



**HAL**  
open science

## **Prehistoric migrations through the Mediterranean basin shaped Corsican Y-chromosome diversity**

Julie Di Cristofaro, Stéphane Mazières, Audrey Tous, Cornelia Di Gaetano, Alice Lin, Paul Nebbia, Alberto Piazza, Roy King, Peter Underhill, Jacques Chiaroni

### ► **To cite this version:**

Julie Di Cristofaro, Stéphane Mazières, Audrey Tous, Cornelia Di Gaetano, Alice Lin, et al.. Prehistoric migrations through the Mediterranean basin shaped Corsican Y-chromosome diversity. PLoS ONE, 2018, 13 (8), pp.e0200641. <10.1371/journal.pone.0200641>. <hal-01863090>

**HAL Id: hal-01863090**

**<https://hal.science/hal-01863090v1>**

Submitted on 17 Sep 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

RESEARCH ARTICLE

# Prehistoric migrations through the Mediterranean basin shaped Corsican Y-chromosome diversity

Julie Di Cristofaro<sup>1,2</sup>, Stéphane Mazières<sup>2</sup>, Audrey Tous<sup>1</sup>, Cornelia Di Gaetano<sup>3</sup>, Alice A. Lin<sup>4</sup>, Paul Nebbia<sup>5</sup>, Alberto Piazza<sup>3</sup>, Roy J. King<sup>4</sup>, Peter Underhill<sup>6</sup>, Jacques Chironi<sup>1,2\*</sup>

**1** Etablissement Français du Sang PACA Corse, Biologie des Groupes Sanguins, Marseille, France, **2** Aix Marseille Univ, CNRS, EFS, ADES, "Biologie des Groupes Sanguins", Marseille, France, **3** Department of Medical Sciences, University of Turin, Turin, Italy, **4** Department of Psychiatry, Stanford University School of Medicine, Stanford, California, United States of America, **5** Sartène Museum, Sartène, France, **6** Department of Genetics, Stanford University School of Medicine, Stanford, California, United States of America

\* [jacques.chironi@efs.sante.fr](mailto:jacques.chironi@efs.sante.fr)



**OPEN ACCESS**

**Citation:** Di Cristofaro J, Mazières S, Tous A, Di Gaetano C, Lin AA, Nebbia P, et al. (2018) Prehistoric migrations through the Mediterranean basin shaped Corsican Y-chromosome diversity. PLoS ONE 13(8): e0200641. <https://doi.org/10.1371/journal.pone.0200641>

**Editor:** Gyaneshwer Chaubey, Banaras Hindu University, INDIA

**Received:** January 4, 2018

**Accepted:** June 30, 2018

**Published:** August 1, 2018

**Copyright:** © 2018 Di Cristofaro et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The authors received no specific funding for this work.

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

**Abbreviations:** kya, Thousand years ago; BCE, Before Christ Era; SNP, Single Nucleotide Polymorphism; Y-STR, Y-chromosome Short

## Abstract

The rarity of human remains makes it difficult to apprehend the first settlements in Corsica. It is admitted that initial colonization could have occurred during the Mesolithic period when glaciations would have shortened the open water travel distance from the continent. Mesolithic sites in Corsica show relatively short and irregular occupation, and suggest discontinuous settling of very mobile groups probably traveling by boat. Previous genetic studies on Corsican populations showed internal differentiation and a relatively poor genetic relationship with continental populations, despite intense historical contacts, however local Mesolithic-based genetic inheritance has never been properly estimated. The aim of this study was to explore the Corsican genetic profile of Y-chromosomes in order to trace the genetic signatures back to the first migrations to Corsica. This study included 321 samples from men throughout Corsica; samples from Provence and Tuscany were added to the cohort. All samples were typed for 92 Y-SNPs, and Y-STRs were also analyzed. Results revealed highly differentiated haplogroup patterns among Corsican populations. Haplogroup G had the highest frequency in Corsica, mostly displaying a unique Y-STR profile. When compared with Provence and Tuscany, Corsican populations displayed limited genetic proximity. Corsican populations present a remarkable Y-chromosome genetic mixture. Although the Corsican Y-chromosome profile shows similarities with both Provence and to a lesser extent Tuscany, it mainly displays its own specificity. This study confirms the high level of genetic diversity in Corsican populations and backs genetic contributions from prehistoric migrations associated with the Mesolithic, Neolithic and Metal Age eras, rather than from historical movements to Corsica, respectively attested by frequencies and TMRCA of haplogroups G2a-L91 and G2a-P15, J2a-M241 and J2-DYS445 = 6, R1b-U152 and R1b-U106. These results suggest that marine routes to reach the Corsican coast in many different points may have led to such a genetic heterogeneity.

Tandem Repeat; TMRCA, Time to the Most Recent Common Ancestor.

## Introduction

Corsica is the fourth largest island of the Mediterranean, located 177km south-east of Provence and 85km west of Tuscany. The rarity of human remains due to the acidity of natural sediments, the sea-level rising and uncertainty regarding the Paleolithic era, all make it difficult to apprehend the first settlements in Corsica.

The most widely accepted hypothesis concerning Homo-Sapiens Corsican settlements is the colonization of the Corsico-Sardinian block via Tuscany throughout the different glacial periods and occupation of the south of the block (currently Sardinia) followed by that of the north (Corsica). The initial colonization may have occurred during the Mesolithic period (18–15,000 years ago) [1]. The most ancient presence of humans is attested by the Mesolithic collective burial site of Campo Stefano (South of Corsica) dated at circa 8,940 BP [2]. Other Mesolithic sites are located in southwestern (Filitosa River) and southern Corsica as well as in Sardinia [2–5]. Major demographic changes occurred in Corsica mainly during the Neolithic period, from the 6<sup>th</sup> millenium BC onwards, due to external contributions attested by an importation of foreign stones to the island, flint and obsidian, and Impressed Ware, Cardial and Bell-Beaker potteries [4,6,7] (S1 Table). Corsican prehistory ended when the Greeks built the city of Alalia in 565 BCE. Following a long occupation by the Romans, from 269 BCE, the island underwent several waves of consecutive invasions (by the Vandals, Byzantines and the Roman papacy) [8].

Surname-based studies showed a very low rate of exogamy between the different Corsican micro-regions, especially in the north and far south of the island [9,10]. Corsican genetic studies on autosomal markers and the Y-chromosome confirmed this differentiation between the south and the rest of the island [10–12].

Genetic relationships between Corsica and Sardinia are controversial. Mitochondrial and Y-chromosome studies showed almost no gene flow transmitted by Sardinian men to Corsica or continental Italy [6,12,13]. More particularly, the I-M26 haplogroup which is rare or absent anywhere other than Sardinia, but homogeneously present throughout this island, is the proof of a founder effect [14]. Conversely, other studies found that the Sardinian population seemed to have a high genetic affinity with the Corsicans [11,15,16].

On a Mediterranean scale, genetic studies showed two clusters: one in the west, including north-west and central Corsica and the south of Sardinia, and one with central and northern Sardinia, southern Corsica, Sicily and Turkey [10,17]. On the contrary, some studies found that current Corsican populations were genetically distinct from Mediterranean populations [18] and that continental populations from Tuscany and France did not seem to have significantly contributed towards the genetic structure of the Corsican populations [19,20] which rather reflected contributions from Southern Europe [18].

Thus, contradictory results have been published concerning the Corsican genetic composition and the relationship with Sardinia and other neighboring regions such as France and Tuscany. It seems clear however that there is a high rate of heterogeneity within the Corsican population, especially between the North and South, supported by archaeological, anthropological and linguistic data. Continental contributions to the Corsican populations seem slight despite intensive historical interactions.

The aim of this study is to explore the genetic profile of a sample of the Corsican population in order to retrace the genetic signatures back to the first migrations to the island during pre-historic times. This study explores the paternal side of the population, including 321 male samples throughout Corsica which were genealogically characterized. Genetic analyses of the Y-chromosome were carried out, part of which has already been published [21,22]. In order to place our samples in a Mediterranean context, samples from Provence (France) [22] and

Tuscany (Italy) were added to the cohort. Finally, Y-chromosome results were compared to databases from published literature on Mediterranean populations.

## Material and methods

### Sampling

A total of 841 individuals from Corsica (France), Provence (France) and Tuscany (Italy) were analyzed. Sample names, geographical coordinates and sample sizes are detailed in Table 1. All samples were collected from healthy donors. All participants signed an informed consent in agreement with the guidelines of the ethical committee of the institutions involved. The study protocol was approved by the *Ministère de l'Enseignement Supérieur et de la Recherche (Ministry of Higher Education and Research)* in France (record number DC-2008-164, formally approved on December 15th, 2008 by the French Ministry of Research).

DNA was extracted from whole blood using the Qiamp blood kit (Qiagen, Courtaboeuf, France). The populations surveyed were either evaluated for the first time in this study, or updated to higher levels of phylogenetic resolution than reported in earlier studies [21,22].

### Y-chromosome SNP analyses

All samples were typed for 92 Y-SNPs. First, 19 SNPs of the main haplogroups were typed in a multiplex SnapShot (Applied Biosystems, France) protocol described in S2 Table: A-P97, B-M181, C-M130, D-M174, E-M96, F-M89, G-M201, H-M69, I-M258, J-M304, K-M526,

**Table 1. Description of the 18 Corsican, 1 Provençal and 3 Tuscan populations under study.**

Populations	Micro-regions	Lat.	Long.	Total (N)
CORSICA	Bastia	42.6972	9.45088	31
	Cap	42.7328	9.36609	8
	Saint Florent	42.6818	9.30370	7
	Balagne	42.6121	8.89077	17
	Calvi	42.5676	8.75722	4
	Casinca	42.4671	9.45996	26
	Ponte Leccia	42.4635	9.20707	14
	Campuloru	42.3159	9.54729	12
	Travu	42.5277	9.35638	22
	Bravone	42.1977	9.55611	7
	Corte	42.3094	9.14902	27
	Sevi	42.2234	8.83009	24
	Taravo	41.9353	9.15521	26
	Ajaccio	41.9192	8.73863	42
	Alta Rocca	41.7282	9.21205	28
	Extreme South	41.7340	9.33455	14
Valinco	41.7168	8.91761	7	
Sartene	41.6218	8.97471	5	
<b>Total CORSICA</b>				<b>321</b>
<b>PROVENCE</b>		<b>43.2964</b>	<b>5.36977</b>	<b>259</b>
TUSCANY	Siena	43.1668	11.3658	86
	Arezzo	43.6985	11.8194	62
	Pisa	43.3993	10.8660	113
<b>Total TUSCANY</b>				<b>261</b>

<https://doi.org/10.1371/journal.pone.0200641.t001>

L-M11, M-Page93, N-M231, O-M175, Q-M242, R-M207, S-M230 and T-M70. Secondly, genotyping was designed to refine each main lineage previously identified. Haplogroups R, J and G were genotyped by specific SNaPshot multiplex protocols. Three multiplex assays were developed for respectively 4, 9 and 4 SNPs of the R sub-haplogroups (Multiplex assay for haplogroup R analysis number 1: M9, SRY1532, M449 and M198; Multiplex assay for haplogroup R analysis number 2: M479, M124, M343, M269, M412, M207, M173, P297 and M478; and Multiplex assay for haplogroup R analysis number 3: S116, M529, U152 and M405); 2 multiplex assays for 3 and 13 SNPs of the sub-haplogroups J (Multiplex assay for haplogroup J analysis number 1: M304, M267 and M172; Multiplex assay for haplogroup J analysis number 2: L222, M530, M304, M267, M92, Page8, M205, M12, M322, Page55, M340, M172, M410 and M67) and 1 multiplex assay for 11 SNPs of the G sub-haplogroups (P15, M406, M201, M377, M547, M285, M286, M287, P16, M527 and P303). SNaPshot multiplex assays were performed according to the manufacturer's protocol and recommendations. PCR and extension primers and their concentrations are given in [S2 Table](#). Y Haplogroups or sub-haplogroups were analyzed using GeneMapper® 4.0.

Further derived samples were genotyped in a hierarchical manner for the 31 markers by direct sequencing (Big Dye 1.1 according to the manufacturer's protocol): haplogroup C: PK2, haplogroup E: M33, V38, M215, M35, M78, V13, V22, M81, M123, M34, haplogroup G: L91, PF3147, L497, M377, haplogroup I: M72, M253, M227, M438, P37.2, M26, M423, M436, M223, haplogroup J: M241, haplogroup Q: M346, M378, haplogroup R: L23, U152, L11 and V88.

Genomic specifications for all markers have been previously reported [23–25] or listed on ISOGG (<http://www.isogg.org/tree/>).

Corsican samples had been previously typed for M201, P287, P15 and L91 [21] and for V13 [26]. Samples from Provence had been previously typed for E-V13, M406, Page94, M423, M269 and the following J-lineages: M304, Page55, M267, M12, M410, M67, M530, M92 and DYS445 [26].

### Y-STR based phylogenetic networks

Samples derived for haplogroups DYS445 = 6, L497, L91, M12, M241, M267, M406, M410, M527, M530, M67, M92, M96, P15, P16, P287, P303, Page55, Page8, V13 and V22, i. e. 298 samples, were typed for short tandem repeats of 17 Y-STR markers using the AmpFISTR Yfiler Kit (Applied Biosystems) according to manufacturer recommendations. Three additional markers, DYS388, DYS445 and DYS461, were typed separately ([S3 Table](#)) [27,28].

Phylogenetic networks were constructed for haplogroups represented by more than 10 individuals using the program Network 4.6.1.1 (Fluxus-Engineering) and applying the median joining algorithm with the following STR loci: DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS439 [29]. Networks were constructed both at a Corsican level and at a broader level with data from Provence (this study and [22]), Tuscany and continental Italy [30].

### Time to the Most Recent Common Ancestor (TMRCA)

TMRCA for haplogroups with frequencies above 3% were estimated using eight Y-STR markers (DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393) and a mean evolutionary rate mutation of  $0.000069 \pm 0.000013$  per locus per 25 years [31].

## Spatial frequency maps

Spatial frequency maps were drawn up for haplogroups and related sub-haplogroups that were detected with a frequency of at least 3%. Frequency data were converted into frequency maps using Surfer software (version 8, Golden Software, Inc.), following the *kriging* interpolation method [32].

## Mediterranean genetic relationships

Y chromosome haplogroup frequencies from 83 Mediterranean populations in published data (North Africa, Iberia, Italian Peninsula, Adriatic Balkans and Western Mediterranean islands) totalizing 7588 men were harmonized to a common set of 12 lineages: E-M96 (xM35), E-M35 (xM78), E-M78, G-M201, I-M170, J1-M267, J2-M172, R2-M479, R1-M173, K-LT-NOP-M9 (xM207), CF-M130-M89 (xM201, xM170, xM304, xM9), and Y (xM174, xM130, xM89) [6,14,33–42]. Genetic relationships within the Liguria and Tyrrhenian Seas were analyzed in a subset of 31 populations out of 83 refined for 23 haplogroups: E-M96 (xM78), E-M78, E-M81, G-M201, J1-M267, J2-M172\*, J2-M12, J2-M410 (xM67, xDYS445 = 6), J2-DYS445 = 6, J2-M67\*, J2-M92, R2-M124, R1a-M173\*, R1a1-M17, R1b-M269, I (xM438), I2-M438 (xM26, xM423), I2-M26, I2-M423, I2-M223, Q-M242, Y (xM96, xM522, xM201), and K-M9 (xM74). Y-Chromosome genetic diversity based on haplogroup frequencies was calculated with the ARLEQUIN v3.5.1.2 package [43]. Finally, our Y-STR data was merged with those from Sardinia, Sicily and continental Italy [30]. Variance and Rst genetic distances were respectively plotted through a Principal Component analysis (PCA) and Multidimensional Scaling (MDS) using the XLSTAT tool (Data Analysis and Statistical Solution for Microsoft Excel. Addinsoft, Paris, France, 2017).

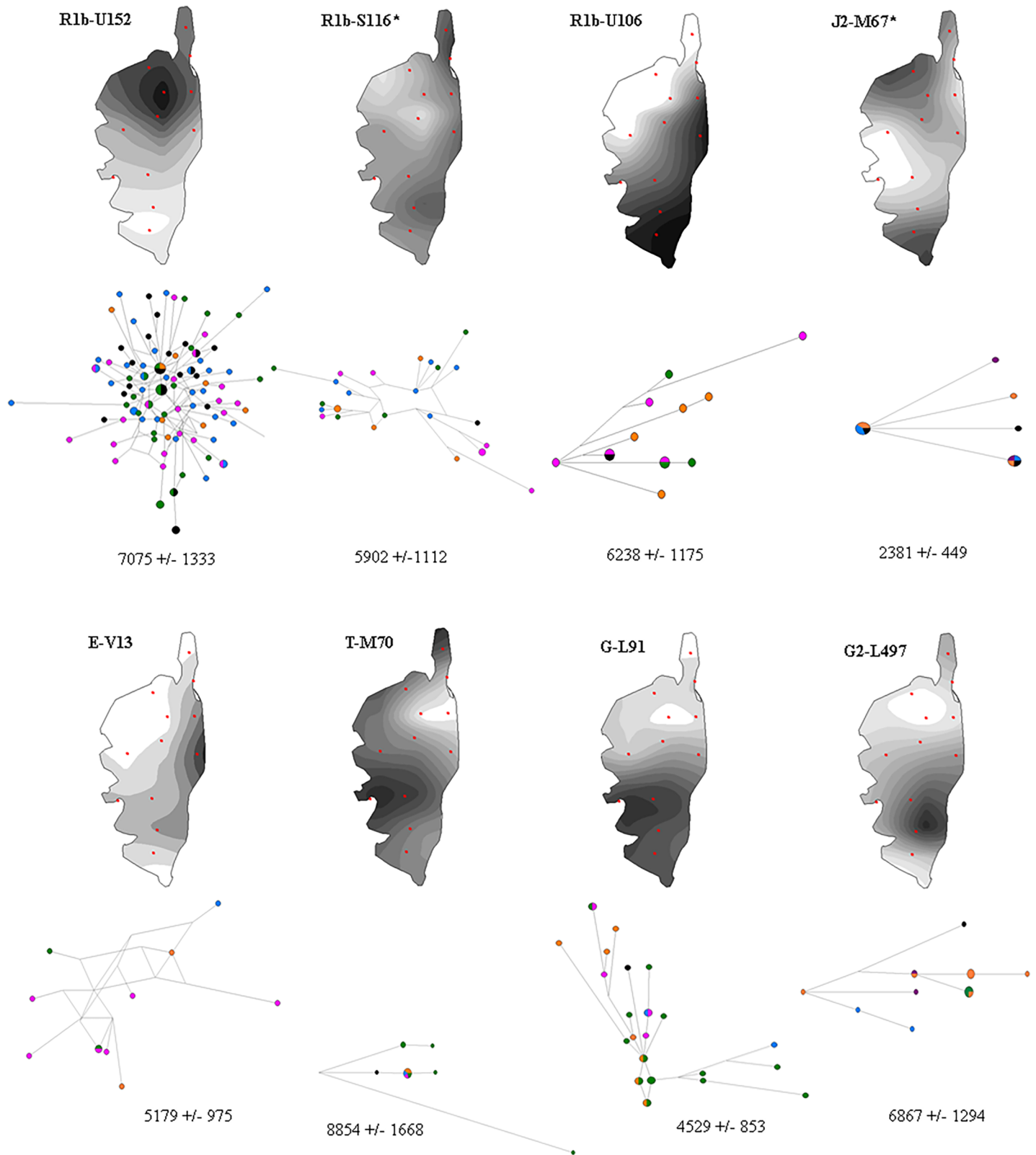
## Results

### Corsican genetic structure and diversity

Hierarchic phylogenetic relationships and frequencies, as well as haplogroup diversity of the 54 paternal haplogroups observed in the 321 Corsican, 259 Provençal and 261 Tuscan samples are described in S1 Fig. Spatial frequency maps performed on haplogroups with frequencies above 3%, their Y-STR based phylogenetic networks and TMRCA are presented in Fig 1.

Haplogroup R represented approximately half of the lineages in both Corsican and Tuscan samples (respectively 51.8% and 45.3%) whereas it reached 90% in Provence. Sub-clade R1b1a1a2a1a2b-U152 predominated in North Corsica whereas R1b1a1a2a1a1-U106 was present in South Corsica. Both SNPs display clinal distributions of frequency variation in Europe, the U152 branch being most frequent in Switzerland, Italy, France and Western Poland. Calibrated branch lengths from whole Y chromosome sequencing [44,45] and ancient DNA studies [46] both indicated that R1a and R1b diversification began relatively recently, about 5 Kya, consistent with Bronze Age and Copper Age demographic expansion. TMRCA estimations are concordant with such expansion in Corsica.

Haplogroup G reached 21.7% in Corsica and 13.3% in Tuscany. Sub-clade G2a2a1a2-L91 accounted for 11.3% of all haplogroups in Corsica yet was not present in Provence or in Tuscany. Thirty-four out of the 37 G2a2a1a2-L91 displayed a unique Y-STR profile, illustrated by the star-like profile of STR networks (Fig 1). G2a2a1a2-L91 and G2a2a-PF3147(xL91xM286) show their highest frequency in present day Sardinia and southern Corsica compared to low levels from Caucasus to Southern Europe, encompassing the Near and Middle East [21,47–50]. Ancient DNA results from Early and Middle Neolithic samples reported the presence of haplogroup G2a-P15 [51–53], consistent with gene flow from the Mediterranean region during



**Fig 1. Spatial frequency maps for haplogroups with frequencies above 3%, their Y-STR based phylogenetic networks in Corsican populations (Blue: North, Green: West, Orange: South, Black: Center and Purple: East) and their TMRCA (in years, +/- SE).**

<https://doi.org/10.1371/journal.pone.0200641.g001>

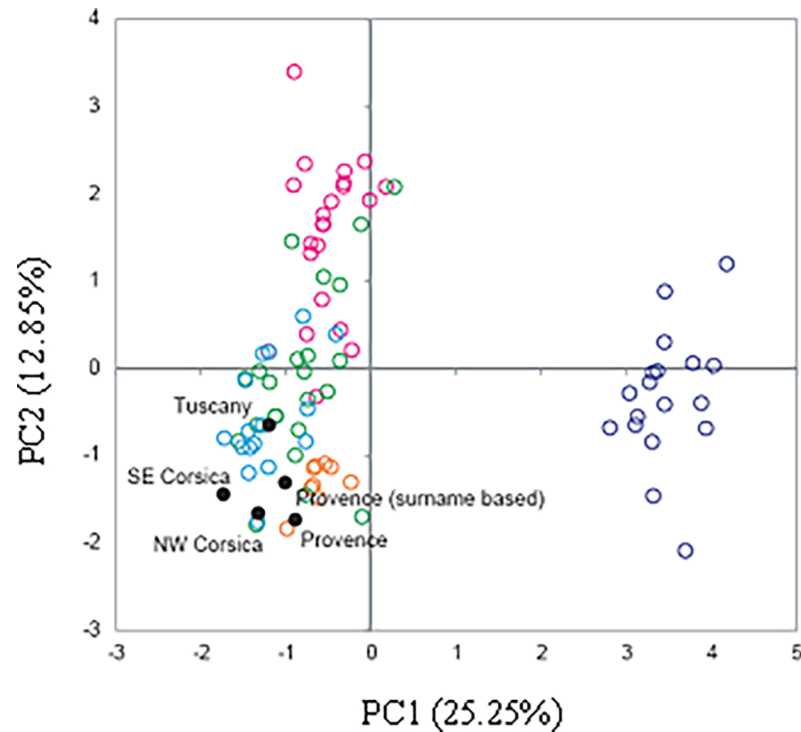
the Neolithic transition. Td expansion time estimated by STR for P15-affiliated chromosomes was estimated to be 15,082 $\pm$ 2217 years ago [49]. Ötzi, the 5,300-year-old Alpine mummy, was derived for the L91 SNP [21]. A genetic relationship between G haplogroups from Corsica and Sardinia is further supported by DYS19 duplication, reported in North Sardinia [14], and observed in the southern part of the Corsica in 9 out of 37 G2a2a1a2-L91 chromosomes and in 4 out of 5 G2a2a-PF3147(xL91xM286) chromosomes, 3 of which displayed an identical STR profile (S4 Table). This lineage has a reported coalescent age estimated by whole sequencing in Sardinian samples of about 9,000 years ago. This could reflect common ancestors coming from the Caucasus and moving westward during the Neolithic period [48], whereas their continental counterparts would have been replaced by rapidly expanding populations associated with the Bronze Age [46,54,55]. Estimated TMRCA for L91 lineage in Corsica is 4529  $\pm$  853 years. G-L497 showed high frequencies in Corsica compared to Provence and Tuscany, and this haplogroup was common in Europe, but rare in Greece, Anatolia and the Middle East. Fifteen out of the 17 Corsican G2a2b2a1a1b-L497 displayed a unique Y-STR profile (S4 Table) with an estimated TMRCA of 6867  $\pm$  1294 years. Haplogroup G2a2b1-M406, associated with Impressed Ware Neolithic markers, along with J2a1-DYS445 = 6 and J2a1b1-M92 [22,49], had very low levels in Corsica. Conversely, G2a2b2a-P303 was highly represented and seemed to be independent of the G2a2b1-M406 marker. The 7 G2a2b2a-P303(xL497xM527) Corsican chromosomes displayed a unique Y-STR profile (S4 Table).

Haplogroup J, mainly represented by J2a1b-M67(xM92), displayed intermediate frequencies in Corsica compared to Tuscany and Provence. J2a1b-M67(xM92) derived STR network analysis displayed a quite homogeneous profile across the island with an estimated TMRCA of 2381  $\pm$  449 years (Fig 1) and individuals displaying M67 were peripheral compared to North-western Italians (S2 Fig). The haplogroup J2a1-Page55(xM67xM530), characteristic of non-Greek Anatolia [22], was found in the north-west of Corsica. Haplogroup J2a1-DYS445 = 6 was found in the north-west with DYS391 = 10 repeats, and in the far south with DYS391 = 9 repeats, the former was associated with Anatolian Greek samples, whereas the second was found in central Anatolia [22]. The 7 J2b2a-M241 displayed a unique Y-STR profile (S4 Table), they were only detected in the Cap Corse region, this sub-haplogroup shows frequency peaks in both the southern Balkans and northern-central Italy [56] and is associated with expansion from the Near East to the Balkans during Neolithic period [57].

Haplogroup E, mainly represented by E1b1b1a1b1a-V13, displayed intermediate frequencies in Corsica compared to Tuscany and Provence. E1b1b1a1b1a-V13 was thought to have initiated a pan-Mediterranean expansion 7,000 years ago starting from the Balkans [52] and its dispersal to the northern shore of the Mediterranean basin is consistent with the Greek Anatolian expansion to the western Mediterranean [22], characteristic of the region surrounding Alaria, and consistent with the TMRCA estimated in Corsica for this haplogroup. A few E1b1a-V38 chromosomes are also observed in the same regions as V13.

Haplogroup I was not found in Provence whereas it was present in both Corsica and Tuscany, respectively as I1 and I2 lineage, although at low levels. Haplogroup I is widespread throughout Europe with a strong geographic differentiation but virtually absent elsewhere. Haplogroup I1 is mostly found in northern Europe; haplogroup I2 has two sub-branches defined by P37.2 and M436 markers, respectively. The latter sub-branch is widespread Balkan Peninsula [58]. Ancient DNA results from Neolithic samples reported the presence of haplogroups I2-M438 and I2a1-P37.2 in southern France [52]. The I2a1a1-M26 marker, found in 30% of Sardinian samples [14,48], was present at very low levels in Corsica.

Haplogroup Q-M242 was mostly found in Sevi (5 of the 6 chromosomes) with quite similar STR patterns (S4 Table). Haplogroup Q, reported to have originated in Central Asia, is



**Fig 2. First and second axes of the PCA based on 12 Y-chromosome haplogroup frequencies in 83 west Mediterranean populations.** North Africa: Purple, Balkans: Pink, Iberia: Orange, Continental Italy: Blue, West Mediterranean Islands: Green.

<https://doi.org/10.1371/journal.pone.0200641.g002>

distributed widely in North Eurasia and at low frequencies in Europe, East Asia, and the Middle East [59].

Haplogroup T1a-M70, known as Thomas Jefferson’s Y chromosome and relatively rare in other Near Eastern populations [60], was observed in Corsica with an estimated TMRCA of 8854 +/- 1668 years.

### Mediterranean populations’ genetic relationships

The Corsican population displayed a significant portion of variance when compared with Provence and Tuscany ( $F_{st} = 11.57\%$ ,  $p = 0.000$ ). The greatest variance was found when four groups were made up ( $F_{st} = 12.11\%$ ,  $p = 0.000$ ) with Pisa, north-western Corsica (Corte, Balagne-Calvi and Ponte Leccia), Provence, and the rest of Corsica, labeled East/South Corsica.

PCA (Fig 2) shows the genetic relationships across the western Mediterranean basin with Corsican populations presented according to SAMOVA results (SAMOVA results not shown). PC1 and 2 account for 38% of total variance. PC1 shows strong departure of Balkan, Iberian, Italian and insular populations from North Africa, due to the exclusive J1-M267 and E-M35 (xM78) in the latter. PC2 shows an east-to-west dispersal of the continental populations along the northern rim of the Mediterranean basin. In this continental pattern, the Tuscan samples under study cluster with those from continental Italy, whereas those from Provence are placed between samples from Iberia and continental Italy. Western Mediterranean islands Sicily and Sardinia cover all continental populations from the Balkans to Italy. North-west Corsican populations coincide with North-West Apulia (Italy) [34] and with Ibiza, whereas

East/South Corsican populations are the furthest away from all of the western Mediterranean populations.

Haplogroup frequency analysis supports genetic similarities between Corsican, Cagliari and northern Sardinian Tempio populations, whereas Tuscan and Provençal samples cluster with continental Italy. Y-STR data showed genetic relationships of Provence and northern Corsican samples with Tuscan, central Italian and Ligurian populations, while southern Corsicans clustered with southern Italians and Sicily (S3 and S4 Figs).

## Discussion

The genetic profile of Corsican Y-chromosomes was explored in order to retrace the genetic signatures back to the first migrations to the island during prehistoric times. A total of 842 individuals including 321 from Corsica, 259 from Provence (France) and 261 from Tuscany (Italy) were analyzed for Y-chromosome haplogroups. Their results were compared to databases from published literature on Mediterranean populations.

Our results showed a highly heterogeneous distribution of Y-chromosome haplogroups in the 183km long and 83km-wide island, with a mosaic of numerous small-scale, micro-regional population groups.

The most ancient Corsican genetic signatures might be represented by haplogroup I2 derived individuals, as this haplogroup was reported to be associated to human expansion phase of Sardinia during pre-Neolithic period [14]. Neolithic migrations to Corsica from the Caucasus moving westward may be supported by the high frequency of haplogroup G-L91 and its estimated TMRCA [21,47–49]. The Alta Rocca region is well known for its archaeological records, particularly in Curacchiaghiu, attesting human presence during the early Neolithic (5650 BCE) and late Neolithic (2980 years BCE) periods [4,61]. Early Neolithic culture in Curacchiaghiu may have encountered two trends, one restricted in space and time, characterized by “pseudo-cardial” pottery, and one, characterizing most of early Neolithic remains, with stamped-impressed pottery. Notably, this region is characterized by a concentrated G-L497 frequency with a dispersed STR profile and an estimated TMRCA concordant with the Neolithic period. Our results show that L91-G2a and L497-G2a lineages are quite unique to the Corsican population. This might be a signature of Neolithic migration conserved over time due to particular Corsican geography and a low rate of exogamy whereas on the Continent they were replaced by expansion during the Bronze Age [21,46]. Haplogroups associated to Neolithic Impressed Ware (J-DYS445 = 6 and J-M92) [22,49] found in the north and south of the island were respectively associated to STR described in Anatolian Greek samples and central Anatolia which further supported independent settlements on the island during the Neolithic period. Furthermore, the haplogroup J-M12, only detected in the Cap Corse region and described to correlate with the distribution of archaeological painted pottery and anthropomorphic figurines [26] suggested an additional Neolithic migration route.

Differential distribution of R1b-U152 and R1b-U106 haplogroups and their respective TMRCA seem to coincide with the two groups of Menhir-statue mostly erected during the Bronze Age. The northern Menhir-statue group has slim figures and prominent ears, whereas in the larger Menhir-statue group, located south of the Ajaccio-Solenzara line, they are rougher and display warrior attributes [4,5,61].

Greek Anatolian historical settlement in Alaria coincided strikingly with haplogroup E-V13 distribution [22]. Haplogroup Q-M242 is mostly represented by sub-group M346 in the Sevi area. M346 appears in Central, Western and Southern Asia, and most parts of Europe [62].

These results support the hypothesis of multiple discontinuous settling in Corsica and may hint a presence of Mesolithic lineages in Corsican patrilineal gene pool, in accordance with

archaeological data from Mesolithic, Neolithic and Bronze Age periods. They also confirmed, as previously reported by genetic and surname-based studies, a very low rate of exogamy within the island leading to Y-chromosome pattern distribution structured by geography, due to the mountainous landscape of Corsica.

Concerning genetic relationships between Corsica and Sardinia, our results corroborated with different genetic heritage for both islands. However, high genetic affinity between the Sardinian population and the Corsicans was previously reported [11,15,16] and is supported by the similarity of the southern Corsican and Sardinian Mesolithic sites [2,3], as well as the similarity of the Corsican « Torrean » and Sardinian « Nuragic » cultures. Our results on haplogroup G samples and their STR profile are in accordance with a genetic continuity between South Corsica and Sardinia [14]. Conversely, our results for haplogroup I support a genetic distinction between these two islands as previously reported [6,12,13].

The Corsico-Sardinian block is characterized by climatic contrast: glacial sediments found in the north of Corsica reveal three glacial periods, whereas these sediments are absent in Sardinia where the climate was characterized by rainy periods. It is reasonable to assume that the first groups of humans would have preferentially settled in regions where the climate was more clement. Then, only when the climate became warmer, would these populations have moved towards Corsica from the north of Sardinia [16]. Overall, archaeological and genetic data support a common Mesolithic genetic background, estimated in Sardinia at ~7700 years ago [47] and transformed by Neolithic migrations [63] and Bronze Age heritage [4,5,61].

On a western Mediterranean scale, genetic studies have published controversial results, with studies showing on the one hand affinities between southern Corsica, central and northern Sardinia, Sicily and Turkey, and on the other hand between north-west and central Corsica and the south of Sardinia [10,17]; on the contrary, other studies found that current Corsican populations were genetically distinct from other Mediterranean populations [18–20].

Our results support genetic affinities between north-western Corsica and continental populations whereas East/South Corsican populations originated from western Mediterranean populations. These two clusters are in accordance with the dramatic divergence of populations from north and south Corsica described here and supported by genetic, linguistic and archaeological studies and possibly explained by a divergence of population history accumulated since Neolithic times between the north and south of the island. These results are in accordance with the hypothesis of different genetic destinies for populations from mainland Europe and those from the islands of the Tyrrhenian sea, reported to have begun to diverge at least 5,000 years ago [21].

In conclusion, whereas it is widely believed that settlers in Corsica arrived from mainland Italy, due to a difference in the sea level during glaciations and the formation of a natural bridge with Tuscany, our results support that Corsica might have been colonized by many different waves of migration since Neolithic period and Bronze Age and led to such a heterogeneous Y-chromosome profile of the current Corsican population. The overall complexity of the genetic profile of Corsicans remains to be fully addressed with the study of complementary mitochondrial and autosomal markers.

## Supporting information

**S1 Fig. Corsica, Provence and Tuscany Y-chromosome tree.** Hierarchic phylogenetic relationships and frequencies (percentages), haplogroup diversity observed in the 321 Corsican, 259 Provençal and 261 Tuscan samples. (XLSX)

**S2 Fig. Y-STR based phylogenetic networks.** Populations from Corsica, Provence and Tuscany under study and Italian populations from Boattini A, et al. 2013 [30].

(TIF)

**S3 Fig. First and second axes of the PCA based on 23 Y-chromosome haplogroup frequencies in 31 populations from the northern rim of the Mediterranean basin.**

(TIF)

**S4 Fig. Multidimensional Scaling (MDS) of Rst genetic distances between populations of the Tyrrhenian Sea (stress: 0.659).** Black: present study, open squares: Oltremontano speakers, filled squares: Cismontano speakers. Others colors refer to the eight Italian areas in Boattini A, et al. 2013 [30] (Light Green: Northwestern Italy, Grey: Northeastern Italy, Red: Bologna, Pink: Tuscany, Light Blue: Central Italy, Yellow: Southern Italy, Blue: Sicily, Green: Sardinia).

(TIF)

**S1 Table. Chronological List of Major Archaeological Strata adapted from D'Anna A et al. 2007 [61].**

(DOCX)

**S2 Table. Primers and their concentrations used for multiplex SNaPshot assays to simultaneously analyze: main Y-chromosome haplogroups, haplogroups R (3 multiplex analyses), haplogroup J (2 multiplex analyses) and haplogroup G.** F and R stand for Forward and Reverse respectively.

(XLSX)

**S3 Table. Primers and their concentrations used for multiplex fragment analysis assay to simultaneously analyze STR DYS388, DYS445 and DYS461.** F and R stand for Forward and Reverse respectively.

(DOC)

**S4 Table. Y-STR data from this study, Boattini et al. 2013 and King et al. 2011.**

(XLSX)

## Acknowledgments

The authors are grateful to the Corsican men who donated DNA.

SM thanks Avril Meffray for her help in Y-chromosome and archaeological database building.

## Author Contributions

**Conceptualization:** Cornelia Di Gaetano, Alice A. Lin, Alberto Piazza, Roy J. King, Peter Underhill, Jacques Chiaroni.

**Data curation:** Julie Di Cristofaro, Stéphane Mazières, Audrey Tous, Cornelia Di Gaetano, Alice A. Lin, Paul Nebbia, Alberto Piazza.

**Formal analysis:** Julie Di Cristofaro, Stéphane Mazières, Alice A. Lin, Peter Underhill.

**Funding acquisition:** Jacques Chiaroni.

**Investigation:** Julie Di Cristofaro, Alice A. Lin, Paul Nebbia, Roy J. King, Peter Underhill, Jacques Chiaroni.

**Methodology:** Julie Di Cristofaro, Stéphane Mazières, Jacques Chiaroni.

**Project administration:** Jacques Chiaroni.

**Resources:** Jacques Chiaroni.

**Software:** Julie Di Cristofaro, Stéphane Mazières.

**Validation:** Julie Di Cristofaro, Alice A. Lin.

**Writing – original draft:** Julie Di Cristofaro.

**Writing – review & editing:** Stéphane Mazières, Cornelia Di Gaetano, Alice A. Lin, Paul Nebbia, Alberto Piazza, Roy J. King, Peter Underhill, Jacques Chiaroni.

## References

1. Bonifay E, Gauthier A, Weiss MC, Camps G, Cesari J, et al. (1990) Préhistoire de la Corse. 126 p.
2. Courtaud P, Cesari J., Léandri, Nebbia P, Perrin T, et al. La sépulture mésolithique de Campu Stefanu (Sollacaro, Corse du Sud, France). *Chronologie de la préhistoire récente dans le Sud de la France.*; 2012; Porticcio (20) France. *Archives d'écologie préhistorique Toulouse.* pp. 275–287.
3. Costa LJ (2004) Nouvelles données sur le mésolithique des îles Tyrrhéniennes (Corse et Sardaigne) [Peut-on parler d'un « mésolithique insulaire »?] *Gallia préhistoire* 46: 211–230.
4. Graziani AM, Cesari J, Bonifay E, Demouche F, Pasquet A, et al. (2013) Histoire de la Corse; Piazzola A, editor.
5. Nebbia P, Ottaviani JC (1988) Signification paléthnologique et interprétation des statues Menhirs de Corse. *Bulletin de la société des sciences historiques et naturelles de la Corse.* Editions Cynes et Méditerranée Ajaccio ed. pp. 69–77.
6. Francalacci P, Morelli L, Underhill PA, Lillie AS, Passarino G, et al. (2003) Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. *Am J Phys Anthropol* 121: 270–279. <https://doi.org/10.1002/ajpa.10265> PMID: 12772214
7. Tozzi CW, J.-C. (2001) Nouvelles données sur le Néolithique ancien de l'aire corso-toscane. *Bulletin de la Société préhistorique française* 98: 445–458.
8. Arrighi JM (2002) Histoire de la langue Corse: Pollina.
9. Morelli L, Paoli G, Francalacci P (2002) Surname analysis of the Corsican population reveals an agreement with geographical and linguistic structure. *J Biosoc Sci* 34: 289–301. PMID: 12117210
10. Ghiani ME, Varesi L, Mitchell RJ, Vona G (2009) Significant genetic differentiation within the population of the Island of Corsica (France) revealed by y-chromosome analysis. *Genet Test Mol Biomarkers* 13: 773–777. <https://doi.org/10.1089/gtmb.2009.0075> PMID: 19810819
11. Grimaldi MC, Crouau-Roy B, Amoros JP, Cambon-Thomsen A, Carcassi C, et al. (2001) West Mediterranean islands (Corsica, Balearic islands, Sardinia) and the Basque population: contribution of HLA class I molecular markers to their evolutionary history. *Tissue Antigens* 58: 281–292. PMID: 11844138
12. Scozzari R, Cruciani F, Pangrazio A, Santolamazza P, Vona G, et al. (2001) Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Hum Immunol* 62: 871–884. PMID: 11543889
13. Morelli L, Grosso MG, Vona G, Varesi L, Torrioni A, et al. (2000) Frequency distribution of mitochondrial DNA haplogroups in Corsica and Sardinia. *Hum Biol* 72: 585–595. PMID: 11048788
14. Contu D, Morelli L, Santoni F, Foster JW, Francalacci P, et al. (2008) Y-chromosome based evidence for pre-neolithic origin of the genetically homogeneous but diverse Sardinian population: inference for association scans. *PLoS One* 3: e1430. <https://doi.org/10.1371/journal.pone.0001430> PMID: 18183308
15. Vona G, Memmi MR, Varesi L, Mameli GE, Succa V (1995) A study of several genetic markers in the Corsican population (France). *Anthropol Anz* 53: 125–132. PMID: 7668841
16. Varesi L, Memmi M, Cristofari MC, Mameli GE, Calo CM, et al. (2000) Mitochondrial control-region sequence variation in the Corsican population, France. *Am J Hum Biol* 12: 339–351. [https://doi.org/10.1002/\(SICI\)1520-6300\(200005/06\)12:3<339::AID-AJHB4>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1520-6300(200005/06)12:3<339::AID-AJHB4>3.0.CO;2-U) PMID: 11534024
17. Quintana-Murci L, Veitia R, Fellous M, Semino O, Poloni ES (2003) Genetic structure of Mediterranean populations revealed by Y-chromosome haplotype analysis. *Am J Phys Anthropol* 121: 157–171. <https://doi.org/10.1002/ajpa.10187> PMID: 12740959

18. Gonzalez-Perez E, Moral P, Via M, Vona G, Varesi L, et al. (2007) The ins and outs of population relationships in west-Mediterranean islands: data from autosomal Alu polymorphisms and Alu/STR compound systems. *J Hum Genet* 52: 999–1010. <https://doi.org/10.1007/s10038-007-0206-6> PMID: 17957332
19. Vona G, Moral P, Memmi M, Ghiani ME, Varesi L (2003) Genetic structure and affinities of the Corsican population (France): classical genetic markers analysis. *Am J Hum Biol* 15: 151–163. <https://doi.org/10.1002/ajhb.10133> PMID: 12621603
20. Grimaldi MC, Crouau-Roy B, Contu L, Amoros JP (2002) Molecular variation of HLA class I genes in the Corsican population: approach to its origin. *Eur J Immunogenet* 29: 101–107. PMID: 11918634
21. Keller A, Graefen A, Ball M (2012) New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat Commun*.
22. King RJ, Di Cristofaro J, Kouvatsi A, Triantaphyllidis C, Scheidel W, et al. (2011) The coming of the Greeks to Provence and Corsica: Y-chromosome models of archaic Greek colonization of the western Mediterranean. *BMC Evol Biol* 11: 69. <https://doi.org/10.1186/1471-2148-11-69> PMID: 21401952
23. Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, et al. (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65: 43–62. PMID: 11415522
24. Underhill PA, Myres NM, Rootsi S, Metspalu M, Zhivotovsky LA, et al. (2010) Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur J Hum Genet* 18: 479–484. <https://doi.org/10.1038/ejhg.2009.194> PMID: 19888303
25. Rootsi S, Behar DM, Jarve M, Lin AA, Myres NM, et al. (2013) Phylogenetic applications of whole Y-chromosome sequences and the Near Eastern origin of Ashkenazi Levites. *Nat Commun* 4: 2928. <https://doi.org/10.1038/ncomms3928> PMID: 24346185
26. King R, Underhill PA (2002) Congruent distribution of Neolithic painted pottery and ceramic figurines with Y-chromosome lineages. *Antiquity* 76: 707–714.
27. White PS, Tatum OL, Deaven LL, Longmire JL (1999) New, male-specific microsatellite markers from the human Y chromosome. *Genomics* 57: 433–437. <https://doi.org/10.1006/geno.1999.5782> PMID: 10329011
28. Kayser M, de Knijff P, Dieltjes P, Krawczak M, Nagy M, et al. (1997) Applications of microsatellite-based Y chromosome haplotyping. *Electrophoresis* 18: 1602–1607. <https://doi.org/10.1002/elps.1150180920> PMID: 9378129
29. Cooper G, Amos W, Hoffman D, Rubinsztein DC (1996) Network Analysis of Human Y Microsatellite Haplotypes. *Human Molecular Genetics* 5: 1759–1766. PMID: 8923004
30. Boattini A, Martinez-Cruz B, Sarno S, Harmant C, Useli A, et al. (2013) Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata. *PLoS One* 8: e65441. <https://doi.org/10.1371/journal.pone.0065441> PMID: 23734255
31. Zhivotovsky LA, Underhill PA, Cinnioğlu C, Kayser M, Morar B, et al. (2004) The Effective Mutation Rate at Y Chromosome Short Tandem Repeats, with Application to Human Population-Divergence Time. *American Journal of Human Genetics* 74: 50–61. <https://doi.org/10.1086/380911> PMID: 14691732
32. Chiaroni J, Underhill PA, Cavalli-Sforza LL (2009) Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc Natl Acad Sci U S A* 106: 20174–20179. <https://doi.org/10.1073/pnas.0910803106> PMID: 19920170
33. King RJ, Özcan SS, Carter T, Kalfoglu E, Atasoy S, et al. (2008) Differential Y-chromosome anatolian influences on the greek and cretan neolithic. *Annals of Human Genetics* 72: 205–214. <https://doi.org/10.1111/j.1469-1809.2007.00414.x> PMID: 18269686
34. Capelli C, Brisighelli F, Scarnicci F, Arredi B, Caglia A, et al. (2007) Y chromosome genetic variation in the Italian peninsula is clinal and supports an admixture model for the Mesolithic-Neolithic encounter. *Molecular Phylogenetics and Evolution* 44: 228–239. <https://doi.org/10.1016/j.ympev.2006.11.030> PMID: 17275346
35. Battaglia V, Fornarino S, Al-Zahery N, Olivieri A, Pala M, et al. (2009) Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe. *European Journal of Human Genetics* 17: 820–830. <https://doi.org/10.1038/ejhg.2008.249> PMID: 19107149
36. Di Gaetano C, Cerutti N, Crobu F, Robino C, Inturri S, et al. (2009) Differential Greek and northern African migrations to Sicily are supported by genetic evidence from the Y chromosome. *European Journal of Human Genetics* 17: 91–99. <https://doi.org/10.1038/ejhg.2008.120> PMID: 18685561
37. Peričić M, Lauc LB, Klarić IM, Janičijević B, Rudan P (2005) Review of Croatian genetic heritage as revealed by mitochondrial DNA and Y chromosomal lineages. *Croatian Medical Journal* 46: 502–513. PMID: 16100752

38. Zalloua PA, Xue Y, Khalife J, Makhoul N, Debiane L, et al. (2008) Y-Chromosomal Diversity in Lebanon Is Structured by Recent Historical Events. *American Journal of Human Genetics* 82: 873–882. <https://doi.org/10.1016/j.ajhg.2008.01.020> PMID: 18374297
39. Fadhlouli-Zid K, Garcia-Bertrand R, Alfonso-Sanchez MA, Zemni R, Benammar-Elgaaied A, et al. (2015) Sousse: extreme genetic heterogeneity in North Africa. *J Hum Genet* 60: 41–49. <https://doi.org/10.1038/jhg.2014.99> PMID: 25471516
40. Triki-Fendri S, Sanchez-Diz P, Rey-Gonzalez D, Ayadi I, Carracedo A, et al. (2015) Paternal lineages in Libya inferred from Y-chromosome haplogroups. *Am J Phys Anthropol* 157: 242–251. <https://doi.org/10.1002/ajpa.22705> PMID: 25677690
41. Marjanovic D, Fornarino S, Montagna S, Primorac D, Hadziselimovic R, et al. (2005) The peopling of modern Bosnia-Herzegovina: Y-chromosome haplogroups in the three main ethnic groups. *Ann Hum Genet* 69: 757–763. <https://doi.org/10.1111/j.1529-8817.2005.00190.x> PMID: 16266413
42. Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, et al. (2004) A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. *Am J Hum Genet* 75: 338–345. <https://doi.org/10.1086/423147> PMID: 15202071
43. Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10: 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x> PMID: 21565059
44. Karmin M, Saag L, Vicente M, Wilson Sayres MA, Jarve M, et al. (2015) A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res* 25: 459–466. <https://doi.org/10.1101/gr.186684.114> PMID: 25770088
45. Underhill PA, Poznik GD, Rootsi S, Jarve M, Lin AA, et al. (2015) The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur J Hum Genet* 23: 124–131. <https://doi.org/10.1038/ejhg.2014.50> PMID: 24667786
46. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, et al. (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522: 207–211. <https://doi.org/10.1038/nature14317> PMID: 25731166
47. Francalacci P, Morelli L, Angius A, Berutti R, Reinier F, et al. (2013) Low-Pass DNA Sequencing of 1200 Sardinians Reconstructs European Y-Chromosome Phylogeny. *Science* 341: 565–569. <https://doi.org/10.1126/science.1237947> PMID: 23908240
48. Francalacci P, Sanna D, Useli A, Berutti R, Barbato M, et al. (2015) Detection of phylogenetically informative polymorphisms in the entire euchromatic portion of human Y chromosome from a Sardinian sample. *BMC Research Notes* 8: 174. <https://doi.org/10.1186/s13104-015-1130-z> PMID: 25926048
49. Rootsi S, Myres NM, Lin AA, Jarve M, King RJ, et al. (2012) Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. *European Journal of Human Genetics* 20: 1275–1282. <https://doi.org/10.1038/ejhg.2012.86> PMID: 22588667
50. Voskarides K, Mazieres S, Hadjipanagi D, Di Cristofaro J, Ignatiou A, et al. (2016) Y-chromosome phylogeographic analysis of the Greek-Cypriot population reveals elements consistent with Neolithic and Bronze Age settlements. *Investig Genet* 7: 1. <https://doi.org/10.1186/s13323-016-0032-8> PMID: 26870315
51. Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, et al. (2010) Ancient DNA from European early Neolithic farmers reveals their near eastern affinities. *PLoS Biology* 8.
52. Lacan M, Keyser C, Ricaut FX, Brucato N, Tarrús J, et al. (2011) Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proceedings of the National Academy of Sciences of the United States of America* 108: 18255–18259. <https://doi.org/10.1073/pnas.1113061108> PMID: 22042855
53. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, et al. (2015) Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528: 499–503. <https://doi.org/10.1038/nature16152> PMID: 26595274
54. Allentoft ME, Sikora M, Sjogren KG, Rasmussen S, Rasmussen M, et al. (2015) Population genomics of Bronze Age Eurasia. *Nature* 522: 167–172. <https://doi.org/10.1038/nature14507> PMID: 26062507
55. Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, et al. (2014) Genome flux and stasis in a five millennium transect of European prehistory. *Nat Commun* 5: 5257. <https://doi.org/10.1038/ncomms6257> PMID: 25334030
56. Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, et al. (2004) 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. *American Journal of Human Genetics* 74: 1023–1034. <https://doi.org/10.1086/386295> PMID: 15069642

57. Battaglia V, Fornarino S, Al-Zahery N, Olivieri A, Pala M, et al. (2009) Y-chromosomal evidence of the cultural diffusion of agriculture in Southeast Europe. *Eur J Hum Genet* 17: 820–830. <https://doi.org/10.1038/ejhg.2008.249> PMID: 19107149
58. Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, et al. (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *American Journal of Human Genetics* 75: 128–137. <https://doi.org/10.1086/422196> PMID: 15162323
59. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, et al. (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research* 18: 830–838. <https://doi.org/10.1101/gr.7172008> PMID: 18385274
60. King TE, Bowden GR, Balaesque PL, Adams SM, Shanks ME, et al. (2007) Thomas Jefferson's Y chromosome belongs to a rare European lineage. *Am J Phys Anthropol* 132: 584–589. <https://doi.org/10.1002/ajpa.20557> PMID: 17274013
61. D'Anna A, Cesari J, Ogel L, Vaquer J (2007) Corse et Sardaigne préhistorique. Relations et échanges dans le contexte Méditerranéen.; scientifiques EdCdthe, editor.
62. Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, et al. (2006) 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. *American Journal of Human Genetics* 78: 202–221. <https://doi.org/10.1086/499411> PMID: 16400607
63. Morelli L, Contu D, Santoni F, Whalen MB, Francalacci P, et al. (2010) A comparison of Y-chromosome variation in Sardinia and Anatolia is more consistent with cultural rather than demic diffusion of agriculture. *PLoS One* 5: e10419. <https://doi.org/10.1371/journal.pone.0010419> PMID: 20454687