Unde venisti? The Prehistory of Italic through its Loanword Lexicon

Wigman, A.M.

Citation

Wigman, A. M. (2023, November 1). *Unde venisti?: The Prehistory of Italic through its Loanword Lexicon*. Retrieved from https://hdl.handle.net/1887/3655644

Version: Publisher's Version

6 Population Genetics of Italy

6.1 Genetics Introduction

The genetic diversity of today's Italian population attests to a history characterized by human migrations. Each event contributed to the complex mosaic of alleles, beginning with the first anatomically modern humans to arrive on the Italian peninsula and continuing to this day. Some historical events are well documented, such as the migration during the Roman Empire and after its fall as well as the settlement of ethnic Albanians and Greeks (both ancient and medieval). Others have only been uncovered due to the work of archaeologists and geneticists.

The term Indo-European is first and foremost a linguistic one, and languages are spoken by people. But given the landmark publications that revolutionized the field of research beginning in 2015 (discussed in detail later), a connection between the Indo-European proto-language and the populations of the Pontic-Caspian steppe is becoming the dominant theory in the research on the Indo-European homeland. No discussion of Indo-European homelands and migrations can be complete without a discussion of these results.

This chapter is an excursion into the background and results of genetic research on Italian populations, serving as a segue between the linguistic patterns identified and the final conclusions. After setting the background for the population genetics of Europe, this chapter details the development of Italian population genetics research with the goal of ultimately identifying the genetic events that can be linked to the arrival of the Indo-European languages in Italy.

6.2 European Genetics from the Origins of Agriculture to the Homeland Debate

6.2.1 Understanding the Origin and Spread of Agriculture

Archaeological inquiry has been making headway into the origins of agriculture for over a century, with the term 'Neolithic Revolution' being coined in 1936 by V. Gordon Childe. We know agriculture developed in Southwest Asia (gradually and originally regionally, cf. Arranz-Otaegui et al. 2016), spreading first to Cyprus, Crete, and mainland Greece (cf. Hofmanová et al. 2016, Douka et al. 2017), then forking into two routes. Via the Cardial Ware Culture, agriculture reached the Mediterranean coast and via the Linearbandkeramik Culture (LBK), it spread through mainland Europe.

Understanding how agriculture spread is not as easy of a question to answer. Scholarly opinions differed as to whether farmers from the Near East themselves migrated into Europe (demic diffusion) or whether the technological aspects of farming were adopted and passed on by the people already living in Europe (cultural diffusion). In 1971 Albert J. Ammerman and Luigi Luca Cavalli-Sforza put forward an argument in favor of the former. They plotted the radiocarbon dates that had been collected for the earliest appearance of agriculture at several archaeological sites across Europe against the distance of those sites from areas in the Near East. The results showed a slow but steady 1 km/year spread of the technology of farming from Southwest to Northeast across Europe. Based on anthropological and archaeological observations along with statistical models, this suggested a "wave of advance" model best explained by demic diffusion rather than cultural diffusion.

Cavalli-Sforza was an influential force in the field of population genetics. In 1964[543], with Anthony Edwards, he used 18 alleles of classical markers (mainly blood groups and antigen systems) to examine how 15 widespread population groups in the world diverged from each other. They popularized the then only infrequently used principal components analysis (PCA) and worked on the best ways to generate trees (Edwards and Cavalli-Sforza 1965, 1967). In the 1971 paper that used archaeological material, he and Ammerman had already postulated how genetic information could be used to inform their work in the future. In 1975, in a report with Alberto Piazza, the results of a tree analysis using 58 alleles and a larger number of populations basically corroborated the results of the 1964 study. The distribution of classical markers amongst modern populations seemed to be powerful enough, when used in combination with each other, to perform large-scale studies of the history of human migrations. Interestingly, up until this point, the main focus of the articles was on the development and presentation of the statistical methods. The genetic analyses were used as examples to show how the methods worked.

In 1978, Menozzi, Piazza, and Cavalli-Sforza did what the 1971 article had proposed, applying genetic methodology to the question of the spread of agriculture. The statistical approach they employed had been used for clustering analysis in genetics before, but this time they would display the results in the form of a map. The distribution of individual classical markers across Europe was beginning to be understood by this time. The first principal component of 38 genetic alleles analyzed produced a pattern that matched the Southeast to Northwest cline of the spread of farming known from archaeological research. This first principal component explained a full third of the genetic variation. Assuming that the hunter-gatherers of Europe and the agriculturalists of the Near East had been in isolation from each other long enough to be genetically quite different, this cline across the geography most likely represented the result of one population beginning in the Near East, and spreading to the Northwest, intermarrying with local populations along the way such that it came to be genetically more and more similar to those populations. In other words, this is precisely what one would expect if farming had spread via demic rather than cultural diffusion.

---

[543] Edwards and Cavalli-Sforza 1964; after two unpublished conference talks in 1963.

The second principal component revealed a general East to West cline and the third principal component revealed a Northeast (centering in Ukraine) to Southwest cline, with all three together explaining 50% of the variation. It was clear that more investigation was needed to fully understand what had occurred, but the stage had been set for the use of genetics to understand the population history of Europe. Ammerman and Cavalli-Sforza augmented their wave of advance model with statistics and examples from archaeological investigation in 1979 and combined this augmented approach with the genetic data and a detailed explanation of how it was analyzed in a capstone book in 1984.

6.2.2 The Indo-European Homeland Debate

Ever since the realization that the relatedness of the Indo-European languages meant that they had all descended from a parent language, hypotheses about the homeland of this language came into circulation. This homeland or *Urheimat* was the geographic location where the speakers of the reconstructed proto-language would have lived. In the century or so of research, several locations have been proposed as the potential homeland. The two that have widely been considered the most compelling are the “kurgan hypothesis” (particularly its later revision as the “steppe hypothesis”) and the “Anatolian hypothesis”.

The earliest concrete codification of the kurgan hypothesis came in the form of archaeologist Marija Gimbutas’ 1956 book *The Prehistory of Eastern Europe, Part I.* In it and its further developments, she proposed the umbrella term ‘Kurgan Culture’ to denote several archaeological cultures in Eastern Europe in the 4th and 3rd millennia BCE. They all seemed to share features (including the construction of burial mounds called kurgans) and overlap in terms of movement and development through time. The changes in settlement patterns of the cultures of the kurgan horizon including, at some stages, drastic expansion, led Gimbutas to propose a series of waves of expansion, some of which were responsible for bringing the Indo-European languages along with them. Schrader (1883) had already argued on the basis of linguistic methodology, which would come to be called linguistic paleontology, for a similar Pontic-Caspian homeland for Indo-European.

In 1987, a strong contender to the kurgan hypothesis was put forward by archaeologist Colin Renfrew in his book *Archaeology and Language: the Puzzle of Indo-European Origins*. The book was very much in line with the genetic discoveries that had been made up until that point. If there had indeed been a large-scale migration of people from the Near East during the Neolithic, we can assume that they would have brought their language with them. They brought farming technology with them that resulted in geographically enormous cultural horizons that spread across Western Europe. If we are looking for a group of people spreading in unison with closely related languages across the European continent, we need look no further than this. In proposing that the Indo-European languages spread with Neolithic farmers from Anatolia, this position came to be called the Anatolian hypothesis.

Despite seeming like a smoking gun both archaeologically and genetically, there were major problems with the Anatolian hypothesis from the linguistic side. In a line of research that has culminated most recently in Kroonen et al. (2022), the society for which Proto-Indo-European is reconstructed could not have been fully agricultural.

The kurgan hypothesis as laid out by Marija Gimbutas, while generally in agreement with linguistic data, received criticism in the field. Older explanations for culture change that frequently involved the assumption of migration had given way to a swing of the pendulum in which such explanations were viewed with suspicion (cf. Anthony 2007: 214).[544] But given the linguistic problems with the Anatolian hypothesis, the field of Indo-European studies never fully rejected Gimbutas' hypothesis. Jim Mallory's 1989 *In Search of the Indo-Europeans: Language, Archaeology and Myth* followed the explanation she presented, and David Anthony's *The Horse, the Wheel and Language* presents a revised steppe hypothesis.

6.2.3 Refining and Overturning the Understandings

While archaeologists and linguists were refining and expanding homeland theories, geneticists continued, with the aid of improving technology, to sample and sequence human genomes.

6.2.3.1 Building Genetic Databases

With the hypothesis that farmers had spread into Europe from the Near East via demic diffusion, i.e. in person, during the Neolithic, it was initially unclear how much of a genetic footprint they would have left on the population. Richards et al. (1996) examined a sequence of mitochondrial DNA in living individuals from Europe and the Middle East. The divergence times that they calculated for the lineages led them to conclude that most mitochondrial alleles spread to Europe during the Upper Paleolithic, with only around 10% of the alleles dating to the spread of farming. Perhaps farming had been more of a local development after all. Cavalli-Sforza and Minch (1997) disagreed with the methodology behind this conclusion. The segment of mitochondrial DNA used was subject to higher rates of mutation, and the mitochondrial patterns might actually be revealing a sex-bias in terms of who was migrating. Staying true to the earlier conclusions about demic diffusion, Cavalli-Sforza and Minch estimated the genetic influence of farmers at the more significant but still quite small amount of ca. 27%. Barbujani, Bertorelle, and Chikhi (1998) sided with the latter on this, emphasizing the archaeological data on the spread of farming and the arguments generated by linguists on the post-Neolithic age of many of the language families in question. Additionally, they argued that the age of a haplotype does not equate to the age of the population in which that haplotype is found, especially when the haplotype is not found exclusively within

---

[544] There was less ideological criticism of her work too. Anthony (2007: 306-7) for example points out the problematic way that she grouped several different burial mound-building archaeological cultures into her one "Kurgan culture".

that population. This would become an important argument in understanding the data, repeated for instance by Chikhi et al. (1998) when they found using modern autosomal DNA that all population separation times were less than 10 kya and thus unlikely to be pre-Neolithic.

Torroni et al. (1998) examined modern European mitochondrial DNA and proposed that it demonstrated a large genetic component from a Mesolithic (post-glacial) Iberian population. Mitochondrial haplogroup V was found in its highest concentrations among Berbers, Basque, and Saami, and so was most likely to have spread from an Iberian refugium after the ice sheets melted. Its sister haplogroup H, though it originated in the Near East, seems to have spread to Europe during the Paleolithic, perhaps along with V from Iberia. This was countered by Simoni et al. (2000a; criticized by Torroni et al. 2000 and defended in Simoni et al. 2000b). They found that the Saami were mitochondrially quite unique, and their inclusion in analysis was skewing the data. When removed, European populations were mitochondrially extremely similar to one another, with little geographic structuring. However, when Southern Europe was examined separately, there was a clinal distribution. While the alleles and their frequencies did not differ between North and South, along the Mediterranean, they were distributed in a very structured way. In combination with the autosomal cline from the Near East to Northwest Europe, the results in theory only supported a large-scale directional expansion. If all human populations had been forced into glacial refugia and re-expanded during the Mesolithic, this would have erased the original Paleolithic distribution patterns. As the patterns persist to the present day, they could most likely only have resulted from a directional expansion during the Neolithic.

The debate continued. Richards et al. (2000) took back-migration into the Near East from Europe into account, suggesting that the Neolithic migration contributed at most 20% of the mitochondrial genepool of Europe—substantial but in the minority. Most of the mitochondrial haplogroups arrived in Europe during the Paleolithic and their distribution dates back to founder effects during the post-glacial re-expansion. Using Y-chromosome data, Semino et al. (2000) agreed. They found that ten lineages account for more than 95% of the Y-chromosomes in their study. Clustering analysis seemed to identify two Paleolithic post-glacial migration events and one Neolithic migration event, with most of the genepool nevertheless being of Paleolithic origin. Chikhi et al. (2002) concluded that it was instead the *Paleolithic* component of the genepool that was less than 30%. They used the Y-chromosome dataset from Semino et al. (2000) but instead of looking for clines, they directly calculated admixture with Near Eastern populations. With an average of 50% across samples, they found European populations closest to the Near East had a Neolithic component of between 85 and 100%, decreasing by France, Germany, and Catalonia to 15-30%.[545] They also identified differences between

---

[545] These results should however be taken with caution, as the descendant population used to represent the Paleolithic genome was the Basques, which was a big assumption. When Sardinia was used as the Paleolithic descendant population, the admixture with Neolithic genes increased to 65%.

Mediterranean and non-Mediterranean populations.

6.2.3.2 A Paradigm Shift Waiting to Happen: the Sequencing of Ancient DNA

It is crucial to bear in mind that all results up until this point had been reached through analysis of modern populations. The problem with this is that the modern population is the result of every migration and all the gene flow that has occurred up until the present moment. The only way to actually determine which genes were Paleolithic, which were Neolithic, and what percent of each made up the European population through time was to actually test ancient samples. Developing genome sequencing technology eventually made this possible.[546] Initially, only ancient mitochondrial DNA could be sequenced, as each cell contains hundreds of copies of the mitochondrial genome as opposed to only one copy of the full nuclear genome.

The very first results were fascinating. Haak et al. (2005) sequenced the mitochondrial genomes of 24 early farmers associated with geographically widespread LBK sites. 18 of the 24 individuals belonged to mtDNA haplogroups that are common in modern Europe, the Near East, and Central Asia. The remaining 6 however belonged to haplogroup N1a, today found in only 0.2% of people in the same area as the individuals sampled. Given how geographically widespread the samples were, it was determined that this represented a 150-fold decrease in the frequency of this haplogroup between the Neolithic and the present day and that this could not have been due to genetic drift. Thus, either agriculture spread in way fitting with the cultural diffusion hypothesis, or European farmers were replaced again later. Given the archaeological evidence, the latter scenario seemed unlikely. Who could the farmers have been replaced with? Based on the arguments proposed thus far, Haak et al. postulated that modern Europeans might genetically be mainly of Paleolithic origin, with the incoming farmers having intermarried with the local hunter gatherer populations to the point of being genetically assimilated into them.

Then however, Malmström et al. (2009) found that, at least in Scandinavia, the hypothesis of genetic continuity between hunter gatherers (in this case the Pitted Ware Culture) and modern Swedes, Norwegians, and even Saami could not be upheld.[547] They were not able to collect enough data on farmers (in this region, the Funnel Beaker Culture), and thus concluded that a population replacement must have occurred during the Neolithic or post-Neolithic periods. The former seemed to be ruled out by the Haak et al. (2005) study, but could this have been due to the regional nature of Malmström’s data? Bramanti et al. (2009) were able to confirm the genetic discontinuity between European hunter-gatherers and LBK farmers. The hunter-gatherers belonged mainly to mitochondrial haplotype U, especially type U5. LBK farmers brought haplotype N1a along with some others, but never types U4 or U5. They discovered that hunter-gatherers

---

[546] For a thrilling overview of the developments in the field, in narrative form, see Pääbo 2014.

[547] It could *not* however be rejected between hunter gatherers and modern Baltic populations, suggesting that the eastern Baltic area remained a genetic refugium for some hunter-gatherer populations.

and farmers, although living side-by-side, did not intermarry, at least for a few thousand years. More importantly, modern European mtDNA diversity cannot be explained by admixture between hunter-gatherers and farmers. There were indeed other, post-Neolithic population turnovers.

Haak et al. (2010) sequenced mitochondrial DNA from 43 LBK farmers, finding amongst them 25 haplotypes. Of those, 11 are still found in high frequency among present-day populations, 10 are found with limited distribution (purportedly near the core of the LBK as well as the Near East), and 4 no longer exist. When compared with some other datasets, previous patterns in haplogroup distribution were confirmed. N1a was present in some 14% of the LBK samples, but none of the hunter-gatherers. Conversely, hunter-gatherer mitochondrial DNA consisted of ca. 80% haplogroups U4 and U5, which were almost non-existent in the LBK individuals. Contrary to Torroni et al. (1998), haplogroups H and V were only common in LBK individuals. Haak et al. (2010) were also able to sequence three LBK Y-chromosomal genomes. Rather than the most common haplogroups today (R1a, R1b, I, and E1b1), they found the males belonged to haplogroups G2a and F, rare today but with slightly higher frequencies in the Near East. They concluded that the modern European population was shaped by *both* Mesolithic re-peopling *and* the LBK expansion, with the later having contributed much more. However, it was also shaped by post-Neolithic events, demonstrated by the 4 unique mitochondrial haplogroups and the 3 rare Y-chromosome haplogroups, perhaps having to do with the movement of Y-chromosome haplogroup R1a.

Haak et al. (2008) had identified Y-chromosome haplogroup R1a1 in Late Neolithic Corded Ware individuals along with diverse mitochondrial haplogroups (K1, U5, I, H, and X). Keyser et al. (2009) found R1a1 at very high frequencies in southern Siberia. Out of a group of 26 sequenceable genomes dating from the Middle Bronze Age Andronovo Culture through the Karasuk, Tagar, and Iron Age Tashtyk cultures, all males save for one (of haplogroup C) were of haplogroup R1a1. As with other analyses, the mitochondrial haplogroups were more diverse: the Bronze Age sites were dominated by western Eurasian haplogroups (HV, H, T, I, U, and K) while the eastern Eurasian haplogroups (Z, G2a, C, F1b, and N9a) increased in frequency in the Iron Age. Working with some data and interpretations by Semino et al. (2000), Keyser et al. (2009) posited that the distribution of Y-chromosomal haplogroup R1a1, with its peak in Poland and a decreasing cline to the Northwest and to the East towards Anatolia and the Caucasus, might represent the re-peopling of Europe from an Eastern European glacial refugium, magnified later by the spread of Kurgan people bearing Indo-European languages. Despite being more diverse, the distribution of mitochondrial DNA haplogroups could not rule out the idea that whole populations, rather than just men, had been migrating. Could this have been a part of the post-Neolithic population changes?

Skoglund et al. (2012) used partial whole genome analysis, and though they only included four individuals (three Pitted Ware hunter-gatherers and one Funnel Beaker farmer) the samples were all from the same geographic region and roughly contemporaneous. They found that, for all European populations, allele-sharing with hunter-gatherers is negatively correlated with allele-sharing with Neolithic farmers. The trend is mainly along a North-South axis with hunter-gatherers sharing most alleles with northern Europeans and the farmers with southeastern Europeans. All in all, the fraction of farmer DNA decreases from a high of ca. 95% in Sardinians through ca. 52% in individuals of northwestern European descent to ca. 11% in Russians. The non-farmer percentage is not fully hunter-gather however. On the contrary, despite allele-sharing, Skoglund et al. (2012) found that the genetic profile of the Scandinavian hunter-gatherers they sequenced is not fully represented in any modern population. By this point, it had become clear that farming had spread, at least in some parts of Europe, via a largescale demic diffusion.

Gamba et al. (2014) provided further evidence for these results, sequencing autosomal genomes from a period of about 5,000 years on the Hungarian plain. This longitudinal analysis showed two clear shifts in genomic affinities on either side of a period of about 2,800 years with little change (during the Neolithic). The first change was due to the incoming farmers, but already the two earliest farmers tested showed intermarriage with hunter-gatherers. In fact, the combination of Mesolithic hunter-gatherer Y-chromosomal haplogroups I2 and C6 with Neolithic farmer mitochondrial haplogroup N1a suggested that hunter-gatherer males were being incorporated into farming societies in Central Europe. This stood in potential contrast to the Mediterranean route, where Lacan et al. (2011) had found Y-chromosomal haplogroups G2a and Elb1b1a1b in Spain. Once widespread, as attested by Haak et al. (2010) finding it in Germany, G2a is now very rare. Lacan et al. (2011) suggested that its high frequency at this southern farming site meant that men had played a greater role in the spread of farming along the Mediterranean route. Sampietro et al. (2007) had also noted a difference between the two farming routes, going as far as to suggest that, since Iberian Neolithic remains were still similar to modern Iberians, farming had perhaps spread via demic diffusion along the Mediterranean route but more via acculturation in Central Europe. In any case, Gamba et al. (2014) found that the Neolithic genomes cluster with affinity to southern Europeans, especially Sardinians. After the period of Neolithic genetic stasis, Bronze Age individuals cluster with modern Central Europeans. This suggested some genetic influence from the North. Finally, the Iron Age individuals clustered between modern Eastern Europeans and individuals from the Caucasus. Now that the archaeologically well-attested migration of Neolithic farmers throughout Europe was being genetically fine-tuned, more attention was being granted to these mysterious post-Neolithic population turn-overs.

Patterson et al. (2012) had found that a portion of European ancestry was shared with Amerindians. The source and time-depth of this added to the mystery. Lazaridis et al. (2014) were able to further contextualize those results. They added seven individuals to the ancient whole genomes that had been sequenced to date and then performed an admixture analysis on modern European and Near Eastern populations to understand how the modern and ancient genomes were related. In a PCA plot, they found that modern European and Near Eastern populations formed two separate North-South clines, bridged by a few Mediterranean populations. The ancient genomes formed several clusters in relation to the modern genomes. Western European hunter-gatherers formed two clusters (Western and Scandinavian hunter-gatherers), beyond present-day Europeans in the direction of European differentiation from the Near East. Early European farmers, including the Tyrolean Ice Man, clustered together. Finally, at the far northern end of the Europeans on the plot clustered a group of Siberian Upper Paleolithic hunter-gatherers, including Mal'ta Boy. Lazaridis et al. (2014) called this cluster Ancient North Eurasian (ANE). Admixture analysis shows that this ANE component is a better representation of the mysterious component of European ancestry than Amerindian ancestry (of which it is also a part, leading to the similarity noticed). The admixture analysis furthermore revealed that nearly all Europeans were genetically a mix of three ancestral populations: 1) Western European hunter-gatherers (WHG) who contributed ancestry to Europeans but not Near Easterners, 2) early European farmers (EEF) mainly of Near Eastern descent but with small amounts of hunter-gatherer ancestry as well, and 3) this ANE Siberian hunter-gatherer component, which contributed to both Europeans and Near Easterners. Haak et al. (2010), mentioned above, had confirmed the role of both hunter-gatherer and farmer ancestry in making up the modern European genome, along with post-Neolithic factors. Now Lazaridis et al. (2014) showed how important these post-Neolithic factors really were. Early European farmer ancestry ranges from 30% in the Baltic region to 90% in the Mediterranean. But the ANE ancestry, absent from both hunter-gatherers and farmers and therefore post-Neolithic, is found in all Europeans at at least 20%. Nearly a quarter of the European genome comes from a *third* ancestral population. This conclusion, resulting from the ability to sequence ancient whole genomes, set the stage for a major paradigm shift. What role did ANE ancestry play in post-Neolithic Europe?

6.2.3.3 The 2015 Paradigm Shift

2015 was an enormously important year for the field of ancient genomics as well as for Indo-European studies. Two major papers, Haak et al. (2015) and Allentoft et al. (2015) appeared in the same issue of *Nature* and provided convincing answers to many of the questions that had surfaced up to that time.

Haak et al. (2015) sequenced whole genomes for 69 ancient individuals. This large sample allowed many trends to become visible. The earliest European farmers from across Europe cluster with present-day Sardinians, suggesting that both the farming expansion across the Mediterranean and the one into Central Europe had similar ancestral origins. Middle Neolithic Europeans are intermediate between the earlier farmers and Western hunter-gatherers, meaning that hunter-gatherer ancestry began to increase again after the farmers first arrived.

In European Russia, Y-chromosome haplogroups R1a and R1b are found in 100% of samples from all periods. Before the Late Neolithic period, outside of Russia, these haplogroups were virtually absent. Then, in late Neolithic and Bronze Age sites, these haplogroups are found in 60% of the male populations, suggesting that they spread from the East in ca. 3000 BCE. Furthermore, Late Neolithic/Bronze Age Europeans plot between the Middle Neolithic farmers and the Yamnaya steppe herders from ca. 3000 BCE, indicating that they are the result of admixture between these two populations.[548] By 2500 BCE, genetic signatures consisting of up to 79% Yamnaya-like ancestry appear in the Corded Ware Culture. As individuals of the Corded Ware Culture are some of the earliest to differentiate from the Middle Neolithic population of Europe and because they also do so most strongly, it appears that Yamnaya people migrated into Europe quite suddenly. This represented, as they titled it, a "massive migration from the steppe". The once-common mitochondrial haplogroup N1a and Y-chromosomal haplogroup G2a virtually disappear at this time and are replaced by mitochondrial haplogroups I, T1, U2, U4, U5a, W, and subgroups of H as well as Y-chromosomal haplogroups R1a and R1b. In fact, the Yamnaya are likely the source of the ANE ancestry component found throughout all Europeans.[549]

Research had focused on the Neolithic expansion since it was archaeologically so clearly a major force of change across Europe. Now though, a second expansion, the full extent of which had been invisible, was brought into the light of day. Current European populations were all the result of a three-way admixture of Western hunter-gatherers, Early Neolithic farmers, and Yamnaya steppe populations. The proposals of Gimbutas, Mallory, Anthony and the linguists since Schrader as to the origin and spread of the Indo-European language family had to be reckoned with in a huge way. Further support continued to shift the paradigm.

Haak et al. (2015) further found that Bell Beaker and Únětice populations had reduced Yamnaya ancestry compared with earlier Corded Ware Culture individuals, suggesting that Early European farmer ancestry began to rise again after the initial incursion of steppe peoples (in the same way that hunter-gatherer ancestry rose again after the initial appearance of farmers). Allentoft et al. (2015) confirmed and expanded these. Individuals from the Afanasievo Culture are genetically indistinguishable from the Yamnaya and demonstrate that the expansion went East as well as West. The later Sintashta, although in geographic proximity to the Afanasievo, are not genetically descended from them. Instead, their mix of Yamnaya and Neolithic farmer ancestry matches the Corded Ware, and requires that a population from Western Europe moved

---

[548] This makes Keyser et al. (2009) seem extremely prescient.

[549] The Yamnaya ancestry signature is composed of admixture between ANE ancestry (from Eastern hunter-gatherers [EHG]) and what Haak et al. (2015) identified as Near Eastern ancestry (cf. Mathieson et al. [2015] who call it Armenian-like Near Eastern but without additional EEF). The latter has been demonstrated by Jones et al. (2015) to most likely be Caucasus hunter-gatherer (CHG) ancestry. Yamnaya, Afanasievo, and Poltavka individuals are homogenous in comprising 48-58% of this component (Mathieson et al. 2015). Lazaridis et al. (2016) find that this component is related to Chalcolithic farmers from Iran.

back across Europe toward the steppe and provided the ancestry of the Sintashta. The Andronovo Culture then represents a temporal and geographic expansion of the Sintashta gene pool. Bronze Age cultures that came after the Andronovo began to introduce East Asian ancestry. At the same time, Mathieson et al. (2015) found that male individuals of the Srubnaya Culture to the West of Andronovo were all of Y-chromosome haplogroup R1a, but that some were of haplotype R1a-Z93. This haplotype is not found among ancient central Europeans, but is found today in Central/South Asians. This suggested a potentially more eastern source for the Early European Farmer ancestry component on the steppe than the Corded Ware.

Further developments saw the refining of our understanding of past populations movements in Eurasia. Each newly sequenced ancient whole genome provided greater resolution. Fu et al. (2016) took a closer look at Paleolithic and Early Neolithic genomics and concluded that population change was a common occurrence in Eurasian history. Surprisingly, they found an R1b male from Villabruna in Italy 14,000 years ago, which, when taken with an R1b farmer from 7,000 years ago in Iberia (cf. Haak 2015), shows that the distribution of Y-haplogroup R1b, as is the case with R1a above, has yet to be fully understood. Their analysis supported the conclusions of Haak et al. (2015) and Allentoft et al. (2015) that ANE ancestry spread with steppe migrations.

Also around this time, the genetic results being published in journals like *Science* and *Nature* were causing concern in archaeological circles, where it was feared that the interpretations were too simplistic, too generalized, and ignored the understanding of human interactions that had been gathered from decades of archaeological inquiry. Vander Linden (2016) writes that “the narrative set forth in several high-profile publications seems as first sight to rise from the darkest depths of culture history.” Other issues include the patchy sampling, particularly in the Mediterranean region and the supra-regional conclusions being made on this potentially unrepresentative data. Individuals are listed as belonging to the Corded Ware Culture or the Bell Beaker Culture, both of which are characterized by their variability over wide-spread geographic areas. He does not reject that the ancient DNA analyses offer a wealth of new information but instead argues for more nuance, more collaboration, and the recognition of the important role of regional differences.

As if in response to these critiques, some subsequent studies addressed regional phenomena and targeted regions with a lack of data. Lipson et al. (2017) sampled Neolithic and Chalcolithic individuals from Hungary, Germany, and Spain in order to better understand the interactions between hunter-gatherers and incoming farmers. They found that each region showed a distinct pattern of interaction between EEFs and hunter-gatherers. WHG ancestry increased over time in all regions, but more slowly in Hungary. The genetic distribution that resulted showed that the admixture occurred locally, rather than suggesting that farmers were quick to admix and then carry that admixture across Europe. After the migrating, settling, and admixing, populations did not seem to be mobile enough to recreate the original homogeneity of the farmers.[550]

Mathieson et al. (2018) helped fill in the picture for Southeastern Europe. The Balkan peninsula was an important genetic transition zone throughout prehistory. Hunter-gatherers from the Iron Gates region show ancestry from both WHG and EHG populations, but admixture analyses show that it is not a perfect fit, suggesting other influences. Amongst them are also some early individuals of entirely Northwestern Anatolian ancestry, which means that region was a zone of interaction. Individuals from the Peloponnese and some associated with the Minoans show more CHG-oriented than WHG-oriented ancestry. Once farming spread to the area, Neolithic populations in modern Bulgaria, Croatia, North Macedonia, Serbia, and Romania are of up to 98% Near Eastern ancestry. The Copper Age sees a rise in hunter-gatherer ancestry, as it does elsewhere in Europe. Mathieson et al. (2018) analyzed this resurgence and concluded that a bias toward male hunter-gatherer ancestry is strong in Iberia and Central Europe at this time, but weak in the Balkans. Some Balkan individuals, including one from a burial at the site of Varna, show steppe ancestry components two millennia before the advent of Corded Ware, but steppe ancestry otherwise remains rare until the Late Bronze Age. Although they suggested it might turn up later, Bronze Age Anatolians have the CHG component but not the EHG component of the steppe signature. This has implications for the time and manner of the spread of the Anatolian Indo-European languages.

Olalde et al. (2018) made an important contribution to the understanding of the Iberian peninsula, where it had been noted that populations did not show steppe ancestry at the time when the rest of Europe had begun to (Mathieson et al. 2015). Outside of Iberia, 84 of 90 Bell Beaker males analyzed belonged to Y-chromosome haplotype R1b-M269. Within Iberia, this haplotype was found in four males with genome-wide steppe ancestry. But otherwise there was a higher proportion of haplogroups like I and G2, known to have been common across Europe during the earlier Neolithic. Steppe ancestry was only present in 8 of the 32 Iberian Bell Beaker individuals sequenced, though it is with the Bell Beaker that steppe ancestry components first appear in Iberia. Across Europe, Bell Beaker-associated individuals are heterogenous, even sometimes within sites, across a cline with one extreme being Yamnaya individuals and the other extreme being Middle Neolithic and Copper Age Europeans. But no Iberia-related ancestry could be discerned in Beaker-complex-associated individuals outside of Iberia. Thus this early spread of the Bell Beaker complex seems to have involved cultural diffusion rather than migrations. This was not the case in the British Isles, where steppe ancestry arrives with the Bell Beaker complex from the mainland, with genetic affinities to Central European individuals. By ca. 2450 BCE, an approximately 90% population turnover had occurred in Britain in autosomal DNA, previously absent Y-chromosomal haplogroup R1b, and previously absent mitochondrial haplogroups I, R1a, and U4. By 2000 BCE, as happened

---

[550] Both the Danubian (LBK) and Mediterranean (Cardial Ware) agricultural waves seem to have originated from one single population of farmers in the Balkans (Haak et al. 2015, Mathieson et al. 2015, Mathieson et al. 2018).

elsewhere, Neolithic-related ancestry modestly increased again. Olalde et al. (2019) further demonstrated the uniqueness of the hunter-gatherer population of Iberia, with a substructuring between northwestern and southeastern hunter-gatherers having existed. It was ca. 2000 BCE that steppe ancestry became entrenched in Iberia through male-biased admixture. Steppe ancestry replaced 40% of the previous ancestry components but nearly 100% of the Y-chromosomes (with R1b-M269). By the Iron Age, steppe ancestry had even spread to the non-Indo-European-speaking Basque.

6.2.4 Summary

The results of ancient DNA analysis provide strong support for the steppe hypothesis of Indo-European origins. Many of the details will continue to be fine-tuned in the future, for example the precise relationship of the R1a and R1b Y-chromosome haplogroups. But it has justified a number of assumptions that relate to the substrate lexicon of Latin:

The Italic language family spread through Europe from the steppe. Its ancestors arrived in a Europe that had been settled for centuries by farmers from Southwest Asia. But they had also admixed to a significant extent with the original inhabitants of Europe: the hunter-gatherers. Thus there is the potential for great substrate language diversity but also the possibility that some aspects could be widely distributed.

6.3 The Italian Peninsula

Genetic results pertaining to the Italian peninsula have been sequestered to a second section of this chapter for two reasons. The first, most obvious reason, is that Italy is the subject of this thesis and it is therefore worth discussing them separately in detail. The second reason is that, in comparison to the rest of the Europe and as lamented by Vander Linden (2016), not nearly as much work has been done on ancient Mediterranean genomes until extremely recently. Population genetic studies with the aim of adding to medical understanding abound in the earliest literature, and a research tradition in line with that of Cavalli-Sforza evolved from that. It serves best to discuss the Italian data after the general stage has been set for Europe, so the particularities of what the data suggest can be most visible. Ancient DNA studies on remains found in Italy are very recent, having been performed even later than for other regions. But more are forthcoming.

6.3.1 Studies on Modern Populations

6.3.1.1 The Earliest Studies

As with population genetics in general, the earliest studies of this sort on the Italian peninsula had begun already in the 1970s, investigating “classical” genetic markers. The purpose was often principally medical and involved understanding the distributions of blood group and protein markers (cf. Bargagna, Domenici, and Morali 1975). Soon however, recurring patterns began to appear. Menozi, Piazza, and Cavalli-Sforza (1978) had already noticed that the distribution of traits in Sardinian populations was closer to that of Middle Eastern populations than anywhere else in the Mediterranean. They attributed this to Phoenician and Punic immigration, assuming that the first farmers would have come from southern Italy. Southern Italy grouped together with southern France, Greece, and Turkey, which they took to be the influence of Greek colonizations.

Zappacosta et al. (1980), when they compared HLA frequencies in Bergamo and Campania, found that frequencies of certain antigens in Campania were similar to the Middle East whereas in Bergamo they parallel the European distribution. Giari, Domenici, and Bargagna (1986) found that the frequency of red cell esterase D (EsD*5) in Tuscany is similar to that expected in Northern and Central Europe, while the frequency of EsD*1 is lower in Italy as a whole than in Europe. Olivetti et al. (1986) found from HLA-A, -B, and -C allelic frequencies throughout Italy that "genetic distances of Sardinia from the rest of Italy are one order of magnitude higher than all the other distances between Italian regions" and that "the Southern regions (with the possible exception of the island of Sicily) are genetically more distant from the Northern and Central regions than from each other." Some alleles showed a North-South gradient.

Piazza et al. (1988) performed an analysis in which they also factored in linguistic and archaeological evidence. They produced a PCA from several blood group gene frequencies, and saw in the first principal component that Northern Italy clusters with Central and northern European countries, while Southern Italy clusters with Greece. The second principal component separated a Northern European cluster from a Mediterranean one (Italy, Greece, France, Spain, Portugal) plus Hungary and Czechoslovakia. As the other studies had done, this multiple allele analysis showed a North-South gradient. Piazza et al. interpreted the genetic similarity of southern Italy to Greece in the same way as Menozzi et al. (1978): possible genetic input from Magna Graecia. Proportional population estimates of the Greeks in Italy led them to believe they should almost certainly see a Greek genetic signature. The rest of their conclusions are based on contemporary understandings of the archaeological and linguistic landscape of Italy (cf. §7.2). They suggested that an area of genetic distinctness in the West could be the influence of Etruscans. Furthermore, they found the population of Liguria to be distinct, and found potential affinities between Oscans and Picenes. They concluded from all of this that the Italic languages were most likely brought by small groups from the area of the Danube, either over the Alps or from the Adriatic, in two waves of first cremating and then inhuming Italici. Citing the lack of any Proto-Messapic archaeological traces along the peninsula, they proposed that Messapic could have been brought to Apulia from Illyria across the Adriatic. This analysis was the first to incorporate genetic lines of evidence into the arguments that had been built using archaeology and linguistics.

6.3.1.2 More Modern Methods

Fascinatingly, the major trends that had been observed from classical/protein markers were *not* overturned, but rather confirmed by DNA testing. Barbujani et al. (1995) found that around one quarter of the mitochondrial haplogroups they identified (12 of 42) across Italy and Sardinia showed substantial variation in frequency among populations. Some of these took the form of North-South gradients, with Sardinia being quite distinct. They took these patterns to be the result of a rapid demographic expansion, but they calculated a date of 8,200 - 20,525 years ago for the expansion. They could not decide if this was from the Neolithic spread of farming or from population growth after the Last Glacial Maximum. In any case however, Sardinia produced other results, suggesting slow if any growth and then isolation. Semino et al. (2000) found that Sardinian Y-chromosomes cluster with Greece and Albania, but not Basques. Chikhi et al. (2002) concluded that Sardinia is an outlier with a significantly higher proportion of Paleolithic genes, and that its pattern represented genetic drift from the Paleolithic rather than a Neolithic immigration. These conclusions fit into the debate at the time that was centered around the true contribution of the Neolithic migrants.

Di Giacomo et al. (2003) examined the distributions of 9 Y-chromosome haplogroups from 30 populations in Italy and Greece, finding major differences between them. In Italy, only 3 haplogroups have frequencies above 10% while in Greece, the three most common haplogroups account for only 55% of chromosomes. The only haplogroup to show a significant distribution was P*(xR1a)[551] in Italy, decreasing from Northwest to Southeast. Because this haplogroup appeared in Basque and Celtic populations, they thought it was a Paleolithic group. But recall that Olalde et al. (2018) would show that this spread to Basque groups relatively late, by the time of the Iron Age. Capelli et al. (2006) sampled more markers with a larger sample size and found that in fact 70% of the Y-chromosome diversity of Italy is structured along a North-South axis. 80% of chromosomes were in haplogroups R1*(xR1a1)(itself 40%), J2 (20%), G (11%), and E3b1 (10%). R1 frequencies decrease from North to South, while J2 and E3b1 frequencies increase in from North to South. Like Di Giacomo et al. (2003), Capelli et al. (2006) assumed Iberian populations were a good proxy for Paleolithic ancestry. Thus when they tested admixture using Iberians to represent Paleolithic Europeans and Anatolians to represent Neolithic farmers, they interpreted the 70-90% Anatolian contribution in the South and 50-70% Anatolian contribution in the North as well as the decreasing frequencies of R1 from North to South to represent Mesolithic-Neolithic contact effects. What they were truly seeing however was steppe influence descending in decreasing frequency from the North.

Di Gaetano et al. (2012) were able to replicate the patterns of 1) Northern and Central Italy clustering separately from Southern Italy, the former with the rest of Europe and the latter with the Middle East and 2) Sardinia clustering separately from all groups using genome-wide data. Keller et al. (2012) found that the Neolithic Ice Man of the Öztal Alps clusters with modern Sardinian populations, suggesting that Sardinia's signature

---

[551] Haplogroup R (to which R1a and R1b belong) is a subclade of P. The *x* indicates that, though the exact subclade was not determined, it is not R1a.

was not Paleolithic, but rather Neolithic. Boattini et al. (2013) looked again at uniparental markers, finding that 62% of modern Italian Y-chromosomes fit into 8 haplogroups, including 4 R1b lineages. Three R1b haplogroups and G-P15 (an EEF haplogroup) were mainly responsible for the Northwest-Southeast cline, with the Southeast showing more intra-population variation and the Northwest showing more inter-population variation. Sardinia differed in having the lowest Y-chromosome haplogroup diversity, and had I-M26, which was virtually absent from the peninsula. In contrast to the Y-chromosome picture and suggesting sex-biased population movements, mitochondrial DNA haplogroups are much more homogenously distributed. De Fanti et al. (2015a) determined that the Italian mitochondrial HV* lineages split very early, and some are unique to Italy, suggesting an ancient local presence perhaps from the time of a glacial refugium in Southern Italy. Sarno et al. (2014) agreed with these results, emphasizing that NW Italy clusters Iberia and Central Europe while SE Italy (including Sicily) clusters with the Balkans and the Levant, and that males seem to have been the mobile ones. The NW to SE gradient was again replicated by De Fanti (2015b), this time with the genes for lactase persistence (North Italy was intermediate between high frequencies of lactose intolerance in Southern Italy to the highest level of lactase persistence in Finland, showing that European populations appear more homogenous than those of the Italian peninsula), and by Fiorito et al. (2016) again with whole genome analysis. Grugni et al. (2018) refined the Y-chromosome picture, finding that in Northern Italy, 69% of haplogroups were R1b (specifically R1b-U152) with no other haplogroup reaching 10%. In the South, only 27.5% were R1b, representing a demographic expansion from North-West and Central-North Europe that greatly affected Northern Italy.[552] This was the traces of steppe ancestry having entered Italy.

6.3.2 Ancient DNA

Ancient DNA analyses with relevance to Italy took considerable time to appear, and there is still a great need for more results in order to understand how and when the patterns found in the modern population data came to be. Mathieson et al. (2018) confirmed the presence of Western hunter-gatherers in Sicily between 12000 and 6100 BCE. Emery (2017) and Emery et al. (2018) confirmed that mitochondrial genomes from Iron Age and Roman period South Italy are comparable to elsewhere in Europe.[553]

---

[552] They also noted, following a paper by Günther et al. 2015, that Basques may not be the “living fossils” from the Paleolithic they have long been considered to be. Instead they might represent a long-lasting isolated population that originated from admixture of local HGs and early farmers. Olalde et al. (2018) would come out in that same year, showing the importance of ancient DNA in understanding population history. The presence of R1b in Basque populations does not date to the Neolithic.

[553] In fact, in Emery (2017) when pre-Late Glacial Maximum and pre-Neolithic mitochondrial genomes were removed from comparison, the Iapygian samples shared similar haplogroup compositions with Armenian Iron Age populations. Emery concluded that, if the results are not due to Late Glacial Maximum gene flow, the Iron Age southern Italians descended from Neolithic farmers (from Anatolia and possibly as far East as the Caucasus) and from migrants arriving from eastern Europe in the Late Neolithic or Early Bronze Age (consistent with the ancestors of the Iapygians arriving in Southern Italy from the Balkans, or at least sharing a source population with them). Emery et al. (2018) show that

Serventi et al. (2018) got the same result for Novilara, the area where North Picene is found, showing that Iron Age Picene populations show similar mitochondrial profiles to modern populations from the same region. Antonio et al. (2019) contextualized the results by performing an analysis on whole genomes from Rome, spanning from the Mesolithic through the Iron Age. Mesolithic individuals from Sicily and the Italian mainland indeed clustered with Western hunger-gatherers. Between 7000 and 6000 BCE, agriculture arrived with the appearance of individuals clustering with Anatolian farmers. Interestingly, they carried a small amount of a component lacking in central European and Iberian individuals and found in high quantities in Iranian farmers and Caucasus hunter-gatherers. This suggests a different or at least additional source population from that involved central and western Europe was involved in Italy's Neolithic transition. Then, just like elsewhere in Europe, hunter-gatherer ancestry began to rise again starting in the Late Neolithic.

Allentoft et al. (2015) had determined the Copper Age Remedello Culture in Northern Italy (ca. 3400-2400 BCE, cf. De Marinis & Pedrotti 1997: 298) did not show steppe ancestry, suggesting a potential *terminus post quem* for the arrival of steppe-derived populations in Italy (though they also note that it may have in some way been unaffected by the Yamnaya expansion). Antonio et al. (2019) were able to push this date to ca. 2900 BCE. A gap in their samples due to cremation meant that they had no data between 2900 BCE and 900 BCE, at which point steppe ancestry is already present (with individuals exhibiting highly variable ancestries, suggesting several sources of migration). Olalde et al. (2018) had sampled two individuals in a Bell Beaker context in Northern Italy (Parma) that showed components of steppe ancestry, dating to between 2200 and 1930 BCE. Saupe et al. (2021) were able to fill in the gaps, with samples dating between 3200 and 1500 BCE. They found the first traces of steppe ancestry in individuals from North Italy around 2000 BCE (a Bell Beaker individual from 2195-1940 calBCE as opposed to two without steppe ancestry, an individual from Broion from 1952-1752 calBCE, and even one of the Remedello individuals described by Allentoft et al. [2015], from 2134-1773 calBCE), with the amount increasing with time. The first traces of steppe ancestry in Central Italy were found four centuries later.[554]

Sarno et al. (2017) tested modern Italian populations, but they viewed the results as a PCA overlain on other modern and ancient sequences, allowing them to provide one of the most-up-to-date interpretations of the trends that had been repeatedly found. They found that modern southern Italian and southern Balkan populations plotted in "an almost uninterrupted bridge between two parallel clines of distribution where most of the other modern populations are found, one stretching along the East-West axis of Europe

---

pre-Roman Iron Age individuals from Vagnari cluster slightly differently than Roman-period individuals from Vagnari, possibly representing the result of Roman subjugation.

[554] Steppe ancestry appears in Sicily ca. 2200 BCE, but seems to arrive from Iberia (Fernandes et al. 2020).

and the other from the Near East to the Caucasus." Thus, Sicily, southern Italy, and Mediterranean regions stretching East to the Anatolian Greek islands formed a Mediterranean genetic continuum, with a Sardinian (Neolithic-like) ancestry component making up half of their ancestry. Near Eastern-like ancestry is more frequent in Sicily, southern Italy, and the Greek islands, whereas a European-like component appears in increasing frequencies in the populations of Albania, mainland Greece, and the rest of the Balkan peninsula. In contrast to continental Europe, Caucasus-related admixture via Yamnaya is present to a lower degree in the Mediterranean groups, detected mainly in Balkan-related groups. They concluded that "any significant Steppe/northern component may have arrived in the south Balkan mainland and southern Italy only later, by which time Indo-European languages of the Italic, Greek and various Balkan branches had already established themselves there." This suggests that the spread of Italic languages into Southern Italy would have involved a certain amount of indirect diffusion and a large amount of contact with pre-IE languages.[555]

6.3.3 Other Questions

6.3.3.1 Greek Colonization

Sarno et al. (2015) had suggested that remnants of the Greek settlement of Southern Italy might be represented by Y-chromosome haplogroup E-V13. Tofanelli et al. (2016) noted that no Italian populations seem to show a closer affinity with Greek and Greek-related sources from a Y-chromosome perspective, but designed an approach that took mutational processes into account. In this way, they purported to find evidence of Greek colonization from Euboea in East Sicily in Y-chromosome and mitochondrial DNA from modern populations. They found a sex bias, and evidence for low numbers: probably thousands of men and a few hundred women. Interestingly, they found that E-V13 was actually a very poor indicator of Greek origin. Later, Sarno et al. (2017) found that the modern Greek-speaking communities of Apulia and Calabria show no clear signs of a recent continental Greek origin, instead clustering within the Mediterranean continuum they found. This suggests that either a large degree of admixture has occurred, that these Greek-speaking communities date back to antiquity (as opposed to the Middle Ages), or perhaps both. Grugni et al. (2018) also found potential genetic traces of the Greek presence in Italy in form of Y-chromosome haplogroup R1b-M412*, which is found in Turkey, Iran, Cyprus, and Greece along with all of their Southern Italian samples but which is only sporadically found in North Italian samples. They suggested this might date to the colonization of Magna Graecia (especially from Greek islands) or perhaps to the Byzantine era. They also found R1a-M17, with the highest frequencies in Italy being in the South and otherwise found in significant amounts in mainland Greece and Thrace. This might represent traces of Balkan migrations into Southeast Italy.

---

[555] Fernandes et al. (2020) found in Sicily Iranian-related ancestry consistent with a shift towards Minoans and Mycenaeans by the Middle Bronze Age. Raveane et al. (2019) had found this signature in modern southern Italian populations as well.

6.3.3.2 Etruscans

The question of Etruscan origins, whether they are from Anatolia or autochthonous to Italy, has also received attention from geneticists. At a remarkably early date, Vernesi et al. (2004) sequenced the mitochondrial genomes of 30 Etruscans, finding little heterogeneity between sites or time periods sampled, confirming that the Etruscans were a unitary population. But they were unsure how to further interpret the data, because most of the haplotypes they found no longer existed in modern populations. Some evidence pointed to their genomes being typical for Europe or West Asia, but they also found similarities with Turkish genepools. Achilli et al. (2007) found that the mitochondrial haplogroups of the modern population of Murlo, a town of Etruscan origin, are composed of 17.5% of Near Eastern haplogroups. No other Tuscan populations had this high of an amount, but overall 5% of mitochondrial haplotypes in Tuscany were found to be shared exclusively with Near Easterners, supporting a direct and recent input from the Near East.

Ghirotto et al. (2013) found from mitochondrial DNA that the Etruscan individuals they sequenced were the ancestors of modern populations in Casentino and Volterra (but, interestingly, not of the more general population of the formerly Etruscan territory) and that the genetic links between Tuscany and Anatolia dated back to at least 5,000 years ago. Thus the Etruscan civilization likely developed locally. Tassi et al. (2013) repeated the conclusion that the genetic links between Tuscany and Anatolia were too old to have brought the Etruscans, instead possibly dating to the spread of Neolithic farmers. But the analysis of results would continue to go back and forth.

Gómez-Carballa et al. (2015) concluded that the Near Eastern components in Tuscan mitochondrial genomes was 8% and in autosomal DNA was 21%. As the Near Eastern mitochondrial haplogroups do not show local or regional variation, it suggests a recent arrival from the Near East with no subsequent founder effects or genetic bottlenecks. Thus the Etruscans were probably from Anatolia. Fiorito et al. (2016) used an admixture analysis on whole genome data that suggested an event ca. 3000 years ago between populations from the Caucasus and Middle East and populations from Central Italy. Pardo-Seco et al. (2014) had calculated a similarly recent data for an admixture event. This might be the arrival of the Etruscans from Anatolia. Grugni et al. (2018) found that the Y-chromosome haplogroup J2a-M67* had high microsatellite variation specifically in Volterra. Given that its origin might be close to the Middle East, they took this to support the Asia Minor origin of the Etruscans.

More recent ancient DNA studies call into question the possibility of this 3000-year-old admixture event. Antonio et al. (2019) in their small sample size found no significant difference between Latins and Etruscans in their allele-sharing with other populations. Most recently, Posth et al. (2021) sequenced individuals dating back to 800 BCE. They confirmed that the Etruscan genetic profile was the same as that of its Latin neighbors (including a large proportion steppe ancestry), without any recent population influx from the Near East. The Etruscan language seems to have been preserved despite large amounts of admixture, similar to the situation of Basque. It is unclear if an elite dominance situation like that responsible for the spread of Hungarian could instead be responsible.

6.4 Conclusions

There are several important conclusions that we can draw from the picture of Italian population genetics. Firstly, the genetic history of the Italian peninsula is one of a melting pot. Steppe-derived populations begin to appear in Northern Italy around 2000 BCE, their presence increasing gradually. Steppe ancestry reaches Central Italy by around 1600 BCE. Modern genetic profiles show that the spread of steppe-derived populations into the South of Italy involved large amounts of admixture with previous populations. Complete admixture seems to have occurred with the ancestors of the Etruscans as well. All signs point to heavy contact with speakers of pre-Indo-European languages. The possibility of non-IE language communities, especially in the South, existing up until a late date is certainly not ruled out.

Genetic studies on the population history of the Mediterranean continue to come out, and a refined understanding of population movements will certainly provide clues about the linguistic diversity of Italy. How different would the language of the Cardial Ware farmers of the Mediterranean coast be from that of those that established Linearbandkeramik in the North? And does the Mediterranean continuum help explain the existence of a Mediterranean substrate? What about the traces of Iranian-related ancestry (consistent with Minoans and Mycenaeans) in modern southern Italians? We may someday be able to postulate what other languages, perhaps attested ones, might have played a role in contact situations.