sforza, L.L.; Semino, O.; Pes, G.M.; Carru, C.; Ferrucci, L.; Bonaf , M.; Franceschi, C.; Deiana, L. Y chromosome binary markers to study the high prevalence of males in Sardinian centenarians and the genetic structure of the Sardinian population. *Hum. Hered.* **2001**, *52*, 136–139. [[CrossRef](#)]
46. Francalacci, P.; Morelli, L.; Underhill, P.A.; Lillie, A.S.; Passarino, G.; Useli, A.; Madeddu, R.; Paoli, G.; Tofanelli, S.; Cal , C.M. Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. *Am. J. Phys. Anthropol.* **2003**, *121*, 270–279. [[CrossRef](#)] [[PubMed](#)]
47. Cruciani, F.; Trombetta, B.; Sellitto, D.; Massaia, A.; Destro-Bisol, G.; Watson, E.; Colomb, E.B.; Dugoujon, J.-M.; Moral, P.; Scozzari, R. Human Y chromosome haplogroup R-V88: A paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages. *Eur. J. Hum. Genet.* **2010**, *18*, 800–807. [[CrossRef](#)] [[PubMed](#)]
48. Rootsi, S.; Myres, N.M.; Lin, A.A.; J rve, M.; King, R.J.; Kutuev, I.; Cabrera, V.M.; Khusnutdinova, E.K.; Varendi, K.; Sahakyan, H. Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. *Eur. J. Hum. Genet.* **2012**, *20*, 1275–1282. [[CrossRef](#)] [[PubMed](#)]
49. Berger, B.; Niederst tter, H.; Erhart, D.; Gassner, C.; Schennach, H.; Parson, W. High resolution mapping of Y haplogroup G in Tyrol (Austria). *Forensic Sci. Int. Genet.* **2013**, *7*, 529–536. [[CrossRef](#)]
50. Price, T.D. *Europe’s First Farmers*; Cambridge University Press: Cambridge, UK, 2000; pp. 1–412.
51. Rootsi, S.; Kivisild, T.; Benuzzi, G.; Help, H.; Bermisheva, M.; Kutuev, I.; Bara , L.; Periči , M.; Balanovsky, O.; Pshenichnov, A. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am. J. Hum. Genet.* **2004**, *75*, 128–137. [[CrossRef](#)]
52. YFull Tree, I-L158. Available online: <https://www.yfull.com/tree/I-L158/> (accessed on 16 October 2019).
53. Lacan, M.; Keyser, C.; Ricaut, F.-X.; Brucato, N.; Duranthon, F.; Guilaine, J.; Crub zy, E.; Ludes, B. Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 9788–9791. [[CrossRef](#)]
54. YFull Tree. Available online: <https://www.yfull.com/tree/> (accessed on 16 October 2019).
55. Gonz lez, M.; Gomes, V.; L pez-Parra, A.M.; Amorim, A.; Carracedo, A.; S nchez-Diz, P.; Arroyo-Pardo, E.; Gusm o, L. The genetic landscape of Equatorial Guinea and the origin and migration routes of the Y chromosome haplogroup R-V88. *Eur. J. Hum. Genet.* **2013**, *21*, 324–331. [[CrossRef](#)]
56. D’Atanasio, E.; Trombetta, B.; Bonito, M.; Finocchio, A.; Di Vito, G.; Seghizzi, M.; Romano, R.; Russo, G.; Paganotti, G.M.; Watson, E. The peopling of the last Green Sahara revealed by high-coverage resequencing of trans-Saharan patrilineages. *Genome Biol.* **2018**, *19*, 20. [[CrossRef](#)]
57. Zalloua, P.A.; Platt, D.E.; El Sibai, M.; Khalife, J.; Makhoul, N.; Haber, M.; Xue, Y.; Izaabel, H.; Bosch, E.; Adams, S.M. Identifying genetic traces of historical expansions: Phoenician footprints in the Mediterranean. *Am. J. Hum. Genet.* **2008**, *83*, 633–642. [[CrossRef](#)]
58. Kivisild, T. The study of human Y chromosome variation through ancient DNA. *Hum. Genet.* **2017**, *136*, 529–546. [[CrossRef](#)]
59. Haber, M.; Mezzavilla, M.; Xue, Y.; Comas, D.; Gasparini, P.; Zalloua, P.; Tyler-Smith, C. Genetic evidence for an origin of the Armenians from Bronze Age mixing of multiple populations. *Eur. J. Hum. Genet.* **2016**, *24*, 931–936. [[CrossRef](#)] [[PubMed](#)]
60. Lipson, M.; Cheronet, O.; Mallick, S.; Rohland, N.; Oxenham, M.; Pietrusewsky, M.; Pryce, T.O.; Willis, A.; Matsumura, H.; Buckley, H. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* **2018**, *361*, 92–95. [[CrossRef](#)] [[PubMed](#)]
61. Mathieson, I.; Alpaslan-Roodenberg, S.; Posth, C.; Sz cs nyi-Nagy, A.; Rohland, N.; Mallick, S.; Olalde, I.; Broomandkoshbacht, N.; Candilio, F.; Cheronet, O. The genomic history of southeastern Europe. *Nature* **2018**, *555*, 197. [[CrossRef](#)] [[PubMed](#)]
62. Marcus, J.H.; Posth, C.; Ringbauer, H.; Lai, L.; Skeates, R.; Sidore, C.; Beckett, J.; Furtw ngler, A.; Olivieri, A.; Chiang, C.W.K. Population history from the Neolithic to present on the Mediterranean island of Sardinia: An ancient DNA perspective. *BioRxiv* **2019**, 583104. [[CrossRef](#)]

63. Fernandes, D.M.; Mitnik, A.; Olalde, I.; Lazaridis, I.; Cheronet, O.; Rohland, N.; Mallick, S.; Bernardos, R.; Broomandkoshbacht, N.; Carlsson, J. The arrival of Steppe and Iranian related ancestry in the islands of the Western Mediterranean. *BioRxiv* **2019**, 584714. [[CrossRef](#)]
64. Mathieson, I.; Reich, D. Differences in the rare variant spectrum among human populations. *PLoS Genet.* **2017**, *13*, e1006581. [[CrossRef](#)]
65. Kılınç, G.M.; Omrak, A.; Özer, F.; Günther, T.; Büyükkarakaya, A.M.; Bıçakçı, E.; Baird, D.; Dönertaş, H.M.; Ghalichi, A.; Yaka, R. The demographic development of the first farmers in Anatolia. *Curr. Biol.* **2016**, *26*, 2659–2666. [[CrossRef](#)]
66. Mathieson, I.; Lazaridis, I.; Rohland, N.; Mallick, S.; Patterson, N.; Roodenberg, S.A.; Harney, E.; Stewardson, K.; Fernandes, D.; Novak, M. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **2015**, *528*, 499–503. [[CrossRef](#)]
67. Olalde, I.; Brace, S.; Allentoft, M.E.; Armit, I.; Kristiansen, K.; Booth, T.; Rohland, N.; Mallick, S.; Szécsényi-Nagy, A.; Mitnik, A. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **2018**, *555*, 190–196. [[CrossRef](#)]
68. Fu, Q.; Posth, C.; Hajdinjak, M.; Petr, M.; Mallick, S.; Fernandes, D.; Furtwängler, A.; Haak, W.; Meyer, M.; Mitnik, A. The genetic history of ice age Europe. *Nature* **2016**, *534*, 200–205. [[CrossRef](#)]
69. Haber, M.; Doumet-Serhal, C.; Scheib, C.; Xue, Y.; Danecek, P.; Mezzavilla, M.; Youhanna, S.; Martiniano, R.; Prado-Martinez, J.; Szpak, M. Continuity and admixture in the last five millennia of Levantine history from Ancient Canaanite and present-day Lebanese genome sequences. *Am. J. Hum. Genet.* **2017**, *101*, 274–282. [[CrossRef](#)] [[PubMed](#)]
70. Semino, O.; Magri, C.; Benuzzi, G.; Lin, A.A.; Al-Zahery, N.; Battaglia, V.; Maccioni, L.; Triantaphyllidis, C.; Shen, P.; Oefner, P.J. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: Inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am. J. Hum. Genet.* **2004**, *74*, 1023–1034. [[CrossRef](#)] [[PubMed](#)]
71. Battaglia, V.; Fornarino, S.; Al-Zahery, N.; Olivieri, A.; Pala, M.; Myres, N.M.; King, R.J.; Rootsi, S.; Marjanovic, D.; Primorac, D. Y-chromosomal evidence of the cultural diffusion of agriculture in Southeast Europe. *Eur. J. Hum. Genet.* **2009**, *17*, 820–830. [[CrossRef](#)] [[PubMed](#)]
72. Van de Loosdrecht, M.; Bouzouggar, A.; Humphrey, L.; Posth, C.; Barton, N.; Aximu-Petri, A.; Nickel, B.; Nagel, S.; Talbi, E.H.; El Hajraoui, M.A. Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. *Science* **2018**, *360*, 548–552. [[CrossRef](#)] [[PubMed](#)]
73. Cruciani, F.; La Fratta, R.; Trombetta, B.; Santolamazza, P.; Sellitto, D.; Colomb, E.B.; Dugoujon, J.-M.; Crivellaro, F.; Benincasa, T.; Pascone, R. Tracing past human male movements in northern/eastern Africa and western Eurasia: New clues from Y-chromosomal haplogroups E-M78 and J-M12. *Mol. Biol. Evol.* **2007**, *24*, 1300–1311. [[CrossRef](#)] [[PubMed](#)]
74. Fiorito, G.; Di Gaetano, C.; Guarrera, S.; Rosa, F.; Feldman, M.W.; Piazza, A.; Matullo, G. The Italian genome reflects the history of Europe and the Mediterranean basin. *Eur. J. Hum. Genet.* **2016**, *24*, 1056–1062. [[CrossRef](#)] [[PubMed](#)]
75. Parolo, S.; Lisa, A.; Gentilini, D.; Di Blasio, A.M.; Barlera, S.; Nicolis, E.B.; Boncoraglio, G.B.; Parati, E.A.; Bione, S. Characterization of the biological processes shaping the genetic structure of the Italian population. *BMC Genet.* **2015**, *16*, 132. [[CrossRef](#)]
76. Sazzini, M.; Ruscone, G.A.G.; Giuliani, C.; Sarno, S.; Quagliarello, A.; De Fanti, S.; Boattini, A.; Gentilini, D.; Fiorito, G.; Catanoso, M. Complex interplay between neutral and adaptive evolution shaped differential genomic background and disease susceptibility along the Italian peninsula. *Sci. Rep.* **2016**, *6*, 32513. [[CrossRef](#)]
77. Contini, M. Classification phonologique des langues sardes. *Bull. Inst. Phonétique Grenoble* **1979**, *8*, 57–96.
78. Zei, G.; Piazza, A.; Moroni, A.; Cavalli-Sforza, L.L. Surnames in Sardinia| III. The spatial distribution of surnames for testing neutrality of genes. *Ann. Hum. Genet.* **1986**, *50*, 169–180. [[CrossRef](#)]
79. Skorecki, K.; Selig, S.; Blazer, S.; Bradman, R.; Bradman, N.; Waburton, P.; Ismajłowicz, M.; Hammer, M.F. Y chromosomes of Jewish priests. *Nature* **1997**, *385*, 32. [[CrossRef](#)] [[PubMed](#)]
80. Thomas, M.G.; Skorecki, K.; Ben-Amid, H.; Parfitt, T.; Bradman, N.; Goldstein, D.B. Origins of old testament priests. *Nature* **1998**, *394*, 138–140. [[CrossRef](#)] [[PubMed](#)]
81. Sykes, B.; Irven, C. Surnames and the Y chromosome. *Am. J. Hum. Genet.* **2000**, *66*, 1417–1419. [[CrossRef](#)] [[PubMed](#)]

82. King, T.E.; Jobling, M.A. Founders, drift, and infidelity: The relationship between Y chromosome diversity and patrilineal surnames. *Mol. Biol. Evol.* **2009**, *26*, 1093–1102. [CrossRef] [PubMed]
83. Van Oven, M.; Van Geystelen, A.; Kayser, M.; Decorte, R.; Larmuseau, M.H. Seeing the wood for the trees: A minimal reference phylogeny for the human Y chromosome. *Hum. Mut.* **2014**, *35*, 187–191. [CrossRef] [PubMed]
84. ISOGG, International Society of Genetic Genealogy. ISOGG Version 14.177. 2019. Available online: <https://isogg.org/tree/> (accessed on 16 October 2019).
85. Karachanak, S.; Grugni, V.; Fornarino, S.; Nesheva, D.; Al-Zahery, N.; Battaglia, V.; Carossa, V.; Yordanov, Y.; Torroni, A.; Galabov, A.S. Y-chromosome diversity in modern Bulgarians: New clues about their ancestry. *PLoS ONE* **2013**, *8*, e56779. [CrossRef] [PubMed]
86. Family Tree DNA, Y Haplotree. Available online: <https://www.familytreedna.com/public/y-dna-haplotree/A> (accessed on 16 October 2019).
87. Underhill, P.A.; Jin, L.; Lin, A.A.; Mehdi, S.Q.; Jenkins, T.; Vollrath, D.; Davis, R.W.; Cavalli-Sforza, L.L.; Oefner, P.J. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **1997**, *7*, 996–1005. [CrossRef]
88. STRBase. Available online: <https://strbase.nist.gov/index.htm> (accessed on 16 October 2019).
89. Nei, M. Phylogenetic analysis in molecular evolutionary genetics. *Annu. Rev. Genet.* **1996**, *30*, 371–403. [CrossRef]
90. Slatkin, M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **1995**, *139*, 457–462.
91. Excoffier, L.; Smouse, P.E.; Quattro, J.M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **1992**, *131*, 479–491.
92. Bandelt, H.-J.; Forster, P.; Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **1999**, *16*, 37–48. [CrossRef]
93. Fluxus Engineering. Available online: <http://www.fluxus-engineering.com/sharenet.htm> (accessed on 16 October 2019).
94. Bandelt, H.-J.; Forster, P.; Sykes, B.C.; Richards, M.B. Mitochondrial portraits of human populations using median networks. *Genetics* **1995**, *141*, 743–753.
95. Zhivotovsky, L.A.; Underhill, P.A.; Cinnioğlu, C.; Kayser, M.; Morar, B.; Kivisild, T.; Scozzari, R.; Cruciani, F.; Destro-Bisol, G.; Spedini, G. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* **2004**, *74*, 50–61. [CrossRef]
96. Sengupta, S.; Zhivotovsky, L.A.; King, R.; Mehdi, S.Q.; Edmonds, C.A.; Chow, C.E.; Lin, A.A.; Mitra, M.; Sil, S.K.; Ramesh, A. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am. J. Hum. Genet.* **2006**, *78*, 202–221. [CrossRef]

