# Genomic insights into the history of prehistoric Eastern Alps

Valentina Coia, Alice Paladin, Giovanna Cipollini, Christina Wurst, Frank Maixner, Albert Zink
Eurac Research, Institute for Mummy Studies, Bolzano, Italy
Contact: valentina.coia@eurac.edu

## 1. Background

The Tyrolean Iceman (3360-3100 cal. BC) is the best representative of the Copper Age in the eastern Italian Alps. Besides the Iceman[1], only one Mesolithic sample (Veneto Dolomites)[2] has been analysed for the genomic variation. Comparison with published genomic data, have shown that the Iceman clusters with Early Neolithic farmers (ENF) from different parts of Europe and with Neolithic individuals from Anatolia. In addition, European individuals contemporary of the Iceman cluster together. The latter also differ from the Iceman in their ancestry and genomic admixture patterns[3-4]. Since the Iceman alone cannot be considered as representative of the genomic diversity of this alpine area, we are analysing molecularly additional prehistoric samples from the Iceman´s territory (from the Copper to the Early Bronze Age).

## 2. Aims

General aim of the project is to characterize the genomic diversity and the genetic structure of prehistoric samples from the Iceman´s territory. More specific questions are: i) do the alpine prehistoric samples cluster together? ii) do they cluster with the ENF as it happens for the Iceman? Iii) do they show the same proportion of the three ancestral components detected in the Iceman? Comparison with archaeological data will also make it possible to insight the so-called "Alpine cultural group" in the Eastern (italian) Alps during the Copper Age as suggested by archeologists.

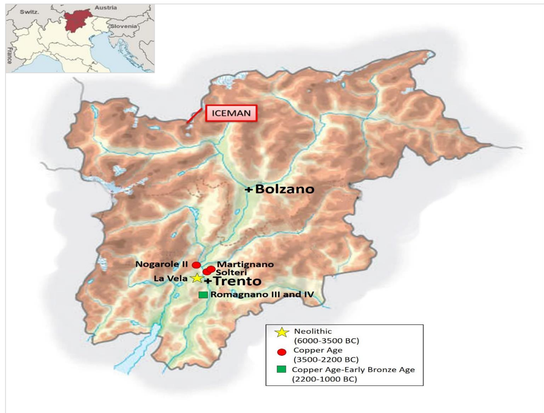



Figure1. Map of Trentino-Alto Adige region, Italy. Distribution of the 5 archaeological sites considered in this study and information about cronology and dating. * = Relative dating. Archaeological excavations were conducted from the 1960 to 1980 (Provincia Autonoma di Trento).

| Site | Cal BC 2 sigma |
|---|---|
| La Vela | 4433-4257 BC |
| Martignano | 3483-3129 BC |
| Iceman | 3360-3100 BC |
| Solteri | 3338-3031 BC** |
| Nogarole II | 2891-2702 BC |
| Romagnano III | 2284-2136 BC** |
| Romagnano IV | 2266-2043 BC** |

## 3. Methods

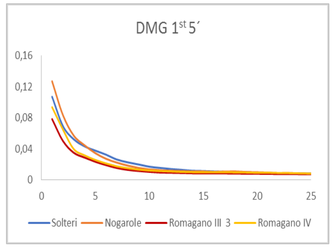
We are analyzing six individuals recovered from five archaeological sites distributed in the Adige valley (Trentino) (Figure 1). Up to know, DNA was extracted[5] from four samples and then single indexed library were constructed using the method reported in Meyer[6] and screened via shotgun sequencing (see Table 1 for more details). Paired-end sequences were merged and trimmed using PEAR[7] and aligned to the human and mitochondrial genome (build Hg19 and rCRS). Minimum mapping and base quality were set to 25. Postmortem damage and contamination estimates were made using mapDamage[8] and Schmutzi[9]. Haplogroup affiliation were performed based on HaploGrep2.0 and Phylotree (Table 1). Principal component analyses was built. SAMtools mpileup[10] and PileupCaller (https://github.com/stschiff/sequenceTools/tree/master/src-pileupCaller) called pseudodiploid genotypes for the individuals, at loci that overlapped with the 1240K targeted SNPs, and merged them to a Human Origins Affymetrix modern European and Middle eastern populations subset[11].

## 4. Results & perspective

All samples showed patterns of postmortem damage characteristic of ancient DNA (Figure3) and low contamination rates estimated based on mtDNA (Schmutzi; Table 1). Percentage of human DNA reads mapped to the human reference genome hg19 ranged from 12,3% to 57,69%. The average coverage estimates were low with a maximum value of 0.184 (SD = 1.022) for the sample of Romagnano (III). Alignment to the mitochondrial reference genome (rCRS) showed an average coverage for the mtDNA from 2,6 to ~25.

| | | nuclear | | | | mtDNA | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Site** | **Sample (S.P)** | **raw reads** | **End. Cont.** | **M. Cov.** | **SD** | **align. reads** | **M. Cov.** | **SD** | | **schmutzi** | | **hapl.** |
| Solteri | P2 upper (HiSeq) | 23.667.750 | 12,35% | 0,046 | 0,340 | 1258 | 4,846 | 2,341 | 0,99 | 0,98 | | J1c8 |
| Nogarole | PP (MiSeq) | 3.415.136 | 46,97% | 0,032 | 0,242 | 630 | 2,628 | 1,664 | 0,99 | 0,98 | | H5a4a1 |
| Romaganano III | PP (HiSeq) | 19.871.102 | 44,90% | 0,184 | 1,022 | 5286 | 24,956 | 6,061 | 0,01 | 0 | 0,02 | J1c3 |
| Romaganao IV | PP (HiSeq) | 16.268.693 | 57,69% | 0,179 | 0,885 | 4652 | 18,584 | 4,583 | 0,02 | 0,01 | 0,03 | H1n |

Table 1. P2 = upper molar; PP = pars petrosa; S.P = sequence Illumina platform; End. Cont. =% of Endogenous content; M.Cov. = Mean Coverage and standard deviation; align. Reads= aligned reads to the reference genome; hapl. = haplogroup. Figure 2 on the right. DMG 1st 5´ = C to T changes at the first position in the 5` end of the sequence read.




In order to further explore the genomic data, we performed principal component analyses and we compared the sample with the highest average coverage in our dataset (Romagnano III, Table 1) with genotypes of other ancient and present-day individuals from European and Middle Eastern populations. The plot (Figure 3) shows that the Romagnano III sample (2284-2136 BC, CA-EBA) fall outside the main group which includes most of the other samples from the same chronological phase from different geographic area (e.g. Germany, Poland). It is also far away from the Iceman, the only other alpine sample included in the dataset. The Iceman, as expected, share more ancestry with Neolithic (Anatolia) and Early Neolithic samples. On the other hand, the Romagnano sample shows genomic affinity with the NE samples from Sweden[12] and with two CA samples from Iberia[13]. However, it should be considered the currently low genome coverage of our sample and especially of that from Sweden (mean coverage 0.01). This first analyses shown good quality of ancient DNA from alpine samples dated to Chalcolithic to the Early Bronze age and suggests high genomic differentiation between alpine samples. However, more samples with highest average coverage would be necessary in order to further explore genomics diversity of alpine individuals. All samples will be further subjected to additional molecular screening and deeper sequencing.

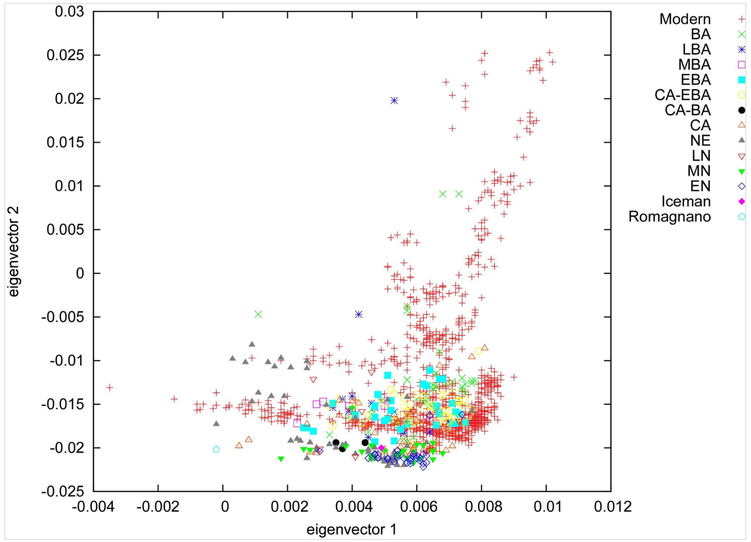



Figure3. PCA plot of Romagnano III individual, modern and ancient selected samples from Europe and Middle est14. BA= Bronze Age; LBA= Late Bronze Age; MBA= Middle Bronze Age; EBA= Early Bronze Age; CA-EBA= Chalcolithic-Early Bronze Age; CA-BA= Chalcolithic-Bronze Age; CA= Chalcolithic; NE= Neolithic; LN= late Neolithic; MN= Middle Neolithic; EN= Early Neolithic.

## References

1. Keller et al. 2012. Nat Commun. 28;3:698; 2.Fu et al. 2016. Nature. 9;534(7606):200-5; 3. Mathieson et al. 2015. Nature. 24;528(7583):499-503; 4. Sikora et al. 2014. PLoS Genet. 8;10(5):e1004353; 5. Gamba et al. 2014. Nat Commun. 5:5257; 6. Meyer and Kircher. 2010. Cold Spring Harb Protoc; 7. Zhang et al. Bioinformatics 30, 614–20 (2014); 8. Ginolhac et al. 2011. Bioinformatics 27, 2153–2155; 9. Renaud et al. 2015. Genome Biol. 16; 10. Li H. & Durbin R. 2009. Bioinformatics 25, 1754–60; 11.Lazaridis et al. 2016. Nature 536, 419–424; 12. Skoglund et al. 2014. Science. 16;344(6185):747-50; 13. Günther et al. 2015.pnas. 22;112(38):11917-22.