Genetic characterization of two North Italian villages: A story of isolation, ancient admixture, and genetic drift

Massimo Mezzavilla[1,2] | Massimiliano Cocca[1] | Pierpaolo Maisano Delser[3] | Margherita Francescatto[4] | Paola de Gemmis[5] | Daniela Segat[5] | Paola Cattelan[5] | Marina Da Meda[5] | Luigi Magnini[6] | Cinzia Bettineschi[7] | Armando de Guio[8] | Paolo Gasparini[1,4]

[1]Institute for Maternal and Child Health IRCCS Burlo Garofolo, Trieste, Italy

[2]Department of Biology, University of Padua, Padua, Italy

[3]Department of Zoology, University of Cambridge, Cambridge, UK

[4]Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, Italy

[5]Mauro Baschirotto Institute for Rare Diseases. B.I.R.D. Foundation, Longare, Italy

[6]Department of History, Human Sciences and Education, University of Sassari, Sassari, Italy

[7]Faculty of Philology and History, University of Augsburg, Augsburg, Germany

[8]Department of Cultural Heritage: Archaeology and History of Art, Cinema and Music, University of Padova, Padua, Italy

Correspondence
Massimo Mezzavilla, Institute for Maternal and Child Health IRCCS Burlo Garofolo, Trieste, Italy.
Email: mzzvilla@gmail.com

Funding information
IRCCS Burlo Garofolo of Trieste

Abstract

Objectives: Population isolates represent a focus of interest because of their particular genetic history and the possibility of mapping peculiar deleterious variants. Here we investigated the pattern of genetic variation in two North Italian villages: Rotzo (RTZ) and Stoccareddo (STC), in the Asiago plateau.

Materials and methods: We genotyped over 800 individuals for more than 600.000 markers. We investigated the isolation level by analyzing runs of homozygosity (ROH) and the level of population structure. Then we estimated the time of admixture and the relationship between ancient genomes and these two villages. Finally, we looked at the effect of genetic drift on deleterious variants.

Results: We highlighted a different isolation level between RTZ and STC; despite the average number of ROH being similar between the two villages, RTZ shows a higher level of total homozygosity. We estimated, from different sources, that the time of admixture for the ancestors of these two populations was between 113 and 88 generations ago.

We discovered that a deleterious variant in *MCUB* gene (rs78025076), which is linked to several lipid traits, is entirely absent in RTZ and at 1%frequency in STC. In contrast, the risk allele frequency is 2% in Europe and 2.4% in North-East Italy.

Discussion: These results show the importance of a genetic characterization of geographically isolated populations. Their vast array of past history could highlight specific events in the past and help describe deleterious variants and traits distribution in different regions.

KEYWORDS
admixture, asiago plateau, genetic drift, genetic isolates, lipid traits, runs of homozygosity

1 | INTRODUCTION

The Asiago plateau is a geographic, linguistic, and cultural enclave located in Veneto, in the north-eastern part of Italy. This region harbors different isolated populations characterized by complex and puzzling history (Colonna et al., 2013; Esko et al., 2013; Sarno et al., 2021). According to the current archeological evidence, Paleolithic hunter-gatherers frequented the area since the Late Epigravettian phase (approx. 16,500–12,000 cal. BP) (Cristiani et al., 2012). The first stable occupation of the Asiago plateau can be referred to the

15th to 12th century BCE in the sites of Longalaita (Rotzo [RTZ] municipality) and Monte Corgnon (Lusiana municipality) (De Guio, 1994). During the first millennium BCE, the material culture can be referred to the Magré group, which is characterized by mixed influences from the Alpine world (i.e., Raetic people, who inhabited an area broadly corresponding to north-eastern Italy, Tyrol, western Switzerland, and southern Bavaria) and the Po plain (ancient Veneti). The main geographical site of this phase is the Bostel settlement (in RTZ municipality); the Monte Corgnon site still existed, possibly functioning as a boundary, ritual area of Este culture, while the Longalita hillfort had been abandoned well before the end of the Bronze Age (De Guio, 1994; De Guio et al., 2011; Magnini et al., 2019).

The Roman conquest is only indirectly testified by the relatively limited Republican and Imperial coins recovered around the slopes on the edges of the plateau. A deafening absence of archeological and written sources suggests a residual human presence in the area up to the 10th to 11th century CE. Only a series of mythopoetic accounts of local scholars of the 16th to 18th century CE advocate the existence of either residual tribes or colonies, and small communities of Raeti, Cimbrains, Goth, Alemanni, Bavarians, Langobards, or generic Germans survived or arrived during the 1st millennium CE, and a complete outline can be found in (Sartori, 1956). On the other hand, the current historical narrative is inclined to accept the idea of a progressive arrival of German groups between 1200 and 1300 CE (Bortolami & Barbierato, 2009). That is echoed nowadays in the long-lasting but gradually disappearing use of Cimbrian, a local German dialect of south Bavarian or Tyrolean origin. (Coluzzi, 2005).

The cultural and political autonomy of the Asiago plateau is well reflected in the foundation in 1311 of the "Federazione dei Sette Comuni" (Confederation of the Seven Communes), with an early start as *League of Seven Sister Lands* in 1259. The Federazione was an independent federal government characterized by the collective ownership of the land. In 1405, a spontaneous dedication to the Republic of Venice guaranteed its privileges for 400 years until Napoleon abolished it in 1807 (Cacciavillani, 1994). In the present-day, in this region, the two villages of RTZ and Stoccareddo (STC), could represent an exciting reservoir of genetic information that could highlight specific elements of the history of the region. Genetic features such as inbreeding, runs of homozygosity (ROH) and distribution of deleterious variants are interesting aspects to study. As we know that they show a significant degree of variation between putative isolated populations (Anagnostou et al., 2017) (Xue et al., 2017), but they could also be implicated in disease and trait distribution in a population (Clark et al., 2019; Johnson et al., 2018).

This research is also intended to provide further hints for unlocking the intricate population dynamics of the Asiago plateau. Offering a reference background to tentatively link language, material culture, and genetic characterization toward a clearer archeological contextualization of the protohistoric sites currently under investigation. This, in turn, will offer solid data for reconstructing and interpreting the history of the ancient and modern inhabitants of the area, and their relevant interconnections.

Summing up, in this work, we tried to characterize these two villages using genetic data coming from several individuals sampled and originally coming from the region. We tried to answer the following questions: (1) What can we infer from the actual genetic data regarding their history? (2) What are the effects of the genetic drift in these populations?

2 | MATERIAL AND METHODS

2.1 | Data preparation and quality control

The Mauro Baschirotto Institute for Rare Diseases, BIRD onlus, as part of a project to characterize genetic isolates from the Asiago highland, enrolled 702 adult volunteers from the STC isolated village and 282 adult volunteers from the RTZ (RTZ) isolated village. (Figure 1). The study was conducted according to the protocol of the Oviedo Convention on Human Rights and Biomedicine. Written informed consent was obtained from all subjects enrolled in the project. The study was evaluated and approved by the Scientific Commission of the Regione Veneto, Giunta Regionale-Ricerca Sanitaria Finalizzata n.176/04, Venice, Italy. It was also funded and approved by the Italian Ministry of Health. The Italian Network of Genetic Isolates project was approved by the Ethical Committee of the IRCCS Burlo-Garofolo.

Each individual underwent a thorough medical examination, including a general visit, the assessment of the main anthropometric parameters (height, weight, waist, and hip circumference), and several lipid traits, including Triglycerides, Total cholesterol mg/dL, HDL cholesterol mg/dl, LDL cholesterol mg/dL. After at least 8 h of fasting, the blood sample was taken by taking 3 EDTA blood tubes: 2 EDTA blood tubes were immediately stored at 4°C. According to the Bioethics Regulations of the European Union and Italy. Written informed consent was required from all participants in this project.

According to the manufacturer's protocol, genomic DNA was isolated from the peripheral venous blood biological sample in EDTA using commercial NucleoSpin Blood Kit (Macherey-Nagel GmbH & Co., Düren Germany). The extracted genomic DNA was quantified using the Nanovue Plus spectrophotometer (GE Healthcare Biosciences - AB) and diluted to 50 ng/ml for subsequent experimental procedures. The DNA quality was evaluated by checking the 260/280 ratio and by the agarose gel electrophoresis run of every single sample.

Samples that did not have an optimal DNA concentration ranged ≥50 ng/ml or whose 260/280 ratio did not have the required range (1.7–1.9) were re-extracted. The blood samples and the extracted genomic DNA were stored at −20°C. All samples were genotyped with the Illumina Infinium Global Screening Array-24 v1 bead chip following manufacturer instructions (Illumina Inc., San Diego, CA, USA). An internal quality control pipeline processed the successfully genotyped data ($N = 889$ samples). Briefly, the pipeline updates SNP ID and strand information, performs a recall step on rare variants (Goldstein et al., 2012) and applies several filters on both variants and individuals. In particular, variants with genotyping missing rate, >1%,

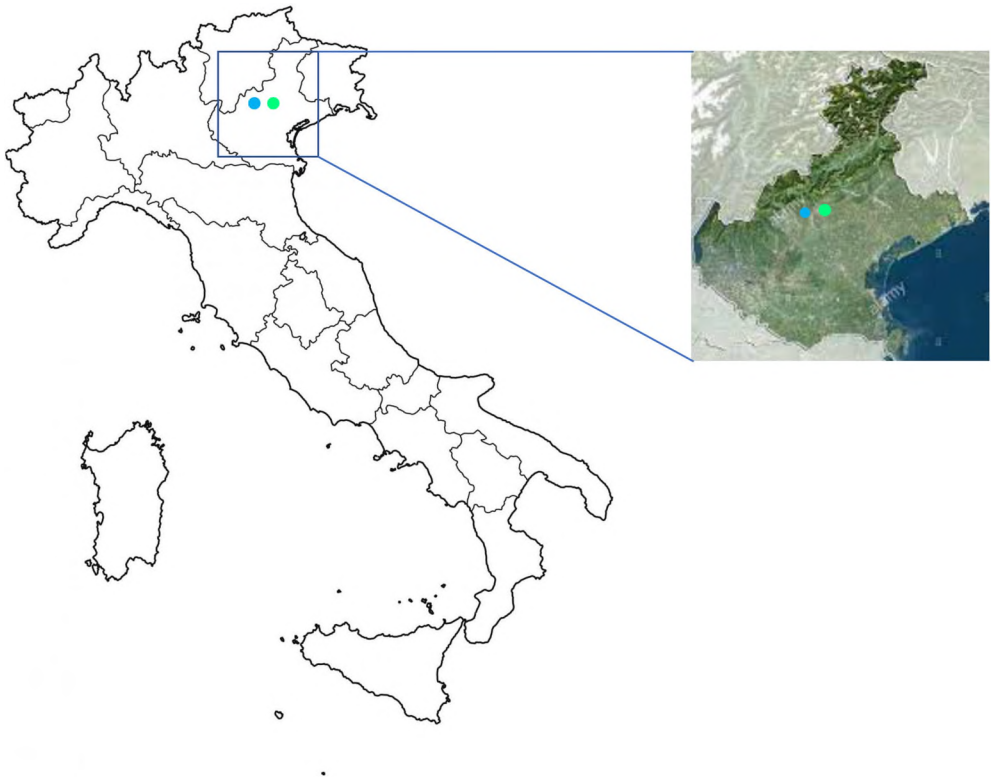

FIGURE 1 Geographical location of Rotzo and Stoccareddo villages. The blue dot represents Rotzo (RTZ) the green one Stoccareddo (STC)

and Hardy–Weinberg equilibrium *p*-value >10–6 were removed. In addition, individuals with genotyping missing rates above 5% and excessive heterozygosity rates (mean +/− 6 standard deviation) were removed. Eight hundred fifty-six samples and 672,947 markers were left for downstream analyses at the end of the QC procedure.

After the QC process, additional highly related individuals (PI_HAT value over 0.25) were removed using the function–genome implemented in PLINK v.1.9 (Purcell et al., 2007), resulting in 163 individuals from STC and 94 individuals from RTZ used for the analyses.

For the analysis using modern genomes, we merged our dataset with the following data: 1000G (Consortium, 2015), HGDP whole genome sequences (Bergström et al., 2020), and INGI isolates (Cocca et al., 2020), and for ancient DNA analyses we merged our dataset with the Reference samples from https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data (v42.4).

2.2 | Population structure and admixture

We extracted 54 unrelated males from RTZ, and 73 unrelated males from STC and Y chromosome haplogroups were assigned using yhaplo (Poznik, 2016) and mitochondrial haplogroup were estimated using HaploGrep 2 (Weissensteiner et al., 2016). Haplotype diversity were estimated using the formula showed in (Nei & Tajima, 1981). Principal component analysis (PCA) was performed on both male and females on autosomal data after removing markers in strong linkage disequilibrium (r2 > 0.4), using the command–indep-pairwise 200 50 0.4 implemented in PLINK v1.9 and with minor allele frequency minor of 0.02. PCA was carried out on worldwide scale and also using only European populations after subsampling 20 random individuals from RTZ and STC.

We used PLINK v1.9 to identify ROH and estimate inbreeding coefficients. The following parameters were used: homozygous-snp 50, homozygous-density 50 homozygous-gap 1000 homozygous-window-snp 5 homozygous-window-het 1 homozygous-window-missing 5 homozygous-window-threshold 0.05 as described in (Ceballos et al., 2018). Differences in total homozygosity level between populations were verified using the Mann–Whitney test.

Then we estimated ROHs with length between 1 and 8 Mb and ROHs higher than 8 Mb. We further investigated the level of genetic drift using IBDseq (Browning & Browning, 2013) and IBDNe (Browning & Browning, 2015) on RTZ and STC in order to estimate the effective population size (Ne) trough time and comparing their Ne with Basque, Sardinian and French populations from HGDP. We using a threshold of 4 centimorgan for IBD segments and default parameters as suggested for SNP array data (Browning & Browning, 2015).

Unsupervised admixture analysis using ADMIXTURE v1.23 (Alexander et al., 2009) was done after subsampling 20 random, unrelated individuals from RTZ and STC. We then explored the genetic similarity between other Europeans populations of HGDP, estimating a population tree using Treemix v.1.1 (Pickrell & Pritchard, 2012). We estimated the standard errors (SEs) using blocks of 500 SNPs. We generated 100 bootstrap replicates by resampling blocks of 500 SNPs to assess the stability of the tree topology. We performed Treemix without adding any migration edge and we coupled the results with the residual plot. Positive residuals indicate pairs of populations where the model underestimates the observed covariance, so the model might improve adding migration edges.

2.3 | Recent and ancient admixture

The admixture time with all possible reference populations combinations was estimated with MALDER (Pickrell et al., 2014) using the default parameters, we tested all possible pairs of reference populations present in the HGDP dataset. In addition, we used fastGLOBETROTTER (Hellenthal et al., 2014) to extrapolate and date admixture events from different populations into RTZ and STC. This approach uses output from ChromoPainter (Lawson et al., 2012) to estimate the pairwise likelihood of being painted by any two surrogate populations at a variety of genetic distances to generate coancestry curves. Genetic data were phased using Eagle v2.4.1 (Loh, Danecek, et al., 2016; Loh, Palamara et al., 2016). We estimated the initial values for switch (−n) and emission (−M) rates with ChromoPainter applying the parameters suggested in the software manual (−f < popsurr>1 10 -s 0 -i 10 -in -iM), obtaining n = 325.278 and M = 0.000381.

In order to create the copy vector file, we run ChromoPainter with the following parameter: -f < popsurr>0 0 -s 0 -n 325.278 -M 0.000381. Last step was to generate the painting sample files with ChromoPainter and the following parameters: -f < popsurr>0 0 -s 10 -n 325.278 -M 0.000381. Finally, fastGLOBETROTTER was run using all the data generated by ChromoPainter as input. Finally, we investigated the relationship between ancient genomes and our two isolated populations, merging our dataset with ancient genomes.

We analyzed the admixture pattern using ancient references using the f3 statistics (Patterson et al., 2012) in the form f3(Source Ancient1, Source Ancient2, Target Isolate); we collected only the results with Z score<= −3. We further investigated the proportions of admixture between our populations (RTZ and STC) and different sources estimated from f3 statistics using qpAdm (Patterson et al., 2012), as outgroup populations, we used: Yoruba, Mbuti, Surui, and Papuan. Furthermore, we investigate the direction of gene flow from different ancient sources to our isolated populations and close related modern population, using the f4 statistics (Patterson et al., 2012; Reich et al., 2009) in the form f4 (Mbuti, X, Isolate, French) and f4 (Mbuti, X, Isolate, North_Italy), where X represents a possible ancient source of gene flow. Finally, we described the relative level of shared drift between European and Italian populations with North and South European references (Orcadian and Sicilian, respectively). We used the outgroup f3 statistics in the form f3(Mbuti; North/South European reference, Target European population), higher the f3 value, higher is the amount of shared drift between the North or South European reference and the target European population.

2.4 | Analysis of genetic drift on low-frequency variants

We analyzed the effects of genetic drift on low-frequency variants. We estimated the pairwise Fst (Weir & Cockerham, 1984) between each isolated village and their closest outbred population (obtained from Treemix analysis) with the largest sample size.

Then we selected the markers that fall over the 95th percentile of the genomic distribution and are present in the GWAS catalog (associated with HDL and LDL traits). Finally; we prioritized variants at low frequency (less than 5%) in the isolated population under study.

2.5 | Phenotypes linked to lipid metabolism

We collected information on sex, age, and lipid traits for the individuals belonging to STC and RTZ. The LDL values were calculated using the following formula: LDL = Total Cholesterol − (HDL + Tryglicerids/5). We then compared the total cholesterol level, triglycerides, LDL, and HDL with other isolated populations from North-East Italy (Cocca et al., 2020), using the Mann–Whitney test. Finally, we performed linear regression analysis on the lipid traits using sex, age, and populations of origin as covariates.

3 | RESULTS

3.1 | Population structure and admixture

The PCA shows that our two isolated populations of RTZ and STC group with other European references from HGDP (Figure 2a), and analysis focused only on Europeans shows that they cluster between North Italians and Adygei (Figure 2b). Admixture analysis reached the lowest cross-validation at K = 9. The major component in RTZ and STC is the one found at a high frequency in the Basque (orange); we also found a reduced amount (∼7%, ∼9%) of the cyan component (Near Eastern component) when compared to the Tuscan from HGDP (∼23%) which is also significantly different between our isolates and the Tuscans (Mann Witney *p*-value <0.05). In contrast, the navy component (Central South Asia) is similar to other Italian reference populations (Figure 3). The Treemix analysis groups STC and RTZ with other European populations, in particular with the earliest split with French and Bergamo (North Italy), thus confirming the ancestral pattern of North-West Europe in RTZ and STC, analysis of residuals did not show any high-positive values for RTZ and STC, which further supports their status of isolate due to reduced gene flow. (Supplementary Figure 1a,b).

Despite RTZ and STC showing similar patterns regarding the ancestry using autosomal markers, we found a contrasting pattern regarding Y haplogroups; in particular, in STC, we found a high frequency of haplogroup J2a1, which is the dominant one (77%), whereas, in RTZ, the frequency of haplogroup J2a1 reaches 20%, and the other major haplogroup is R1b1a2a1a2b (32%). (Supplementary Table 1). The haplotype diversity of the Y chromosome is also different: in RTZ reaches a value of 0.84, instead of in STC, it reaches a value of 0.4. On the other hand, analysis of mitochondrial haplotype diversity showed that RTZ have a haplotype diversity of 0.85 where STC only of 0.6. The ratio of mitochondrial on Y haplotype diversity is ∼1 for RTZ but is ∼1.5 for STC. (Supplementary Figure 2).

The differences between these isolates are also evident in the analysis of ROH. Using two sets of ROH (1-8 Mb and over 8 Mb) we

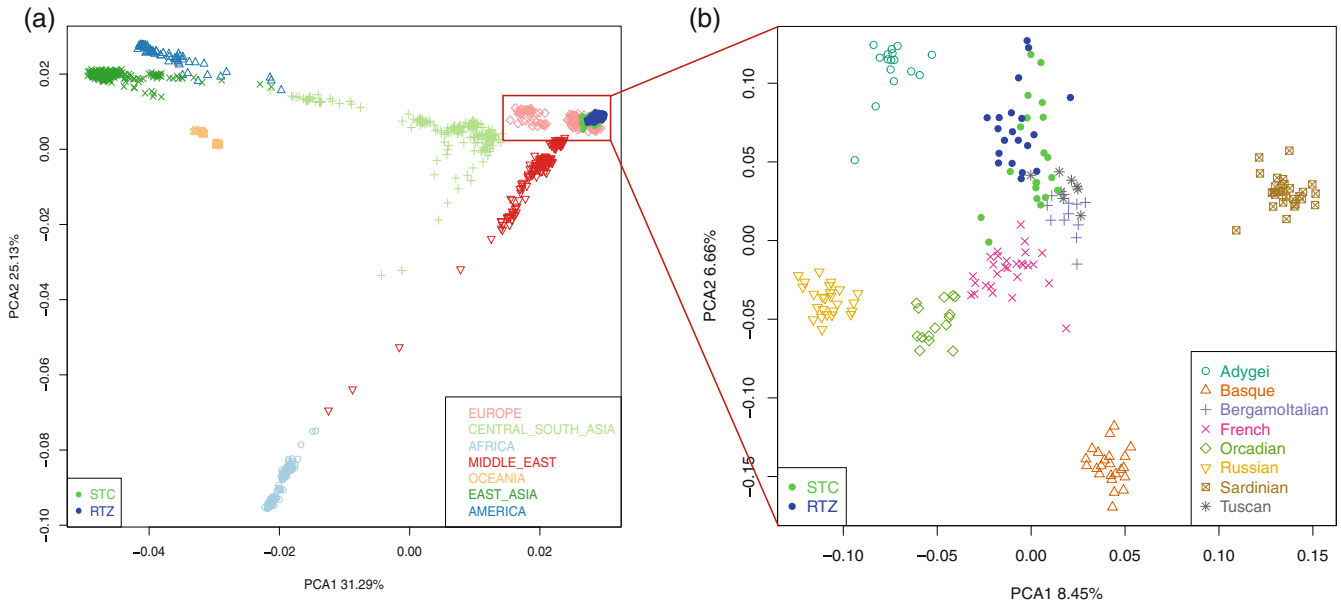

FIGURE 2 Principal component analysis (PCA). (a) Rotzo (blue) and Stoccareddo (green) in a worldwide contest. (b). PCA using only Europeans populations and subsampling 20 random individuals from RTZ and STC. On the x and y-axis are reported the variance explained by PC1 and PC2, respectively. RTZ, Rotzo; STC, Stoccareddo

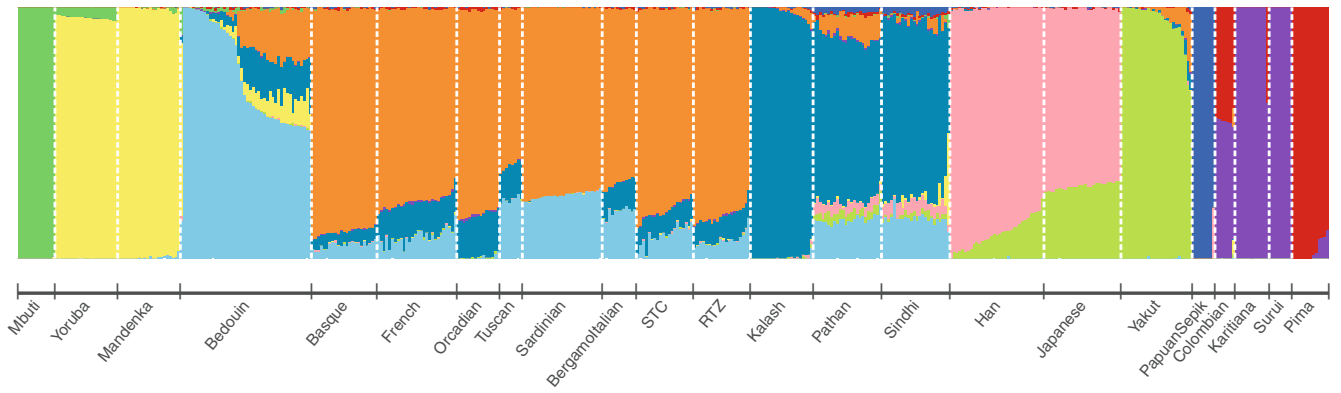

FIGURE 3 Admixture analysis for K = 9. The nine components are subdivided as follows: Light green component (Mbuti population, Africa), yellow component (Yoruba, Africa), cyan component (Bedouin population, West Asia), orange component (Basque population, Europe), navy component (Kalash population, South Asia), olive component (Yakut population, East Asia), blue component (Papuan population, South East Asia), violet component (Surui population, South America), red component (Pima population, South America).The orange component present in RTZ and STC is found in all European populations, especially in the Basque, the cyan component found in the near eastern population is found at a lower frequency in RTZ and STC compared to other Italian reference populations. RTZ, Rotzo; STC, Stoccareddo

highlighted a difference in the level of homozygosity in RTZ and STC. Using ROH between 1 and 8 Mb, we found that RTZ shows a higher level of total homozygosity (Figure 4a). This pattern becomes clearer when looking at longer ROHs (>8 Mb) where RTZ has a similar isolation pattern with other Italian isolates, where STC drift closer to other outbred European populations (Figure 4b). RTZ showed a significant higher level of both medium ROHs (1–8 Mb, Mann–Whitney p = 6.5e-05) and long ROHs (>8 Mb, Mann–Whitney p = 2.3e-4).

Analysis of effective population size (Ne), revealed that the Ne of RTZ and STC was similar in the last 800 years (using a generation time of 28 years). However, in the last 300–400 years we observed a drop in the Ne, which could be the start of an increased isolation, limited grow, and reduced gene flow in the population, additionally, RTZ was showing a lower Ne than STC. (Supplementary Figure 3).

3.2 | Recent and ancient admixture

The investigation of the time of admixture using only the modern population revealed that STC experienced one event of admixture 94+/−12 generations ago, RTZ showed two close admixture events, one 113+/−10 generations ago and a second one 88+/−10 generations ago. If we consider a generation time of

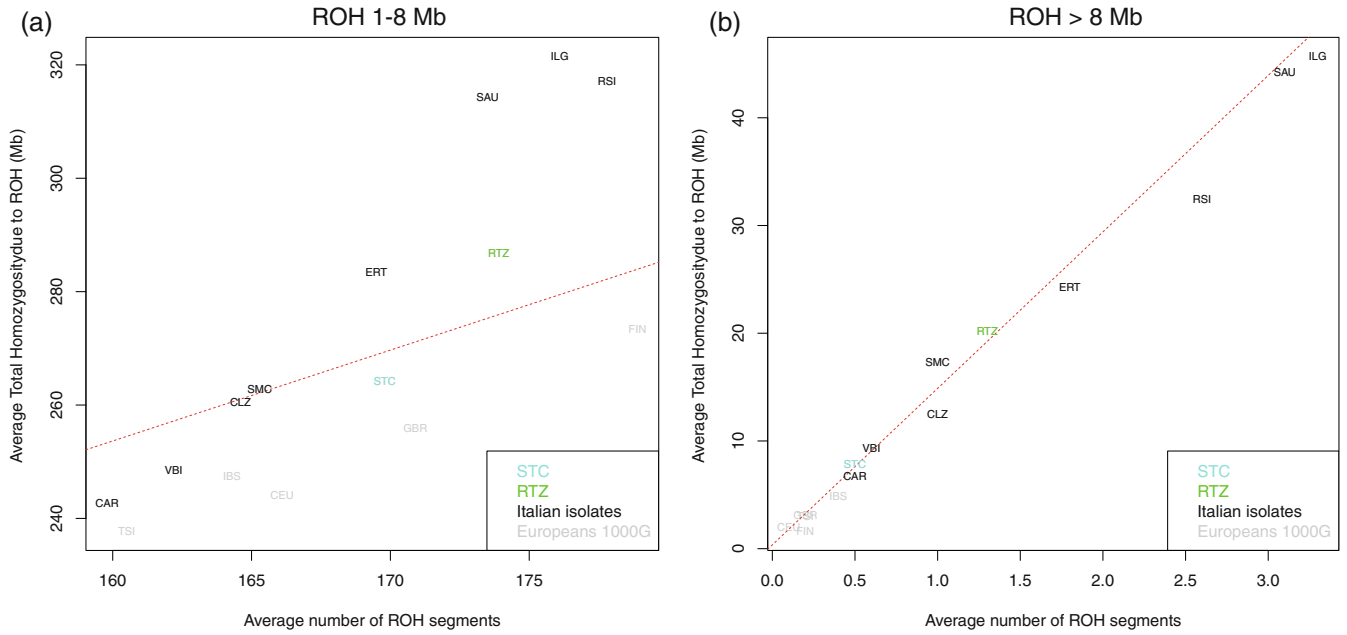

FIGURE 4 Runs of homozygosity pattern for RTZ and STC in the European and Italian isolates' context. On the x-axis is reported the average number of homozygous segments in each population and on the y-axis is reported the average total homozygosity for each population. The red line represents the regression line between total homozygosity and the number of segments using all European populations (a) pattern of homozygosity using ROH of between 1 and 8 Mb, we can observe how RTZ has higher total homozygosity compared to STC which show a pattern similar to GBR, both isolates have less ROHs than the Finnish (FIN) but RTZ shows higher total homozygosity. (b) Pattern of homozygosity using ROH of at least 8 Mb. We can see how STC group with villages with a low level of isolation such as Val Borbera (VBI) and has similar pattern to outbred population. RTZ, instead, shows a pattern of stronger isolation and groups with other more isolated populations, such as Erto (ERT), despite not reaching a level of homozygosity comparable with other north eastern Italian isolates. RTZ, Rotzo; STC, Stoccareddo

28 years, these events range from 1444–864 BCE to 724–164 BCE for RTZ and 948–276 BCE for STC.

However, GLOBETROTTER analysis revealed a set of more recent events for STC (60 generations ago, ~1700 years ago), with Tuscan (proportion 51%) and French (49%) as best fitted donors, where RTZ show an even more recent event (40 generations ago, ~1200 year ago) with French (proportion 29%) and Tuscan (proportion 71%) as best fitted donors.

With the f3 statistics calculated using only the ancient genome, we highlighted a higher number of significant pairs of ancient sources for STC (with Zscore <= −3); probably due to the higher level of isolation found in RTZ. However, the results seem consistent between the two populations, hinting at a shared ancient history of the region. In particular, we found that a population from the Bronze Age from Russia (Russia_BA_Okunevo. SG) is a common source of gene flow for both RTZ and STC when paired with Anatolia Neolithic, qpAdm revealed that the fraction of Russian Bronze Age is 0.14 and 0.15 for STC and RTZ respectively. On the other hand, STC seems to result from a mixture of genomes from Italy during the Imperial Age (admixture proportion: 0.57) and Serbia Mesolithic genomes (admixture proportion of 0.43) (Supplementary Table 2). Furthermore, when we modeled RTZ and STC as a mixture of Anatolia Neolithic and Middle Bronze Age genomes from England we discovered that the admixture fractions with England_MBA are 0.65 for STC and 0.72 for RTZ. All models have a $p > 0.05$ which indicate a good fit with the data. (Supplementary Table 3).

We then analyzed the direction of gene flow with different ancient genomes using the f4 statistics and comparing STC and RTZ with the French and North Italy.

We discovered significant gene flow from Italy Late antiquity and Italy Imperial genomes to RTZ and STC compared to the French, and no significant difference in gene flow when looking at Hungary Langobards and Italy North Early Medieval (Supplementary Figure 4a,b). However, when comparing the same set of genomes and substituting French with North Italians, RTZ shows a different pattern, with more North European ancestry, we can observe a significant gene flow to the village from Hungary Langobard, where no significant differences were found when using ancient Italian genomes. STC shows even less differences from the outbred North Italian reference, with a higher gene flow from North European populations from the Bronze Age. (Figure 5a,b). From outgroup f3 statistics (Supplementary Figure 5) we can see how RTZ and STC, when compared to other Italian populations, show higher shared drift with a North European reference population.

3.3 | Analysis of genetic drift on low-frequency variants

Analysis of the Fst outlier revealed a total of 3 markers in RTZ that have an Fst over the 95th percentile of the genomic distribution and frequency, that are present in the GWAS catalog associated with HDL and LDL traits.

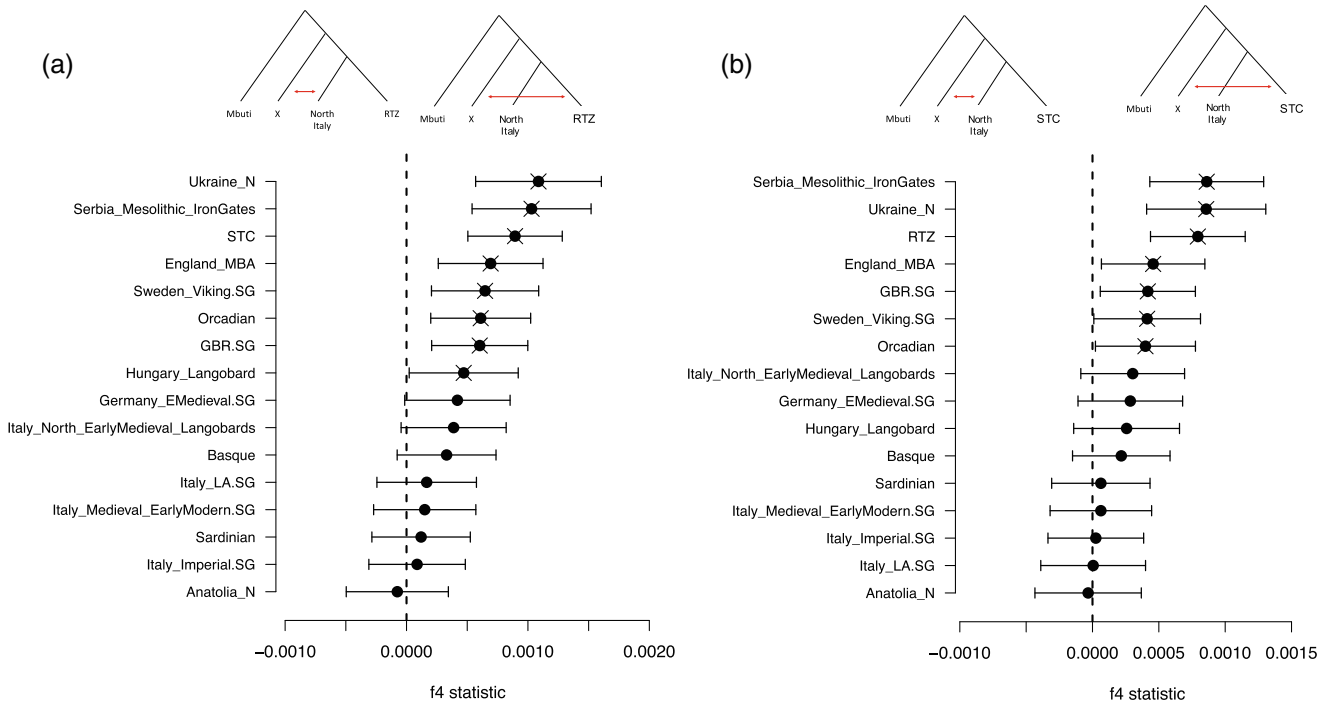

FIGURE 5 Analysis of relationship with ancient genomes using F4 statistics. (a) f4 (outgroup, ancient genome, RTZ, North Italy), a positive value indicates a close relationship between the ancient genome and RTZ compared to North Italy: With a cross is marked the results with Zscore<= −3. (b) f4 (outgroup, ancient genome, STC, North Italy) a positive value indicates a close relationship between the ancient genome and STC compared to North Italy; with a cross are marked the results with Zscore<= −3. The whiskers represent three times the standard errors of the f4 statistics. RTZ, Rotzo; STC, Stoccareddo

One of these variants is rs78025076, which is entirely monomorphic in RTZ and has a frequency of 1% in STC, whereas the risk allele has a frequency of 2% in Europe and 2.3% in North-East Italy. This variant is a missense located in the *MCUB* gene, and it is associated with different types of phenotypes linked to lipid traits(Richardson et al., 2020; Sinnott-Armstrong et al., 2021). In particular, the T allele is associated with increased HDL and lower LDL, and this allele has a frequency of 100% in RTZ. Another variant is rs16928809 in *SLC22A18*. The frequency of the A allele in RTZ is 3%, in STC 7% but in Europe and North-East Italy it's above 10%, with the A allele associated with decreased HDL. Finally, a third variant is rs11076174 in the *CETP* gene. The C allele has only one carrier in RTZ, a frequency of 6% in STC, 9% in FVG and 12.5% in the French population. In our case, the T allele was associated with increased HDL.

3.4 | Phenotypes linked to lipid metabolism

We observed no significant differences (Mann–Whitney adjusted $p > 0.05$) between RTZ and STC in HDL, LDL, total cholesterol, and triglycerides levels. However, we discovered significant differences (Mann–Whitney adjusted $p > 0.05$) between the two villages RTZ and STC, and other populations from North-East Italy (FVG). More in detail, RTZ and STC show lower LDL (median 123 and 122 respectively) with respect to FVG (median 135). In addition, RTZ and STC show higher HDL levels compared to FVG (median 57 and 58 versus 54). (Supplementary Figures 6 and 7). Regression analyses using sex and age as covariates revealed a significant association between the village of origin in HDL and LDL levels; in particular, HDL and LDL are lower in STC and RTZ than FVG.

4 | DISCUSSION

Population isolates represent a focus of interest to geneticists because of their potential to increase the ability to detect disease-causing genetic variants (Hatzikotoulas et al., 2014; Panoutsopoulou et al., 2014; Xue et al., 2017) and their fascinating history. In particular, the Italian peninsula harbors peculiar geography and history that might have produced specific genetic patterns in the populations (Sazzini et al., 2016). Our study focused on people inhabiting two villages in the Asiago plateau: RTZ and STC. Several other isolates were previously described in the northeast part of Italy (Esko et al., 2013). However, the fascinating history of the Asiago plateau prompted us to investigate the genetic characteristics of these two villages. The most striking aspect was discovering a steep difference in haplotype diversity for the Y chromosome and mitochondrial DNA between our two villages. In addition, the ratio of mitochondrial on Y haplogroup diversity is higher than 1 in STC, where is ∼1 in RTZ, which could hint at a recent sex-biased migration in one of the two isolates. The historical sources further emphasize this, indicating the birth of STC as promoted by the Gallio municipality around the 15th cent. CE, when a

few number of patriarchal families settled in the area (Paganin, 2013). Our hypothesis of a sex-biased migration only in STC is also based on the analysis of the genetic structure using principal components, as it did not reveal a significant difference in terms of autosomal ancestry. In addition, the isolate with the highest level of long ROHs, which indicated recent endogamy (Arciero et al., 2021; Ceballos et al., 2018; McQuillan et al., 2008), is the also one with the highest level of Y and mitochondrial haplotype diversity (RTZ), where the one with reduced Y haplotype diversity respect to mitochondrial haplotype diversity (STC) is the one with lower level of isolation. Thus, the absence of long ROHs could result from a small sex-biased gene flow that disrupted the actual pattern of homozygosity but not the distribution of ancestral components. We did not observe the presence of admixture events in the last ~40 generations, even using a haplotype-based approach such as GLOBETROTTER which is geared to detect recent admixture events (Chimusa et al., 2020).This timeline should roughly correspond to the Langobards invasion of Italy (Amorim et al., 2018; Vai et al., 2019). Nonetheless, we did find significant allele sharing with Hungary Langobards in RTZ compared to other North Italian reference populations. We should also note that our populations appear to be a mixture of Italy Imperial genomes described in (Antonio et al., 2019) and Mesolithic and Early Neolithic individuals ('Latvia_HG')(Mathieson et al., 2018), indicating a much more complex history in the formation of the genomic pool of the ancestors of these villages.

The lack of recent admixture events could also be related to their isolation level, as the recent gene flow could be negligible compared to the amount of genetic drift due to the small effective population size.

Our analysis discovered a series of admixture events that can be traced back to 80–100 generations ago. Considering a generation time of 28 years could be fitted between 1444–864 BCE/ 724–164 BCE for RTZ and 948–276 BCE for STC. However, another haplotype approach revealed a set of more recent events around 305 CE for STC and 865 CE for RTZ.

Furthermore, these two villages show higher genetic affinity to North European populations from the Bronze Age than other Italian reference populations. Between the Neolithic and the Ancient Bronze Age, no permanent settlement has been identified on the Asiago plateau, but only on its slopes and inner valleys with direct access to the surrounding plains. As already noted, while the Monte Corgnon site was later frequented with periodic breaks until the end of the 1st millennium BCE, the Longalaita hillfort was soon abandoned in favor of the nearby promontory of Bostel; which constitutes the best-known protohistoric site of the plateau thanks to the systematic excavation campaigns carried out in the last 30-years (De Guio et al., 2011; Magnini et al., 2019; Magnini & Bettineschi, 2021). The acme of the Bostel village can be traced back to the 6th-2nd century BCE, as evidenced by a combination of stratigraphic, typological, and radiocarbon data (Magnini et al., 2020). During this phase, the site is reorganized as a permanent settlement with hybrid cultural characteristics, clearly indicating a connection between the Alpine world (Raetic or Fritzen-Sanzeno culture) and the Po Plain (the so-called ancient Veneti or Este culture).

The most significant admixture events for RTZ broadly correspond to the two key episodes documented in the archeological record, i.e. the first stable occupation of the Asiago plateau (during the Middle Bronze Age, around 1550–1350 BCE), and the main phase of the Bostel village (approx. 550–100 BCE). The higher genetic affinity with Bronze Age populations from northern Europe is strongly supported by the linguistic evidence. As all the epigraphic sources discovered in the Bostel village (including inscribed bones, and ceramic vessels) concur in indicating a local variant of the Raetic alphabet, which is widespread in the Alpine region between Austria, Switzerland, southern Germany, and northern Italy, with those from Bostel among the most southern occurrences (Marchesini, 2015).

Also, recent excavation campaigns in the Bostel village revealed limited, but very relevant material traces dating back to the beginning of the 1st millennium BCE that testify early connections of the site with the Laugen-Melaun culture, which developed between the Late Bronze Age and the early Iron Age in Trentino, South and East Tyrol, and the Engadin valley(Magnini et al., 2020).

These results are expected, as previous works showed the impact of populations of the Bronze Age throughout all of Europe (Allentoft et al., 2015). The third admixture event identified for RTZ around the year 865 CE might relate to the birth of the modern village of RTZ, which can be dated back to the 9th-10th century CE. This dating is confirmed by the construction of the church of St. Margherita – the oldest Christian building of the whole Asiago plateau – that was erected during the 10th century CE.

Further connections with the northern European world can be found during the Middle Ages, given the profound similarities between the traditional Cimbrian language of the Veneto-Trentini plateaus and the Bavarian or Tyrolean idiom used in the 12th century. For RTZ, several other elements seem to indicate the Nordic influence on the local substratum starting from the 12th and 13th century AD: (1) the dedication of the church built in 1231 to St. Engheltrude, an agionym absent in southern Veneto in the same period; (2) the evolution of the local toponymy and onomastics, which show a combination of Romance and Germanic contributions.

Combining all these shreds of evidence, these two populations show characteristics of an isolate with interesting traces of the history of the Italian peninsula not hampered by recent admixture events not hampered by recent admixture events. As previously stated, isolation, selection, and genetic drift affect the distribution of deleterious alleles (Chheda et al., 2017; Mezzavilla et al., 2015; Panoutsopoulou et al., 2014; Xue et al., 2017). Despite the pattern of deleterious variants is better described using whole-genome sequences, our investigation revealed the effect of genetic drift on specific variants in *SLC22A18*, *MCUB*, and *CETP*, linked to several lipid traits (Richardson et al., 2020; Sakaue et al., 2021; Willer et al., 2013). Interestingly the distribution of HDL and LDL in our isolated village is different from other geographically neighboring populations. The relatively different frequency in the European and nearby isolated populations of these variants in comparison to the populations of the Asiago plateau could hint that genetic drift or selection could be the main factor behind such reduction. It is not easy to discern the role of selection when paired with genetic drift, but this result suggests that future genotype–phenotype studies could reveal more population-specific variants linked

to HDL and LDL in these populations. We hypothesize that the future use of whole-genome sequences at high coverage will uncover more of these instances. In conclusion, our work showed the usefulness of characterizing isolated populations by coupling historical sources and genetic data to provide a more precise and detailed picture of the population history and the distribution of specific alleles, improving the understanding of how several genetic factors could influence trait distributions.

AUTHOR CONTRIBUTIONS

Massimo Mezzavilla: Conceptualization (lead); data curation (supporting); formal analysis (lead); investigation (lead); methodology (lead); supervision (lead); writing – original draft (lead); writing – review and editing (lead). **Massimiliano Cocca:** Data curation (lead); methodology (supporting); resources (supporting); writing – review and editing (equal). **Pierpaolo Maisano Delser:** Formal analysis (supporting); methodology (supporting); writing – review and editing (supporting). **Margherita Francescatto:** Formal analysis (supporting); methodology (supporting). **Paola de Gemmis:** Data curation (supporting); methodology (supporting). **Daniela Segat:** Data curation (supporting); methodology (supporting). **Paola Cattelan:** Data curation (supporting); methodology (supporting). **Marina Da Meda:** Data curation (supporting); methodology (supporting). **Luigi Magnini:** Conceptualization (supporting); investigation (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Cinzia Bettineschi:** Conceptualization (supporting); investigation (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Armando De Guio:** Conceptualization (supporting); investigation (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Paolo Gasparini:** Conceptualization (supporting); funding acquisition (lead); investigation (supporting); project administration (equal); supervision (equal); validation (supporting); writing – review and editing (supporting).

ACKNOWLEDGMENTS

We express our gratitude to Anna Albarello and Giuseppe Baschirotto for promoting and supporting the project. We would like to thank all the volunteers and all the physicians, nurses and researchers that kindly contributed to this project. A special thanks to RTZ and STC inhabitants for having participated to the study. This work was also supported by IRCCS Burlo Garofolo of Trieste (5X1000 funding).

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

DATA AVAILABILITY STATEMENT

The genetic data belonging to INGI cohorts and used for comparison in this manuscript have been submitted to the European Variation Archive (EVA) are accessible in Variant Call Format (VCF) at the following link: https://www.ebi.ac.uk/ena/data/view/PRJEB33648 (accessed on November 2021). The genetic data belonging to Rotzo and Stoccareddo cohorts have been submitted to the European Variation Archive (EVA) information on allele frequencies in each cohort are accessible in Variant Call Format (VCF) at the following link: https://www.ebi.ac.uk/ena/browser/view/PRJEB53507 (accessed on August 2022).

ORCID

Massimo Mezzavilla https://orcid.org/0000-0002-9000-4595

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

Allentoft, M. E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P. B., Schroeder, H., Ahlström, T., Vinner, L., Malaspinas, A.-S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa, P. D., Dąbrowski, P., … Willerslev, E. (2015). Population genomics of bronze age Eurasia. *Nature*, *522*(7555), 167–172.

Amorim, C. E. G., Vai, S., Posth, C., Modi, A., Koncz, I., Hakenbeck, S., … Pohl, W. (2018). Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nature Communications*, *9*(1), 1–11.

Anagnostou, P., Dominici, V., Battaggia, C., Pagani, L., Vilar, M., Wells, R. S., … Francalacci, P. (2017). Overcoming the dichotomy between open and isolated populations using genomic data from a large European dataset. *Scientific Reports*, *7*(1), 1–13.

Antonio, M. L., Gao, Z., Moots, H. M., Lucci, M., Candilio, F., Sawyer, S., … Aikens, R. C. (2019). Ancient Rome: A genetic crossroads of Europe and the Mediterranean. *Science*, *366*(6466), 708–714.

Arciero, E., Dogra, S. A., Malawsky, D. S., Mezzavilla, M., Tsismentzoglou, T., Huang, Q. Q., … Martin, H. C. (2021). Fine-scale population structure and demographic history of British Pakistanis. *Nature Communications*, *12*(1), 7189. https://doi.org/10.1038/s41467-021-27394-2

Bergström, A., McCarthy, S. A., Hui, R., Almarri, M. A., Ayub, Q., Danecek, P., … Kamm, J. (2020). Insights into human genetic variation and population history from 929 diverse genomes. *Science*, *367*(6484), eaay5012.

Bortolami, S., & Barbierato, P. (2009). Storia e geografia della colonizzazione germanica medievale. In P. Rigoni & M. Varotto (Eds.), *L'Altopiano dei Sette Comuni* (pp. 144–179). Cierre.

Browning, B. L., & Browning, S. R. (2013). Detecting identity by descent and estimating genotype error rates in sequence data. *The American Journal of Human Genetics*, *93*(5), 840–851.

Browning, S. R., & Browning, B. L. (2015). Accurate non-parametric estimation of recent effective population size from segments of identity by descent. *The American Journal of Human Genetics*, *97*(3), 404–418.

Cacciavillani, I. (1994). L'Autonomia dei Sette Comuni nel dominio della Serenissima. In *Storia dell'Altopiano dei Sette Comuni. Territorio e Istituzioni* (pp. 447–469). Neri Pozza Editore.

Ceballos, F. C., Hazelhurst, S., & Ramsay, M. (2018). Assessing runs of homozygosity: A comparison of SNP array and whole genome sequence low coverage data. *BMC Genomics*, *19*(1), 1–12.

Chheda, H., Palta, P., Pirinen, M., McCarthy, S., Walter, K., Koskinen, S., … Palotie, A. (2017). Whole-genome view of the consequences of a population bottleneck using 2926 genome sequences from Finland and United Kingdom. *European Journal of Human Genetics*, *25*(4), 477–484.

Chimusa, E. R., Defo, J., Thami, P. K., Awany, D., Mulisa, D. D., Allali, I., … Mazandu, G. K. (2020). Dating admixture events is unsolved problem in multi-way admixed populations. *Briefings in Bioinformatics*, *21*(1), 144–155.

Clark, D. W., Okada, Y., Moore, K. H. S., Mason, D., Pirastu, N., Gandin, I., … Wilson, J. F. (2019). Associations of autozygosity with a broad range of human phenotypes. *Nature Communications*, *10*(1), 1–17. https://doi.org/10.1038/s41467-019-12283-6

Cocca, M., Barbieri, C., Concas, M. P., Robino, A., Brumat, M., Gandin, I., … Mezzavilla, M. (2020). A bird's-eye view of Italian genomic variation through whole-genome sequencing. *European Journal of Human Genetics*, *28*(4), 435–444. https://doi.org/10.1038/s41431-019-0551-x

Colonna, V., Pistis, G., Bomba, L., Mona, S., Matullo, G., Boano, R., … Achilli, A. (2013). Small effective population size and genetic homogeneity in the Val Borbera isolate. *European Journal of Human Genetics*, *21*(1), 89–94.

Coluzzi, P. (2005). Language planning for the smallest language minority in Italy: The Cimbrians of Veneto and Trentino-Alto Adige. *Language Problems and Language Planning*, *29*(3), 247–269.

Consortium, G. P. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74.

Cristiani, E., Lemorini, C., & Dalmeri, G. (2012). Ground stone tool production and use in the late upper palaeolithic: The evidence from Riparo Dalmeri (venetian Prealps, Italy). *Journal of Field Archaeology*, *37*(1), 34–50. https://doi.org/10.1179/0093469011Z.0000000003

De Guio, A. (1994). Dal Bronzo Medio all'inizio dell'età del Ferro. In *Storia dell'Altopiano dei Sette Comuni. Territorio e Istituzioni* (pp. 157–177). Neri Pozza Editore.

De Guio, A., Bressan, C., Ferrari, G., Mantoan, R., Gamba, M., Migliavacca, M., … Nicosia, C. (2011). Bostel di Rotzo (VI) - Stato di avanzamento delle ricerche. *Quaderni Di Archeologia Del Veneto*, *XXVII*, 168–183.

Esko, T., Mezzavilla, M., Nelis, M., Borel, C., Debniak, T., Jakkula, E., … D'Adamo, P. (2013). Genetic characterization of northeastern Italian population isolates in the context of broader European genetic diversity. *European Journal of Human Genetics*, *21*(6), 659–665. https://doi.org/10.1038/ejhg.2012.229

Goldstein, J. I., Crenshaw, A., Carey, J., Grant, G. B., Maguire, J., Fromer, M., … Stevens, C. (2012). zCall: A rare variant caller for array-based genotyping genetics and population analysis. *Bioinformatics*, *28*(19), 2543–2545.

Hatzikotoulas, K., Gilly, A., & Zeggini, E. (2014). Using population isolates in genetic association studies. *Briefings in Functional Genomics*, *13*(5), 371–377.

Hellenthal, G., Busby, G. B. J., Band, G., Wilson, J. F., Capelli, C., Falush, D., & Myers, S. (2014). A genetic atlas of human admixture history. *Science*, *343*(6172), 747–751.

Johnson, E. C., Evans, L. M., & Keller, M. C. (2018). Relationships between estimated autozygosity and complex traits in the UK biobank. *PLoS Genetics*, *14*(7), e1007556.

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genetics*, *8*(1), e1002453.

Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., Reshef, Y. A., Finucane, H. K., … Abecasis, G. R. (2016). Reference-based phasing using the haplotype reference consortium panel. *Nature Genetics*, *48*(11), 1443–1448.

Loh, P.-R., Palamara, P. F., & Price, A. L. (2016). Fast and accurate long-range phasing in a UK biobank cohort. *Nature Genetics*, *48*(7), 811–816.

Magnini, L., & Bettineschi, C. (2021). Object-based predictive modeling (OBPM) for archaeology: Finding control places in mountainous environments. *Remote Sensing*, *13*(6), 1197. https://doi.org/10.3390/rs13061197

Magnini, L., Bettineschi, C., De Guio, A., Burigana, L., Colombatti, G., Bettanini, C., & Aboudan, A. (2019). Multisensor-multiscale approach in studying the proto-historic settlement of Bostel in northern Italy. *Archeologia e Calcolatori*, *30*, 347–365. https://doi.org/10.19282/ac.30.2019.20

Magnini, L., Bettineschi, C., De Guio, A., Burigana, L., Pedersoli, S., & Griggio, E. (2020). Non solo scavo: novità dal Bostel di Rotzo (VI) a venticinque anni dall'avvio delle indagini dell'Università di Padova. FOLD&R, (*473*), 1–18. http://www.fastionline.org/docs/FOLDER-it-2020-473.pdf

Marchesini, S. (2015). *Monumenta linguae raeticae*. Scienze e Lettere.

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., … Cheronet, O. (2018). The genomic history of southeastern Europe. *Nature*, *555*(7695), 197–203.

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., … Tenesa, A. (2008). Runs of homozygosity in European populations. *The American Journal of Human Genetics*, *83*(3), 359–372.

Mezzavilla, M., Vozzi, D., Badii, R., Khalifa Alkowari, M., Abdulhadi, K., Girotto, G., & Gasparini, P. (2015). Increased rate of deleterious variants in long runs of homozygosity of an inbred population from Qatar. *Human Heredity*, *79*(1), 14–19. https://doi.org/10.1159/000371387

Nei, M., & Tajima, F. (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics*, *97*(1), 145–163.

Paganin, M. (2013). *Antiche famiglie di Asiago: alle origini di una comunità rurale*. GNG.

Panoutsopoulou, K., Hatzikotoulas, K., Xifara, D. K., Colonna, V., Farmaki, A.-E., Ritchie, G. R. S., … Zeggini, E. (2014). Genetic characterization of Greek population isolates reveals strong genetic drift at missense and trait-associated variants. *Nature Communications*, *5*, 5345. https://doi.org/10.1038/ncomms6345

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., … Reich, D. (2012). Ancient admixture in human history. *Genetics*, *192*(3), 1065–1093.

Pickrell, J. K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., … Reich, D. (2014). Ancient west Eurasian ancestry in southern and eastern Africa. *Proceedings of the National Academy of Sciences*, *111*(7), 2632–2637.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*(11), e1002967.

Poznik, G. D. (2016). Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. *BioRxiv*, 88716.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Daly, M. J. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575.

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*(7263), 489–494.

Richardson, T. G., Sanderson, E., Palmer, T. M., Ala-Korpela, M., Ference, B. A., Davey Smith, G., & Holmes, M. V. (2020). Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLoS Medicine*, *17*(3), e1003062.

Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., … Akiyama, M. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. *Nature Genetics*, *53*(10), 1415–1424.

Sarno, S., Boscolo Agostini, R., De Fanti, S., Ferri, G., Ghirotto, S., Modenini, G., … Boattini, A. (2021). Y-chromosome variability and genetic history of commons from northern Italy. *American Journal of Physical Anthropology*, *175*, 665–679.

Sartori, A. D. (1956). *Toria della federazione dei sette comuni Vicentini*. Amministrazione di Gallio.

Sazzini, M., Ruscone, G. A. G., Giuliani, C., Sarno, S., Quagliariello, A., De Fanti, S., … Catanoso, M. (2016). Complex interplay between neutral and adaptive evolution shaped differential genomic background and disease susceptibility along the Italian peninsula. *Scientific Reports*, *6*, 32513.

Sinnott-Armstrong, N., Tanigawa, Y., Amar, D., Mars, N., Benner, C., Aguirre, M., … Kiiskinen, T. (2021). Genetics of 35 blood and urine biomarkers in the UK biobank. *Nature Genetics*, *53*(2), 185–194.

Vai, S., Brunelli, A., Modi, A., Tassi, F., Vergata, C., Pilli, E., … Baricco, L. P. (2019). A genetic perspective on longobard-era migrations. *European Journal of Human Genetics*, *27*(4), 647–656.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 1358–1370.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., … Schönherr, S. (2016). HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research*, *44*(W1), W58–W63.

Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., … Mora, S. (2013). Discovery and refinement of

loci associated with lipid levels. *Nature Genetics*, *45*(11), 1274–1283.

Xue, Y., Mezzavilla, M., Haber, M., McCarthy, S., Chen, Y., Narasimhan, V., … Southam, L. (2017). Enrichment of low-frequency functional variants revealed by whole-genome sequencing of multiple isolated European populations. *Nature Communications*, *8*, 1–7.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Mezzavilla, M., Cocca, M., Maisano Delser, P., Francescatto, M., de Gemmis, P., Segat, D., Cattelan, P., Da Meda, M., Magnini, L., Bettineschi, C., de Guio, A., & Gasparini, P. (2022). Genetic characterization of two North Italian villages: A story of isolation, ancient admixture, and genetic drift. *American Journal of Biological Anthropology*, *179*(3), 460–470. https://doi.org/10.1002/ajpa.24620