Fast and reliable ancestral reconstruction on ancient genotype data with non-negative Least square and Principal Component Analysis

Luciana de Gennaro*[1], Ludovica Molinaro*[2], Alessandro Raveane[3], Federica Santonastaso[3], Sandro Sublimi Saponetti[1], Michela Carlotta Massi[3], Luca Pagani[4,5], Mait Metspalu[5], Garrett Hellenthal[6], Toomas Kivisild[2,5], Mario Ventura*[1], Francesco Montinaro*[1,5]

[1]Department of Biosciences, Biotechnology and Environment, University of Bari, Bari, Italy
[2]Department of Human Genetics, KU Leuven, Leuven, Belgium
[3]Human Technopole, Milan, Italy
[4]Department of Biology, University of Padova, Padova, Italy
[5]Institute of Genomics, University of Tartu, Tartu, Estonia
[6]Department of Genetics, Evolution & Environment, University College of London, London, UK

Contacts: luciana.degennaro@uniba.it, ludovica.molinaro@kuleuven.be, mario.ventura@uniba.it, francesco.montinaro@gmail.com

Abstract

The history of human populations has been strongly shaped by admixture events, contributing to the patterns of observed genetic diversity across populations. Given its significance for evolutionary and medical studies, many algorithms focusing on the inference of the genetic composition of admixed populations have been developed. In particular, the recent development of new ancestry estimation methods that consider the fragmentary nature of ancient genotype data, such as the f-statistics family and its derivations, have radically changed our understanding of the past. F-statistics capture similar genetic similarity information as Principal Component Analysis (PCA), which is widely used in population genetics to quantify genetic affinity between populations or individuals. In this study, we introduce ASAP (ASsessing ancestry proportions through Principal component Analysis) method that leverages PCA and Non-Negative Least Square (NNLS) to assess the ancestral compositions of admixed individuals given a large set of populations. We tested ASAP on different simulated models, incorporating high levels of missingness. Our results show its ability to reliably estimate ancestry across numerous scenarios, even those with a significant proportion of missing genotypes, in a fraction of the time required when using other tools. When harnessed on Eurasia's genotype data, ASAP helped replicate and extend findings from previous studies proving to be a fast, efficient, and straightforward new ancestry estimation tool.

Introduction

The history of human populations has been strongly shaped by past admixture events that cumulatively have contributed to patterns of genetic diversity observed today (Hellenthal et al., 2014; Montinaro et al., 2015). Several multidisciplinary studies proved that virtually all human populations have interacted throughout their history in complex demographic scenarios, including migration and admixture (Busby et al., 2016; Ongaro et al., 2019; Patterson et al., 2012; Schlebusch et al., 2012). These interactions resulted in a sudden or gradual transfer of genetic material, generating new groups different from their sources (Hellenthal et al., 2014). Given its significance for evolutionary and medical studies, many algorithms focusing on the inference of the genetic composition of admixed populations have been developed. In this context, it has been shown that using phased genotype data can offer higher-resolution description of genetic population structure (Haak et al., 2015; Hellenthal et al., 2014; Lazaridis et al., 2016; Leslie et al., 2015; Montinaro et al., 2015; Narasimhan et al., 2019; Ongaro et al., 2019; Pankratov et al., 2020; Skoglund et al., 2017).

However, existing methods often present limitations when dealing with low-coverage ancient DNA (aDNA) data. Algorithms using haploid-called genotypes to estimate allele frequencies and allele sharing probabilities at limited numbers of overlapping variant positions have been designed to meet these challenges.

Among the others, qpAdm (Haak et al., 2015; Harney et al., 2021) is one of the most widely used approaches on ancient data (Damgaard et al., 2018; Haber et al., 2017; Hajdinjak et al., 2018; Harney et al., 2018; Lazaridis et al., 2017, 2016; Narasimhan et al., 2019; Olalde et al., 2018; Skoglund et al., 2017), given its ability to deal with the challenges of aDNA data and model admixture events involving multiple sources (Haak et al., 2015; Harney et al., 2021; Patterson et al., 2012). This tool takes advantage of the fact that genetic variation within a specific population can be summarised by comparing its allele frequencies to those of three additional groups using a “treeness test” belonging to the F-statistics family, the *f4* metric (Harney et al., 2021; Patterson et al., 2006).

For a given target population T, a set of putative sources of admixture P_i, and a set of “right populations” R_i with different relationships to P_i, qpAdm builds a matrix A of *f4* in the form (T, X, R_1, R_i), in which X can alternatively be T or a P_i population. Given that any *f4* in the state (T, T, R_1, R_i) is 0, qpAdm solves the equation $w.A=0$, where w are

the admixture coefficients (weights), assuming that their sum is equal to 1 (Haak et al., 2015).

QpAdm framework can be iterated multiple times to test several scenarios, allowing the evaluation of the models based on their p-values. However, sifting through all possible proxy sources and the right populations for an admixture event can be overwhelming. In addition, a recent survey has shown that, depending on the approach and the quality of the genetic data analyzed, qpAdm may suffer from high false discovery rates, adding substantial uncertainty to the interpretation of the results of admixture inference (Eren Yüncü et al., 2023).

A similar approach, introduced by Haak et al. 2015, but less frequently employed, uses a Non-Negative Least Square (NNLS) approach on a matrix of *f4*s in the form *f4*(X, R_1, R_i, R_j), where X is either T or any P population (Haak et al., 2015; Lazaridis et al., 2016).

F-statistics results broadly recapitulate genetic relationships emerging from Principal Component Analysis (PCA) (Peter, 2022), widely used in population genetics to quantify genetic affinity between populations or individuals, including ancient ones.

There is indeed a geometric relationship between the two metrics, although they are based on different statistical principles: the f-statistic is based on the measurement of the branch lengths of a hypothetical tree in which the analyzed populations are related, while PCA reduces the dimensionality of the data while maintaining the maximum variance present among individuals. In detail, considering four populations A, B, C, and D projected in a PC space, the *f2*(A, B) is correlated with the Euclidean distance between A and B computed in PC coordinates, while the f3(A; B, C) will be proportional to the orthogonal projection of A-B on A-C. Similarly, the *f4*(A, B; C, D) will be related to the orthogonal projection of A-B onto C-D (Peter, 2022). Moorjani et al. 2011 showed that f4 ratios can be used to estimate the rate of admixture (Moorjani et al., 2011).

Considering these results, it is, in principle, possible to use PC coordinates to infer admixture proportions of a target population using a set of sources. Different attempts and approaches have recently been proposed using principal components (Conley et al., 2023).

In this study, we present ASAP (ASsessing Ancestry proportions through Principal component analysis), in which we aim to leverage PCA and NNLS to assess the ancestral compositions of admixed individuals given a large set of populations. We

test ASAP on different simulated models, incorporating high levels of missingness. We show its ability to reliably estimate ancestry across numerous scenarios, even those with a significant proportion of missing genotypes, in a fraction of the time required when using other tools.

RESULTS

ASAP workflow and datasets

Here we provide an overview of the methodology implemented in ASAP (Fig. 1). We simulated a set of 20 unadmixed and 16 admixed populations (Supp. Table 1). For each admixed group, we simulated an admixture event involving two or three sources (Molinaro et al., 2021) with minor source contribution ranging from 5 to 40%, to test ASAP performance in various conditions and settings, accommodating a wide range of routinely performed approaches. For each of the true sources, we also simulated a sister group that split 3 thousand years ago (KYA) to mimic a proxy source: a group related to the real admixing source but not the direct contributor to the admixture event. These proxy populations allowed us to test whether ASAP could infer the closest proxy sources to the admixing populations.

Specifically, we simulated admixture events between groups with different degrees of affinity, from highly divergent to closely related populations, with pairwise Fst between populations ranging from 0.01 to 0.23, including bottleneck events, expecting a lower assignment accuracy in cases where the source groups are genetically closer (Molinaro et al., 2021). For each scenario, we tested our approach on the average Principal Component (PC) coordinates from each admixed group (population-wise approach) and on each admixed individual separately (individual-wise approach).

We initially tested ASAP performance considering as putative admixture sources the entire panel of the true sources or their sister groups ('proxy sources'), even though only two (two-way admixture) or three (three-way admixture) sources were used to simulate the admixture event. We then ran PCA where the PC space was built by the true sources and their respective sister groups on the first 10 components, while all admixed groups were projected onto it. The number of components was selected after running a preliminary assessment of ASAP performance as described in Supplementary Text 1. Subsequently, using NNLS, we modeled the average PC coordinates across individuals of each admixed group as a mixture of those of all the available sources, considering as sources either the true or the sister groups panel. Standard Errors (SE) were estimated using a chromosome-based jackknife approach (Busing et al., 1999; Montinaro et al., 2015), as described in the methods section.

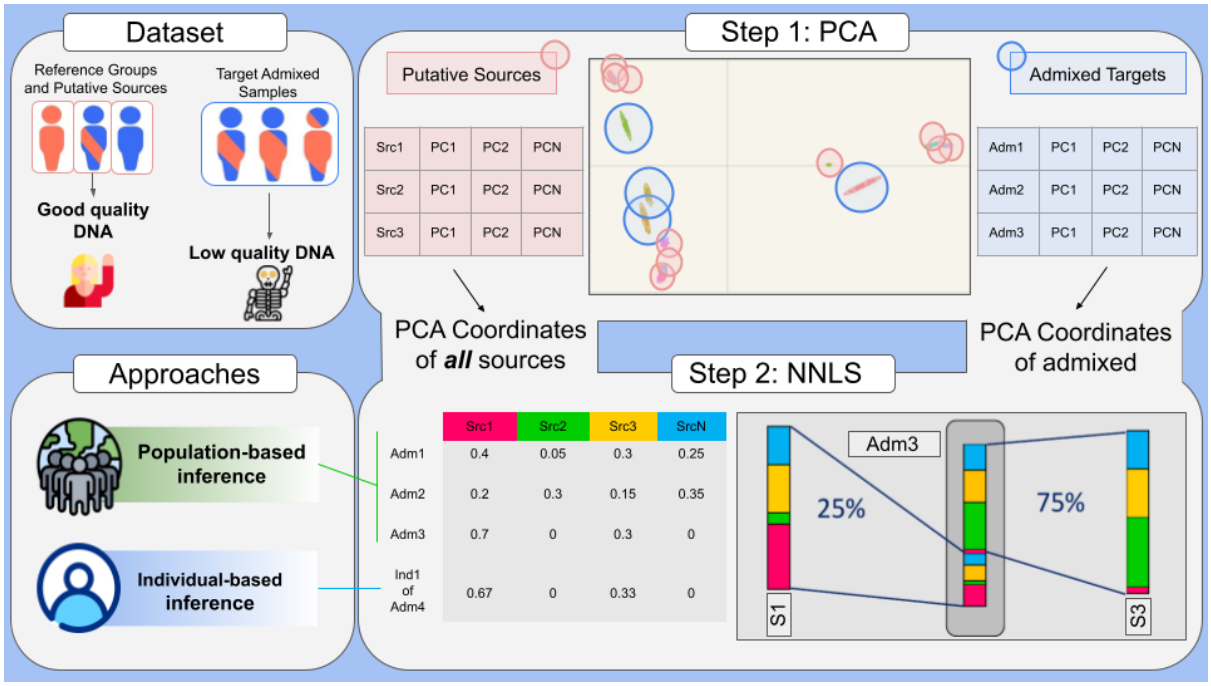


Fig. 1. Schematic representation of ASAP workflow. ASAP harnesses Non Negative Least Square using individual or population Principal Component vectors. PCA analysis can be carried out on no-missing genotype data, or using different approaches which accommodate different degrees of missingness.

Data and source availability

Our approach to ASsess Ancestral composition using Principal Component Analysis and NNLS (ASAP) is available as an R package at https://github.com/lm-ut/ASAP.

Results on Simulated genotypes with no missingness

ASAP performance with true sources: In the population-based approach, for all the 16 simulated admixed populations, ASAP successfully assigned the main ancestry components to the true sources that contributed to the admixture event despite the large panel of potential source groups available. The ancestry proportions of the true source groups (Supp. Table 2a) yielded a maximum error of 0.014 and a maximum jackknife standard error (SE) of 0.012, when two sources contributed to the target population (Fig. 2A-C). Minor additional contributions were assigned to other groups but never exceeding 0.004 (Pop 8 in Fig. 2A). In these cases, the additional ancestral

component was assigned to groups closely related to the true source (Fst ≤ 0.01, (Molinaro et al., 2021)). In three-way admixed populations (Pop 15 and Pop 16 in Fig. 2D-E), the true sources are always recognized with a maximum error of 0.010 (SE < 0.01).

The individual ancestry assignment estimation for (Supp. Table 2) 70-30% admixed populations (Pop 1-8) shows an average error lower than 0.029 when the admixing sources split more than 9 KYA (Kilo Years Ago), which increases (0.038) when the simulated population split was less than or equal to 9 KYA (Pop 6 and 7).

Admixed populations with lower source contributions (Pop 9-14) record an average error of a maximum of 0.022. In this case, lower error values are observed for populations with a less recent split (for example Pop 9 and Pop 12). The highest average error in the individuals-based analysis is observed in the three-way admixed populations (Supp. Table 2b). An over/underestimation exceeding 0.05 of the assigned contribution to the main sources in the 32% of individuals is recorded. Only one individual has an error larger than 0.1.

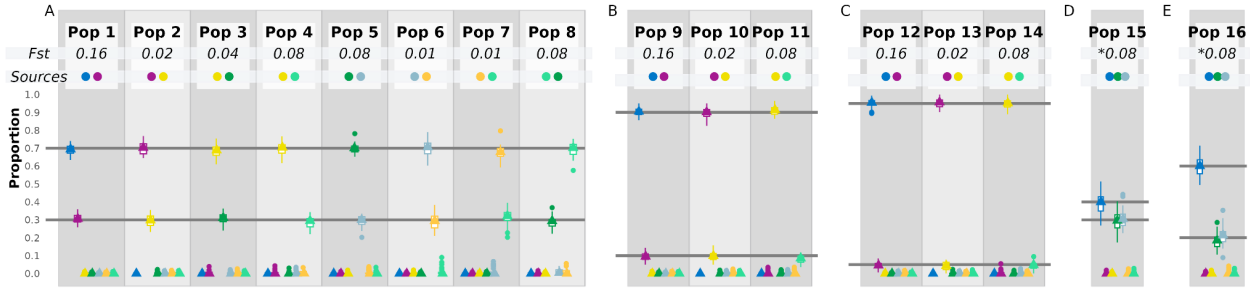


Fig. 2 ASAP assignment using true sources for each admixed group; triangles and boxplots show the population and individual ancestry estimation, respectively. On the upper part of the panel the *Fst* values (minimum values are marked with *) and the *Sources* that contribute to the admixture event in the simulated populations are shown: A) populations obtained from the combination of two sources with proportions 70-30%; B) admixed populations obtained from the combination of two sources with proportions 90-10% and **C)** 95-5%. **D)** Three-way admixed populations generated by combining three sources with 40-30-30% and **E)** 60-20-20% proportions.

ASAP performance with proxy sources:

We evaluated ASAP performance on a PCA where the admixed groups were projected onto the PC space built on all the remaining populations. We then modeled the admixed groups as a mixture of all the proxy sources only. As we knew which simulated proxy group was indeed the sister group of the real admixing source, we

calculated the assignment error by considering differences in observed and expected ancestry proportions and whether ASAP could indeed select the closest sister group. In this scenario, for all the 16 tested populations the proxies of the real sources were recognized without ever assigning even minimal contributions to other populations (Supp. Table 2).

For two-way admixed populations with proportions of 70-30%, the average error is 0.033 (SE < 0.009, Supp. Table 2C). Generally, the error estimates tend to be larger when the admixing source populations (Pop 3, 4, 6, 7 in Fig. 3) are characterized by a higher genetic similarity due to recent split times and bottleneck events. However, the error never exceeds 0.057 (Pop 4).

The ASAP accuracy is also robust in the case of three-way admixed populations, with a maximum error of 0.031 (SE < 0.0087).

The overestimation of the major component becomes more important in strongly imbalanced contribution cases. When the contribution of the minor source is 10% (Fig. 3B), the minor contribution is underestimated (Pop 10-11 in Fig. 3B). For populations where the minor source contributed 5%, ASAP completely misses the minor source contribution and assigns the total of the ancestral component to the main source (SE < 0.007) (Fig. 3C).

In individual-based inferences (Supp. Table 2D), ASAP correctly assigns ancestral proportions in the two- 70%-30% admixed individuals. The estimates obtained for the 50 individuals within each group are characterized by a maximum average error of 0.0582 (Pop 6, Fig. 3A).

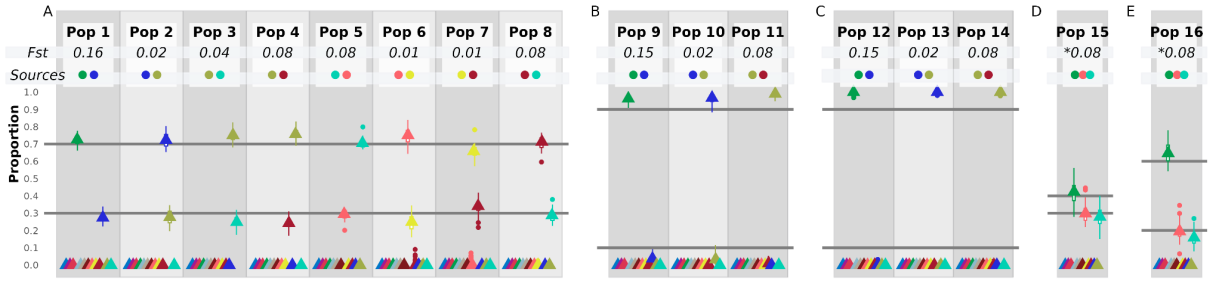


Fig. 3 ASAP assignment using proxy sources for each admixed group. Triangles and boxplots show the population and individual ancestry estimation, respectively. On the upper part of the panel the *Fst* values (minimum values are marked with *) and the proxy *Sources* that contribute to the admixture event in the simulated populations are shown: A) populations obtained from the combination of two sources with proportions 70-30%; B) admixed populations obtained from the combination of two sources with proportions 90-10% and C) 95-5%. D) Three-admixed populations generated by combining three sources with 40-30-30% and E) 60-20-20% proportions.

For admixed populations with minor source contributions of 10% and 5%, the contribution of the minor sources is underestimated or completely missed. In detail, for populations 9 and 10, the average estimated minor contribution is 0.038 and 0.034, with 29% of individuals showing less than 2%. On the other hand, in population 11, the average minor contribution is 0.016, with 64% of individuals showing less than 2%. No relevant contribution from other sources is recorded. For populations 12, 13 and 14, 85% of individuals are modeled as unadmixed, with the remaining individuals showing an average minor contribution of 1.3%.

For the three-way admixed populations, ASAP is always able to recognize the correct sources and assigns them the right proportions with a maximum error of 0.05 (Pop 16, Fig. 3E) due to the fact that for some individuals there is a slight overestimation of the main source (AFR) at the expense of the Asian source, one of the other two minor sources.

ASAP performance using pseudo-ancient data

We tested ASAP using pseudo-haploid samples, simulated by introducing different degrees of missing genotypes (up to 50%) and pseudo-haplodised (see materials and methods) mirroring the fragmentary nature of data commonly adopted in ancient DNA studies. We tested ASAP on a PCA where the PC space is built by the diploid genomes of the proxy sources, onto which we projected the pseudo-haploid genomes of the admixed groups and all possible true sources. In this scenario, we tested whether ASAP could model the admixed groups with the pseudo-haploid true sources.

ASAP correctly detects the closest admixture sources even in a large panel of putative donors, despite the target and source samples being pseudo-haploid and containing missing genotypes (Supp. Table 3, Fig. 4). Indeed, the average maximum assignment error is 0.033. In this case, ASAP always identifies the true sources and assigns a marginal additional component to other sources (maximum 0.004). Furthermore, the jackknife SE is also generally low, with a maximum of 0.017 (Supp. Table 3) seen in the admixed population whose sources split more recently (7.5 KYA). Even when single samples are targeted, the true sources are generally recognized and the major source ancestry assignments show an average error of 0.039. Despite the low average

error, the maximum per sample error can reach 0.248, caused by the misassignment to the most closely related group to the sources (Fst ≤ 0.01,(Molinaro et al., 2021)).

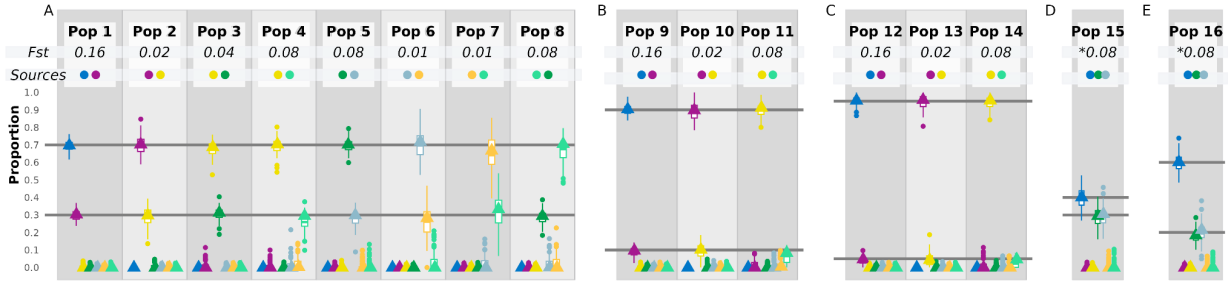


Fig. 4 ASAP assignment using pseudo-haploid simulated data and modeling each admixed group as a mixture of all the available proxy sources. Triangles and boxplots show the population and individual ancestry estimation, respectively. n the upper part of the panel the *Fst* values (minimum values are marked with *) and the proxy *Sources* that contribute to the admixture event in the simulated populations are shown: A) populations obtained from the combination of two sources with proportions 70-30%; B) admixed populations obtained from the combination of two sources with proportions 90-10% and C) 95-5%. D) Three-admixed populations generated by combining three sources with 40-30-30% and E) 60-20-20% proportions.

ASAP performance with limited reference genetic variation availability: We tested ASAP in a scenario where only the proxy, but not the true aDNA-like pseudo-haploid sources of the admixture, were available. The rationale behind this analysis is to mimic the lack of true mixing sources when exploring ancient DNA datasets while leveraging the availability of diploid genomes to build the PC space. In this case, an initial hypothesis of the demographic history of the admixed group is required, given that we are subsetting the donor panel to two or three putative source groups.

In this test, we projected onto the PC space the pseudo-haploid genomes of i) the target admixed group and ii) the closest proxy sources of each real source, two proxy sources in case of a two-way admixture, and three for the three-way admixture. The PC space was built with the diploid genomes of the remaining proxy sources. We modeled the target admixed group's relative admixture proportions given the projected proxy sources, relying on a limited donor panel of two or three groups.

Given the large error in individual analysis, mostly due to the lack of a proper reference dataset, we focused on the population-based approach (Fig. 5). In such a scenario (complete results available in Supp. Table 4), error estimates are lower than 0.043 for

all groups whose sources diverged more than 24 KYA (Pop 1, 4, 5, 8, 9, 11, 12, 14). For the only group whose sources split 24 KYA (Pop 3), the error increases to 0.11. In contrast, for all the other groups with closer sources, the error estimates range between 0.16 and 0.58, with jackknife SE estimation following the same pattern (Supp. Table 4).

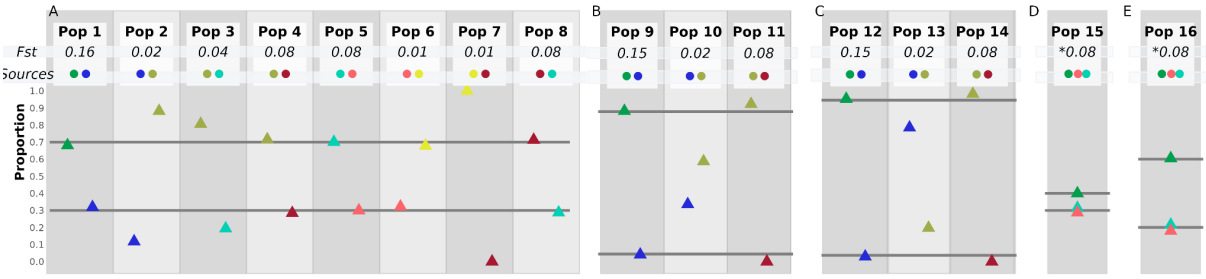


Fig. 5 ASAP performance with limited reference genetic variation availability. Only population-based inferences are shown.n the upper part of the panel the *Fst* values (minimum values are marked with *) and the proxy *Sources* that contribute to the admixture event in the simulated populations are shown: A) populations obtained from the combination of two sources with proportions 70-30%; B) admixed populations obtained from the combination of two sources with proportions 90-10% and C) 95-5%. D) Three-admixed populations generated by combining three sources with 40-30-30% and E) 60-20-20% proportions.

Benchmarking ASAP versus existing global ancestry inference tools

We compared ASAP with qpAdm, Rye, and Unlinked-ChromoPainter NNLS, which harness f4-statistics, PCA, and a modified Li and Stephens model with infinite recombination between SNPs for the ancestry composition inference, respectively (Conley et al., 2023; Haak et al., 2015; Harney et al., 2021; Li and Stephens, 2003). We compared the accuracy in estimating the ancestral proportions of the four approaches using the pseudo-haploid genomes of both the target admixed samples and the true sources of the admixture. Our method behaves similarly to the others (Fig. 6A-C); the correlation of ancestry assignments (Fig. 6D) of ASAP, qpAdm, and Rye is higher than 0.95 (ASAP vs qpAdm R2 = 0.968, p-value < 10e-6; ASAP vs Rye R2 = 0.998, p-value < 10e-6, ASAP vs CP R2 = 0.985). Among the four harnessed algorithms, qpAdm is characterized by the highest average error, and all four approaches show a lower accuracy for the admixed populations characterized by a subcontinental admixture, in which the two admixing sources are generically close (Fst ≤ 0.01) (Molinaro et al., 2021; See Suppl. Table 5).

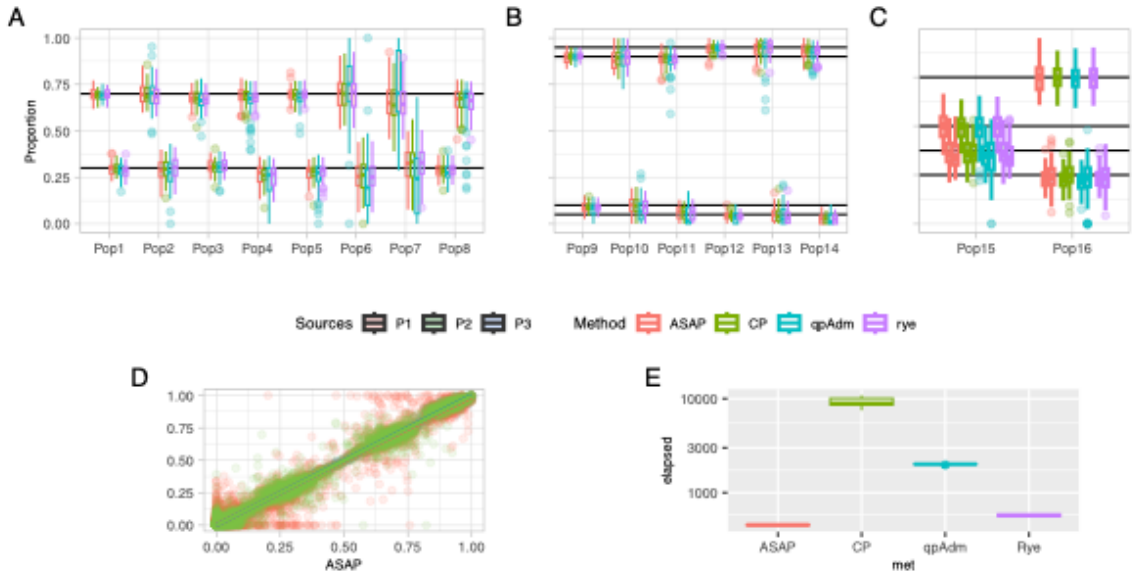


Fig. 6 Comparison between ASAP (red), ChromoPainter NNLS (CP, green) , qpAdm (blue), and Rye (purple) modeling the ancestral proportions of pseudo-haploid admixed populations given a set of pseudo-haploid sources; **A)** ancestry proportions for simulated populations with 70%- 30% sources' contribution; **B)** ancestry proportions for simulated populations with 90% - 10% (Pop 9-11) and 95%-5% (Pop 12-14) sources' contribution; **C)** ancestry proportions for three-way admixture simulated populations with 40%-30%-30% (Pop 15) and 60%-20%-20% (Pop 16) sources' contribution; **D)** correlation of the ancestry proportion assignment between CP, Rye and qpAdm in the y-axis and ASAP on the x-axis and **E)** computational time for a subset of 100 individuals.

We also compared the computational speed of each framework (Fig. 6E) replicating (10 iterations) the ancestry inference for the same set of 100 individuals. For ASAP and Rye, we included the PC (10 analysis) time, while for qpAdm we took into consideration the estimation of the f2. ASAP outperforms the other methods: ASAP computational time stands at 454 seconds (s) (sd = 5 s), while Rye reaches 575 s (sd = 14.1 s), qpAdm 2,011 s (sd = 22.399 s), and ChromoPainter 156 minutes and 23 seconds (9,383 s, sd = 1148 s).

ASAP performance on real data

We tested ASAP on real data using a dataset of different ancient Eurasian populations (Lazaridis et al., 2022). We projected 1,380 ancient individuals into the first 10 Principal components inferred using 1,668 present-day individuals. Following Lazaridis et al. 2022, we applied ASAP on 1,350 target individuals, using five putative sources: Western Hunter-Gatherers (WHG), Caucasus Hunter-Gatherers (CHG), Eastern Hunter-Gatherers (EHG), Anatolia Neolithic and Levant Neolithic.

The ancestry compositions captured by ASAP on real data show a significant correlation ($R=0.92$, $p<.0001$) with F4admix results obtained in the original paper (Lazaridis et al., 2022), confirming the reliability of ASAP in real-world scenarios (Fig. 7, Suppl. Table 6).

Moreover, we explored the individual ancestral composition of specific geographic locations in different time transects as in Lazaridis (Lazaridis et al., 2022). This enabled us to pinpoint the emergence of ancestral influences across different geographical regions and prehistoric periods. First, we examined the Anatolian region and confirmed an increase in Caucasus/Levantine ancestries around 3,000 BCE, accompanied by a subsequent reduction in local Anatolian ancestry (Fig. S1). Then, we confirmed the introduction of CHG-related ancestry into Steppe populations around 5,000 BCE, alongside the absence of Anatolian ancestry in this region prior to 3,000 BCE. We did not observe an increase in Levantine PPN ancestry, suggesting that most Eastern influence is associated with Anatolia Neolithic ancestry. Our approach corroborates again the complex genetic composition observed within the Yamnaya cluster, characterized by consistent CHG admixture (Fig. S2).

ASAP analysis further identified a less pronounced overrepresentation of CHG ancestry if compared to EHG ancestry in Aegean Bronze Age populations. This observation suggests significant gene flow occurring after the Neolithic period, particularly during the Early Bronze Age, across the Aegean and Balkan Peninsula regions (Fernandes et al., 2020; Raveane et al., 2022, 2019; Saupe et al., 2021) (Fig. S3). Similar trends were also observed in Italy, where Iron Age Southern Italian samples exhibited the highest frequency of Caucasus hunter-gatherer ancestry, found almost absent in Central Italian Etruscans Fig. S4 (Aneli et al., 2022).

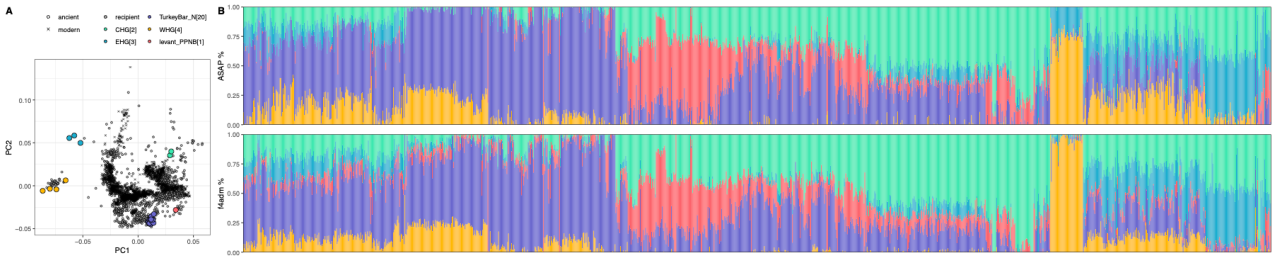


Fig. 7: Ancestry inference using ASAP on the aDNA dataset; **A)** PCA used as input by ASAP. **B)** Admixture plot displaying ancestry proportions for 1,350 ancient individuals (x-axis ordered by k-mean cluster numbers computed on ASAP inferred proportions): the upper panel has been estimated using ASAP, while the lower panel shows estimates reported in the original publication using F4admix (Lazaridis et al., 2022).

Although overall there is a high correlation between the two inferences, we observed 273 (out of 6,750) highly discordant estimates (HDE), in which the ancestral proportion difference exceeds 0.2.

1. When considering Western Hunter-Gatherer ancestry, we observed a correlation Pearson correlation coefficient R=0.9 (p-value<$1e^{-4}$) and 53 HDEs. Many of them include Hunter-Gatherers from Serbia and Romania, which are modeled by ASAP as approximately 70% WHG with the remaining ancestry mainly assigned to EHG, while 90% WHG and 10% EHG were estimated by Lazaridis et al. 2022 (Lazaridis et al., 2022). These samples were first published by Mathieson et al. 2018, who described them as a combination of WHG and EHG using qpAdm (although the estimates are associated with very low p-values) and D-statistics (Mathieson et al., 2018).
2. Concerning the ancestry of Turkey Barcin Neolithic individuals, commonly known as Anatolian Neolithic (AN) (Lazaridis et al., 2022), we observed a correlation R=0.95 (p < 1e-4) and 59 HDEs. Nevertheless, a few individuals exhibit a substantial discrepancy in AN ancestry proportion between the two compared methods, making it challenging to determine which of the two approaches has the highest performance.

   For example, ASAP estimates higher Anatolian Neolithic ancestry for some Mycenaean individuals (Lazaridis et al., 2017) while F4admix gives higher Anatolian ancestry for an Iron Age individual from Lebanon (Haber et al., 2020) (Suppl. Table 6). Both methods can be inaccurate in some cases, as shown by comparisons with previous studies.

3. For Iran Neolithic/CHG ancestry, we observed a correlation R of 0.95 (p-value<1e-4) and 47 HDEs. Most (20) are related to populations from Chalcolithic and Bronze/Iron Age Near East (Iran and Lebanon) individuals. For example, seven Bronze Age individuals from Shahr I Sokhta are modeled as having a substantially smaller Iran Neolithic/CHG ancestry for ASAP estimations (mean=0.65) compared to F4admix (mean=0.94). In Narasimhan et al. (Narasimhan et al., 2019), when Shahr I Sokhta individuals are modeled using qpAdm, they show on average 0.66 IN/CHG (SD=0.05).
4. In the case of EHG, we noted 40 HDEs and an R value of 0.86 (p-value<1e-4). As for WHG, most of the HDEs are related to Hunter-Gatherers from the Iron Gates regions of Serbia, Romania, for which ASAP estimates a higher proportion of EHG when compared to Lazaridis et al. (Lazaridis et al., 2022). Furthermore, in four Bell beaker individuals from Germany, France, and England, ASAP estimates a very low proportion of such ancestry.
5. We observed 74 HDEs and a correlation R of 0.88 (p-value<1e-4) for Levant Neolithic ancestry. Most of the HDEs are related to ancient individuals from the Near East, for which estimates of Levant PPN are always higher than those inferred by Lazaridis et al. (Lazaridis et al., 2022). These results align with previous estimates on the same samples. For example, for the individual I3832, which was modeled as 0.58 Levant PPN and 0.42 Iran Chalcolithic using LINADMIX in its original publication (Agranat-Tamir et al., 2020), ASAP estimated the Levant PPN proportion at 0.77, which was 0.38 when using F4admix (Lazaridis et al., 2022). A possible explanation for this discrepancy is related to the fact that in (Lazaridis et al., 2022), the same individual is modeled to have approximately ~0.2 related to AN. Similarly, ASAP's ancestral composition for individuals from Roman and Iron Age from Lebanon are in line with previous DyStruct inferences (Haber et al., 2020). Furthermore, F4admix (Lazaridis et al., 2017) estimated a substantial proportion of Levant PPN ancestry in two Greek and one Italian Bronze Age samples, in contrast with a series of findings on the same or similar individuals. All these samples are characterized by a missingness rate higher than 40% (I9006), suggesting that in some cases ASAP might be less biased than F4admix estimates.

We also used ASAP to test the robustness of the support behind the hypothesis that the WHG contributions of British farmers came mostly from continental WHG rather than local British WHG (Brace et al., 2019). Therefore, modeled different European Neolithic samples as a mixture of WHG and AN, as in Brace et al., 2019 (Brace et al., 2019). We confirmed that WHG proportions in Iberian Early Neolithic samples similar to those in British Neolithic samples, suggesting a common WHG source (Supp. Table 7). We then tested models with pairwise WHG individuals as possible sources and a single Anatolian Neolithic population (see method). We found that both Iberian and British samples consistently preferred Bichon or Villabruna-associated samples as WHG source, indicating their close relationship to the true source and confirming a shared origin (Supp. Table 8) (Fu et al., 2016).

Discussion

We present ASAP, a global ancestry exploration approach based on PCA and NNLS, that allows an accurate estimation of ancestry proportions in admixed groups or single samples. The approach leverages how the location of samples on the PC space can be related to the mean time of coalescence between pairs of samples (McVean, 2009), and to the recent observation that PC vectors are strongly related to *f4* metrics (Peter, 2022). Specifically, in the case of an admixture event, samples will fall along a gradient and their putative admixture sources will be placed at the ends of the gradient (McVean, 2009; Patterson et al., 2006). Our approach exploits the relative coordinates of the admixed samples and the ones of the putative sources in the PC space and summarises the ancestry proportions of the targets through NNLS.

An advantage of the method is that it can leverage the entire PC space, allowing a large panel of donors to model a given target admixed group or multiple parallel analyses in case several different target groups and their relative proxy sources are analyzed.

We demonstrated that the approach is highly accurate in most of the scenarios tested. Regardless of the availability of true or proxy admixture sources, our approach correctly assigned ancestral proportions, with a low associated error. Moreover, in the rare cases of error, ASAP appointed a group closely related (Fst ≤ 0.01) to the source.

Furthermore, ASAP can accurately assign ancestry proportions also in case of minor contributions as low as 10%. However, when only proxy sources are available it tends to overestimate the major component.

More importantly, ASAP performs well even when pseudo-haploid data with missing variants are analyzed, with a maximum assignment error of 3%. Specifically, when the admixture sources have a split time > 24 KYA, ASAP error estimates are lower than 0.014. On the other hand, for closely related admixture sources, the per-sample misassignment can reach 20% when multiple, closely related putative sources are available.

Compared to other global ancestry assignment tools, the approach is faster in terms of runtime while being as accurate (ChromoPainter) or more accurate than other tools (qpAdm) and, most importantly, provides ancestry estimates based on a straightforward formulation of user-defined ancestry sources with no need for in- or out-groups.

When tested on a real dataset of ancient and modern Eurasian genotypes, ASAP confirmed the trends in ancestry composition observed in previous research, providing relevant information on the complex scenario of the continent. Notably, it estimated significant gene flows after the Neolithic period in Aegean Bronze Age populations and confirmed previous findings about the shared origin of WHG ancestry in British and Iberian farmers.

Our approach relies on the assumption that the target group is indeed admixed. However, the target group might fall within a given cline for a demographic scenario different from admixture. This method should be used as an exploratory tool and subsequent analyses, such as formal tests for admixture, should be performed to test the admixture event further.

A future area of work is to explore and evaluate how the ASAP approach can also be applied to other summary statistics, such as IBS and/or Fst estimates, avoiding the evaluation of the number of PCs to be used.

Materials and Methods

DATASETS

Simulated dataset

- **Genotype data with no missingness**

We used a simulated genotype dataset from Molinaro et al. composed of 13 simulated demes with different population sizes and split times ranging from 250 to 4,000 generations, to represent a simplified scenario for current European (EUR 1-3), East Asian (ASN 1-3) and African (AFR 1-7) groups and 7 sister groups characterized by a split time from their closest population of 100 generations (Molinaro et al., 2021). The data simulation was carried out with mspms and following a modified Van Dorp et al. model (Kelleher et al., 2016; van Dorp et al., 2015). The initial dataset consisted of eight admixed groups, obtained by combining pairs of the simulated Ghost populations (GST), all with ancestry proportions of 70%-30%. The pairs of admixing GST were selected in order to cover a broad spectrum of split times. Specifically, we simulated admixture groups whose sources split time span from 75 KYA to 9 KYA, six sources shared a bottleneck event and for three of these, we simulated an additional one. The initial set also comprised one admixed group characterized by a three-way admixture with the proportions of 60%-20%-20%, with African-like, European-like, and Asian-like ancestries, respectively.

We simulated an additional three-way admixture group, using the same highly divergent sources as above, but different ancestry proportions, namely: 40% for the African-like ancestry, and 20% for both the European and Asian-like ancestries. To test models with strongly imbalanced ancestry proportions, we also simulated three groups with 90%-10% and three groups with 95%-05% ancestry proportions. In this case, as well, we chose the admixture sources (GST) to cover a broad spectrum of split times.

All admixture simulations were carried out with admix-simu (https://github.com/williamslab/admix-simu), creating 50 individuals per each population, using a constant recombination rate (1.25×10^{-8}) and admixture time of 100 generations (Molinaro et al., 2021). We obtained a simulated dataset of 4,745,025 SNPs, 20 non-admixed, and 16 admixed groups. After filtering for minor allele frequency with PLINK (maf 0.01), the final dataset comprised 284.249 SNPs (Chang

et al., 2015). PC analyses were performed on the final dataset, projecting the admixed target samples on the scaffold built from the non-admixed groups.

- **Ancient (Pseudo-haploid) data**

To mimic the data quality of ancient DNA, we manipulated the simulated set by introducing both missing data as well as using pseudo-haploid genotypes. In each population, we introduced a variable missing rate (from 10% to 50%) in randomly selected positions, so every 10 individuals would be characterized by 10, 20, 30, 40 or 50% of missing data. Secondly, we created pseudo-haploid genomes by randomly selecting at each locus one allele and assigning it as a homozygous genotype, eventually obtaining for each simulated group 100 pseudo-haploid genomes from the original 50 diploid individuals. The missingness proportions were maintained after pruning.

PCA was performed after filtering for minor allele frequency (maf 0.01). For the pseudo-haploid datasets containing missing data, pruning was also performed (PLINK v1.9 indep-pairwise 50 10 0.1)(Chang et al., 2015).
After the filtering, the bim file of the modern simulated dataset contains 284.249 SNPs, the one in which only the real sources are pseudo-haploid has 135.211 SNPs, while the one in which the sources are also pseudo-haploid has roughly 100.000 SNPs.

- **Real modern and ancient dataset**

We downloaded 1240K+HO dataset (version V52.2, https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data) in EIGENSTRAT format. Such dataset includes present-day and ancient DNA data and converted it into PLINK format using convertf (Patterson et al., 2006).
Starting from the .anno file and following Aneli et al. method (Aneli et al., 2021), we created a list of ancient and modern samples to keep from the 1240K+HO dataset.
In particular, only the ancient ones (those that in the “Full date “column did not have the string "present") coming from Western Eurasian countries (latitudes higher than 22 and longitudes between − 15 and 60) and the Mbuti individuals from Congo (string “Mbuti” in the “Genetic.ID” column) (Bergström et al., 2020; Patterson et al., 2012)

were selected. Then we removed individuals that in the “Group.ID” column have the string “Ignore”. Finally, we kept only those whose “Assessment” column contained the string “PASS”. After removing any duplicates, we created a preliminary plink file with only 1240K+HO ancient samples to whom we added other ancient samples taken from other published datasets (Aneli et al., 2022; Lazaridis et al., 2022; Posth et al., 2021; Reitsema et al., 2022).

For the modern samples, we selected individuals coming from Western Eurasian countries with latitudes higher than 22 and longitudes between − 15 and 60, but removing those coming from Uzbekistan, Kazakhstan, Algeria, Morocco, Tunisia, Libya, as well as some populations from Russia and others showing “Ignore” within “Group.ID”. Then, we selected the samples flagged as “PASS” in the “Assessment” column. In this way, we created a preliminary plink file with only the 1240K+HO modern individuals to which we merged other modern-day samples taken from the Raveane et al. dataset (Raveane et al., 2019).

From this modern dataset, we extracted only the autosome chromosome SNPs and excluded those monomorphic (--maf .00001) and with more than 5% missing data (--geno .05) using PLINK (Chang et al., 2015).

Then we extracted the bulk of variants built on our modern dataset from the ancient dataset and finally merged all with PLINK1.9 excluding the ancient samples with less than 20,000 SNPs (N=1,381). To assess the relatedness, we computed the kinship coefficient pi-hat, expressing the probability that two randomly selected alleles, belonging to the same locus, are identical-by-descent (IBD) for any two samples. Using --genome --min 0.35 in PLINK1.9 we selected pairs of samples showing a pi-hat of at least 0.35, retaining 7,312 individuals. We then filtered out samples with duplicated IDs retaining the one with a higher number of SNPs. We carried out a Principal Component Analysis (PCA) using smartpca, through which we filtered out outliers obtaining a dataset of 6,408 samples. Our final real dataset, containing 4,740 ancient and 1,668 modern samples with 206,363 SNPs, was converted to the EIGENSTRAT format using convertf.

Principal Component Analysis (PCA)

PCA was performed using smartpca (version 16000) available in the EIGENSOFT package (Patterson et al., 2012).

The admixed (or target) populations were always projected, regardless of the dataset used. In the case of datasets with pseudo-haploid or ancient individuals, we projected all samples' genotypes onto the principal components inferred from the diploid/modern individuals using the lsqproject: YES option. For each analysis on the simulated and real genotype datasets, we run 10 PCs. Furthermore, we performed the same analysis using different numbers of PCs, in order to assess the performance of ASAP (see Supplementary Text 1)

Non Negative Least Squares (NNLS)

To perform the Non-Negative Least Squares, we used the NNLS function, as described in (Hellenthal et al., 2014; Leslie et al., 2015; Ongaro et al., 2019), which is an adaptation of the Lawson–Hanson NNLS implementation of Non-Negative Least Squares (Lawson and Hanson, 1995) available in the statistical software package R 3.5.1 (R Core Team, 2020).

We applied NNLS both population-wise and individual-wise. To estimate the NNLS population-wise, we estimated the average of each of the population PCs and then applied NNLS on the resulting vector. On the other hand, to estimate the NNLS individual-wise, the PCs of each individual were maintained as separate vectors. Error values reported in the text were calculated as the absolute average difference between the observed and the expected proportion assignment. We used a block jackknife approach to resampling our set and estimate the standard errors. Given that the simulated set consisted of only chromosome 1, we could not use chromosomes as blocks, as usually it's done when the entire genome is available. We thus estimated the number of SNPs available after filtering and divided them into 20 blocks. For each resampling step, we removed one of such blocks and performed PCA on the remaining ones.

Standard errors were estimated on chromosome-based jackknife replications (Büsing et al., 2011).

qpAdm

To validate the results obtained from ASAP, we performed the most widely used approach to assess the ancestry components and the relative proportions of the admixed population: qpAdm programs in the ADMIXTOOLS package (Patterson et al., 2012).

For each admixed individual, we tested as "left" populations all the possible true sources and used all the others as right populations. Subsequently, we selected the inference characterized by the largest p-value, irrespectively to their significance. Although there are many ways to harness qpAdm to obtain more reliable results, we decided to use a strategy comparable to the other tools harnessed here.

Rye

We applied Rye (Conley et al., 2023) converting the PCA output obtained by smartpca using a custom R script. We performed five different rounds using the first ten PCs.

ChromoPainter

ChromoPainter (CP) (Lawson et al., 2012) was applied using the unlinked (-u) model, where, for each SNP in the target, we assign a score of 1/K to each reference haplotype that carries the same allele, where K is the total number of reference haplotypes that carry the same allele.

Analysis of Eurasian ancient and modern genotype data

We carried out a PCA computing 10 Principal Components (PCs) per each individual in our final dataset projecting ancient samples on the top of present-day genome variability (N = 1668). We used smartpca version 16000 with autoshrink lsqproject options for this analysis. Subsequently, we selected ancient individuals previously analyzed in Lazaridis et al., 2022 (Lazaridis et al., 2022) and filtered out samples with less than 180K SNPs. This resulted in a dataset comprising 1,380 ancient donors, of which 30 were chosen as recipients for ASAP. The selection of recipient samples was based on the five main ancestral sources identified in Lazaridis 2022 (Lazaridis et al., 2022), namely Western Hunter-Gatherers (WHG), Eastern Hunter-Gatherers (EHG), Caucasus Hunter-Gatherers (CHG), Anatolian Neolithic, and Levant Neolithic. ASAP

was run using 10 PCs, and the estimates were correlated with F4admix results using all samples combined, as well as stratified by different ancestral sources. Pearson correlation analysis was performed using the ggpubr library in R. To explore ancestry over time, we visualized single ancestry trends by either selecting individuals as indicated in the publication or by visually inspecting populations present in (Lazaridis et al., 2022) figures.

Acknowledgements

LDG and MV were supported by #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006) – A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022). FM was supported by Fondazione con il Sud (2018-PDR-01136) and by the Italian Ministry of University and Research (2022P2ZESR). MV was supported by the Italian Ministry of University and Research (2022E8NN2N). FS is a PhD student within the European School of Molecular Medicine (SEMM). GH was supported by the Wellcome Trust (224575/Z/21/Z). LP is funded by the Italian Ministry of University and Research (PRIN 2022B27XYM). LM and TK were supported by KU Leuven BOF-C24 grant ZKD6488 C24M/19/075 and FWO grant G0A4521N (TK).

We would like to thank Nicole Soranzo for advice on the final stage of the manuscript preparation.

Bibliography

Agranat-Tamir, L., Waldman, S., Martin, M.A.S., Gokhman, D., Mishol, N., Eshel, T., Cheronet, O., Rohland, N., Mallick, S., Adamski, N., Lawson, A.M., Mah, M., Michel, M., Oppenheimer, J., Stewardson, K., Candilio, F., Keating, D., Gamarra, B., Tzur, S., Novak, M., Kalisher, R., Bechar, S., Eshed, V., Kennett, D.J., Faerman, M., Yahalom-Mack, N., Monge, J.M., Govrin, Y., Erel, Y., Yakir, B., Pinhasi, R., Carmi, S., Finkelstein, I., Carmel, L., Reich, D., 2020. The Genomic History of the Bronze Age Southern Levant. Cell 181, 1146-1157.e11. https://doi.org/10.1016/j.cell.2020.04.024

Aneli, S., Caldon, M., Saupe, T., Montinaro, F., Pagani, L., 2021. Through 40,000 years of human presence in Southern Europe: the Italian case study. Hum. Genet. 140, 1417–1431. https://doi.org/10.1007/s00439-021-02328-6

Aneli, S., Saupe, T., Montinaro, F., Solnik, A., Molinaro, L., Scaggion, C., Carrara, N., Raveane, A., Kivisild, T., Metspalu, M., Scheib, C.L., Pagani, L., 2022. The Genetic Origin of Daunians and the Pan-Mediterranean Southern Italian Iron Age Context. Mol. Biol. Evol. 39, msac014. https://doi.org/10.1093/molbev/msac014

Bergström, A., McCarthy, S.A., Hui, R., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., Blanché, H., Deleuze, J.-F., Cann, H., Mallick, S., Reich, D., Sandhu, M.S., Skoglund, P., Scally, A., Xue, Y., Durbin, R., Tyler-Smith, C., 2020. Insights into human genetic variation and population history from 929 diverse genomes. Science 367, eaay5012. https://doi.org/10.1126/science.aay5012

Brace, S., Diekmann, Y., Booth, T.J., van Dorp, L., Faltyskova, Z., Rohland, N., Mallick, S., Olalde, I., Ferry, M., Michel, M., Oppenheimer, J., Broomandkhoshbacht, N., Stewardson, K., Martiniano, R., Walsh, S., Kayser, M., Charlton, S., Hellenthal, G., Armit, I., Schulting, R., Craig, O.E., Sheridan, A., Parker Pearson, M., Stringer, C., Reich, D., Thomas, M.G., Barnes, I., 2019. Ancient genomes indicate population replacement in Early Neolithic Britain. Nat. Ecol. Evol. 3, 765–771. https://doi.org/10.1038/s41559-019-0871-9

Busby, G.B., Band, G., Si Le, Q., Jallow, M., Bougama, E., Mangano, V.D., Amenga-Etego, L.N., Enimil, A., Apinjoh, T., Ndila, C.M., Manjurano, A., Nyirongo, V., Doumba, O., Rockett, K.A., Kwiatkowski, D.P., Spencer, C.C., Network, M.G.E., 2016. Admixture into and within sub-Saharan Africa. eLife 5, e15266. https://doi.org/10.7554/eLife.15266

Büsing, C., Koster, A.M.C.A., Kutschka, M., 2011. Recoverable robust knapsacks: the discrete scenario case. Optim. Lett. 5, 379–392. https://doi.org/10.1007/s11590-011-0307-1

Busing, F.M.T.A., Meijer, E., Leeden, R.V.D., 1999. Delete-m Jackknife for Unequal m. Stat. Comput. 9, 3–8. https://doi.org/10.1023/A:1008800423698

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4. https://doi.org/10.1186/s13742-015-0047-8

Conley, A.B., Rishishwar, L., Ahmad, M., Sharma, S., Norris, E.T., Jordan, I.K., Mariño-Ramírez, L., 2023. Rye: genetic ancestry inference at biobank scale. Nucleic Acids Res. 51, e44. https://doi.org/10.1093/nar/gkad149

Damgaard, P. de B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E., Baimukhanov, N., Loman, V., Hedeager, L., Pedersen, A.G., Nielsen, K., Afanasiev, G., Akmatov, K., Aldashev, A., Alpaslan, A., Baimbetov, G., Bazaliiskii, V.I., Beisenov, A., Boldbaatar, B., Boldgiv, B., Dorzhu, C., Ellingvag, S., Erdenebaatar, D., Dajani, R., Dmitriev, E., Evdokimov, V., Frei, K.M., Gromov, A., Goryachev, A., Hakonarson, H., Hegay, T., Khachatryan, Z., Khaskhanov, R., Kitov, E., Kolbina, A., Kubatbek, T., Kukushkin, A., Kukushkin, I., Lau, N., Margaryan, A., Merkyte, I., Mertz, I.V., Mertz,

V.K., Mijiddorj, E., Moiyesev, V., Mukhtarova, G., Nurmukhanbetov, B., Orozbekova, Z., Panyushkina, I., Pieta, K., Smrčka, V., Shevnina, I., Logvin, A., Sjögren, K.-G., Štolcová, T., Taravella, A.M., Tashbaeva, K., Tkachev, A., Tulegenov, T., Voyakin, D., Yepiskoposyan, L., Undrakhbold, S., Varfolomeev, V., Weber, A., Wilson Sayres, M.A., Kradin, N., Allentoft, M.E., Orlando, L., Nielsen, R., Sikora, M., Heyer, E., Kristiansen, K., Willerslev, E., 2018. 137 ancient human genomes from across the Eurasian steppes. Nature 557, 369–374. https://doi.org/10.1038/s41586-018-0094-2

Eren Yüncü, Ulaş Işıldak, Matthew P. Williams, Christian D. Huber, Olga Flegontova, Leonid A. Vyazov, Piya Changmai, Pavel Flegontov, 2023. False discovery rates of qpAdm-based screens for genetic admixture. bioRxiv 2023.04.25.538339. https://doi.org/10.1101/2023.04.25.538339

Fernandes, D.M., Mittnik, A., Olalde, I., Lazaridis, I., Cheronet, O., Rohland, N., Mallick, S., Bernardos, R., Broomandkhoshbacht, N., Carlsson, J., Culleton, B.J., Ferry, M., Gamarra, B., Lari, M., Mah, M., Michel, M., Modi, A., Novak, M., Oppenheimer, J., Sirak, K.A., Stewardson, K., Mandl, K., Schattke, C., Özdoğan, K.T., Lucci, M., Gasperetti, G., Candilio, F., Salis, G., Vai, S., Camarós, E., Calò, C., Catalano, G., Cueto, M., Forgia, V., Lozano, M., Marini, E., Micheletti, M., Miccichè, R.M., Palombo, M.R., Ramis, D., Schimmenti, V., Sureda, P., Teira, L., Teschler-Nicola, M., Kennett, D.J., Lalueza-Fox, C., Patterson, N., Sineo, L., Coppa, A., Caramelli, D., Pinhasi, R., Reich, D., 2020. The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. Nat. Ecol. Evol. 4, 334–345. https://doi.org/10.1038/s41559-020-1102-0

Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., Nickel, B., Peltzer, A., Rohland, N., Slon, V., Talamo, S., Lazaridis, I., Lipson, M., Mathieson, I., Schiffels, S., Skoglund, P., Derevianko, A.P., Drozdov, N., Slavinsky, V., Tsybankov, A., Cremonesi, R.G., Mallegni, F., Gély, B., Vacca, E., Morales, M.R.G., Straus, L.G., Neugebauer-Maresch, C., Teschler-Nicola, M., Constantin, S., Moldovan, O.T., Benazzi, S., Peresani, M., Coppola, D., Lari, M., Ricci, S., Ronchitelli, A., Valentin, F., Thevenet, C., Wehrberger, K., Grigorescu, D., Rougier, H., Crevecoeur, I., Flas, D., Semal, P., Mannino, M.A., Cupillard, C., Bocherens, H., Conard, N.J., Harvati, K., Moiseyev, V., Drucker, D.G., Svoboda, J., Richards, M.P., Caramelli, D., Pinhasi, R., Kelso, J., Patterson, N., Krause, J., Pääbo, S., Reich, D., 2016. The genetic history of Ice Age Europe. Nature 534, 200–205. https://doi.org/10.1038/nature17993

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Bánffy, E., Economou, C., Francken, M., Friederich, S., Pena, R.G., Hallgren, F., Khartanovich, V., Khokhlov, A., Kunst, M., Kuznetsov, P., Meller, H., Mochalov, O., Moiseyev, V., Nicklisch, N., Pichler, S.L., Risch, R., Rojo Guerra, M.A., Roth, C., Szécsényi-Nagy, A., Wahl, J., Meyer, M., Krause, J., Brown, D., Anthony, D., Cooper, A., Alt, K.W., Reich, D., 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522, 207–211. https://doi.org/10.1038/nature14317

Haber, M., Doumet-Serhal, C., Scheib, C., Xue, Y., Danecek, P., Mezzavilla, M., Youhanna, S., Martiniano, R., Prado-Martinez, J., Szpak, M., Matisoo-Smith, E., Schutkowski, H., Mikulski, R., Zalloua, P., Kivisild, T., Tyler-Smith, C., 2017. Continuity and Admixture in the Last Five Millennia of Levantine History from Ancient Canaanite and Present-Day Lebanese Genome Sequences. Am. J. Hum. Genet. 101, 274–282. https://doi.org/10.1016/j.ajhg.2017.06.013

Haber, M., Nassar, J., Almarri, M.A., Saupe, T., Saag, L., Griffith, S.J., Doumet-Serhal, C., Chanteau, J., Saghieh-Beydoun, M., Xue, Y., Scheib, C.L., Tyler-Smith, C., 2020. A Genetic History of the Near East from an aDNA Time Course Sampling Eight Points in the Past 4,000 Years. Am. J. Hum. Genet. 107, 149–157. https://doi.org/10.1016/j.ajhg.2020.05.008

Hajdinjak, M., Fu, Q., Hübner, A., Petr, M., Mafessoni, F., Grote, S., Skoglund, P., Narasimham, V., Rougier, H., Crevecoeur, I., Semal, P., Soressi, M., Talamo, S.,

Hublin, J.-J., Gušić, I., Kućan, Ž., Rudan, P., Golovanova, L.V., Doronichev, V.B., Posth, C., Krause, J., Korlević, P., Nagel, S., Nickel, B., Slatkin, M., Patterson, N., Reich, D., Prüfer, K., Meyer, M., Pääbo, S., Kelso, J., 2018. Reconstructing the genetic history of late Neanderthals. Nature 555, 652–656. https://doi.org/10.1038/nature26151

Harney, É., May, H., Shalem, D., Rohland, N., Mallick, S., Lazaridis, I., Sarig, R., Stewardson, K., Nordenfelt, S., Patterson, N., Hershkovitz, I., Reich, D., 2018. Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. Nat. Commun. 9, 3336. https://doi.org/10.1038/s41467-018-05649-9

Harney, É., Patterson, N., Reich, D., Wakeley, J., 2021. Assessing the performance of qpAdm: a statistical tool for studying population admixture. Genetics 217, iyaa045. https://doi.org/10.1093/genetics/iyaa045

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, 747–751. https://doi.org/10.1126/science.1243518

Kelleher, J., Etheridge, A.M., McVean, G., 2016. Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. PLOS Comput. Biol. 12, e1004842. https://doi.org/10.1371/journal.pcbi.1004842

Lawson, C.L., Hanson, R.J., 1995. Solving Least Squares Problems. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611971217

Lawson, D.J., Hellenthal, G., Myers, S., Falush, D., 2012. Inference of population structure using dense haplotype data. PLoS Genet 8, e1002453. https://doi.org/10.1371/journal.pgen.1002453

Lazaridis, I., Alpaslan-Roodenberg, S., Acar, A., Açıkkol, A., Agelarakis, A., Aghikyan, L., Akyüz, U., Andreeva, D., Andrijašević, G., Antonović, D., Armit, I., Atmaca, A., Avetisyan, P., Aytek, A.İ., Bacvarov, K., Badalyan, R., Bakardzhiev, S., Balen, J., Bejko, L., Bernardos, R., Bertsatos, A., Biber, H., Bilir, A., Bodružić, M., Bonogofsky, M., Bonsall, C., Borić, D., Borovinić, N., Morante, G.B., Buttinger, K., Callan, K., Candilio, F., Carić, M., Cheronet, O., Chohadzhiev, S., Chovalopoulou, M.-E., Chryssoulaki, S., Ciobanu, I., Čondić, N., Constantinescu, M., Cristiani, E., Culleton, B.J., Curtis, E., Davis, J., Davtyan, R., Demcenco, T.I., Dergachev, V., Derin, Z., Deskaj, S., Devejyan, S., Djordjević, V., Carlson, K.S.D., Eccles, L.R., Elenski, N., Engin, A., Erdoğan, N., Erir-Pazarcı, S., Fernandes, D.M., Ferry, M., Freilich, S., Frînculeasa, A., Galaty, M.L., Gamarra, B., Gasparyan, B., Gaydarska, B., Genç, E., Gültekin, T., Gündüz, S., Hajdu, T., Heyd, V., Hobosyan, S., Hovhannisyan, N., Iliev, I., Iliev, L., Iliev, S., İvgin, İ., Janković, I., Jovanova, L., Karkanas, P., Kavaz-Kındığılı, B., Kaya, E.H., Keating, D., Kennett, D.J., Kesici, S.D., Khudaverdyan, A., Kiss, K., Kılıç, S., Klostermann, P., Valdes, S.K.B.N., Kovačević, S., Krenz-Niedbała, M., Škrivanko, M.K., Kurti, R., Kuzman, P., Lawson, A.M., Lazar, C., Leshtakov, K., Levy, T.E., Liritzis, I., Lorentz, K.O., Łukasik, S., Mah, M., Mallick, S., Mandl, K., Martirosyan-Olshansky, K., Matthews, R., Matthews, W., McSweeney, K., Melikyan, V., Micco, A., Michel, M., Milašinović, L., Mittnik, A., Monge, J.M., Nekhrizov, G., Nicholls, R., Nikitin, A.G., Nikolov, V., Novak, M., Olalde, I., Oppenheimer, J., Osterholtz, A., Özdemir, C., Özdoğan, K.T., Öztürk, N., Papadimitriou, N., Papakonstantinou, N., Papathanasiou, A., Paraman, L., Paskary, E.G., Patterson, N., Petrakiev, I., Petrosyan, L., Petrova, V., Philippa-Touchais, A., Piliposyan, A., Kuzman, N.P., Potrebica, H., Preda-Bălănică, B., Premužić, Z., Price, T.D., Qiu, L., Radović, S., Aziz, K.R., Šikanjić, P.R., Raheem, K.R., Razumov, S., Richardson, A., Roodenberg, J., Ruka, R., Russeva, V., Şahin, M., Şarbak, A., Savaş, E., Schattke, C., Schepartz, L., Selçuk, T., Sevim-Erol, A., Shamoon-Pour, M., Shephard, H.M., Sideris, A., Simalcsik, A., Simonyan, H., Sinika, V., Sirak, K., Sirbu, G., Šlaus, M., Soficaru, A., Söğüt, B., Sołtysiak, A., Sönmez-Sözer, Ç., Stathi, M., Steskal, M., Stewardson, K., Stocker, S., Suata-Alpaslan, F., Suvorov, A., Szécsényi-Nagy, A., Szeniczey, T., Telnov, N., Temov, S., Todorova, N., Tota, U., Touchais, G., Triantaphyllou, S., Türker, A., Ugarković, M., Valchev, T., Veljanovska, F., Videvski,

Z., Virag, C., Wagner, A., Walsh, S., Włodarczak, P., Workman, J.N., Yardumian, A., Yarovoy, E., Yavuz, A.Y., Yılmaz, H., Zalzala, F., Zettl, A., Zhang, Z., Çavuşoğlu, R., Rohland, N., Pinhasi, R., Reich, D., 2022. The genetic history of the Southern Arc: A bridge between West Asia and Europe. Science 377, eabm4247. https://doi.org/10.1126/science.abm4247

Lazaridis, I., Mittnik, A., Patterson, N., Mallick, S., Rohland, N., Pfrengle, S., Furtwängler, A., Peltzer, A., Posth, C., Vasilakis, A., McGeorge, P.J.P., Konsolaki-Yannopoulou, E., Korres, G., Martlew, H., Michalodimitrakis, M., Özsait, M., Özsait, N., Papathanasiou, A., Richards, M., Roodenberg, S.A., Tzedakis, Y., Arnott, R., Fernandes, D.M., Hughey, J.R., Lotakis, D.M., Navas, P.A., Maniatis, Y., Stamatoyannopoulos, J.A., Stewardson, K., Stockhammer, P., Pinhasi, R., Reich, D., Krause, J., Stamatoyannopoulos, G., 2017. Genetic origins of the Minoans and Mycenaeans. Nature 548, 214–218. https://doi.org/10.1038/nature23310

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., Connell, S., Stewardson, K., Harney, E., Fu, Q., Gonzalez-Fortes, G., Jones, E.R., Roodenberg, S.A., Lengyel, G., Bocquentin, F., Gasparian, B., Monge, J.M., Gregg, M., Eshed, V., Mizrahi, A.-S., Meiklejohn, C., Gerritsen, F., Bejenaru, L., Blüher, M., Campbell, A., Cavalleri, G., Comas, D., Froguel, P., Gilbert, E., Kerr, S.M., Kovacs, P., Krause, J., McGettigan, D., Merrigan, M., Merriwether, D.A., O'Reilly, S., Richards, M.B., Semino, O., Shamoon-Pour, M., Stefanescu, G., Stumvoll, M., Tönjes, A., Torroni, A., Wilson, J.F., Yengo, L., Hovhannisyan, N.A., Patterson, N., Pinhasi, R., Reich, D., 2016. Genomic insights into the origin of farming in the ancient Near East. Nature 536, 419–424. https://doi.org/10.1038/nature19310

Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E.C., Cunliffe, B., Lawson, D.J., Falush, D., Freeman, C., Pirinen, M., Myers, S., Robinson, M., Donnelly, P., Bodmer, W., Wellcome Trust Case Control Consortium 2, International Multiple Sclerosis Genetics Consortium, 2015. The fine-scale genetic structure of the British population. Nature 519, 309–314. https://doi.org/10.1038/nature14230

Li, N., Stephens, M., 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165, 2213–2233. https://doi.org/10.1093/genetics/165.4.2213

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., Fernandes, D., Ferry, M., Gamarra, B., Fortes, G.G., Haak, W., Harney, E., Jones, E., Keating, D., Krause-Kyora, B., Kucukkalipci, I., Michel, M., Mittnik, A., Nägele, K., Novak, M., Oppenheimer, J., Patterson, N., Pfrengle, S., Sirak, K., Stewardson, K., Vai, S., Alexandrov, S., Alt, K.W., Andreescu, R., Antonović, D., Ash, A., Atanassova, N., Bacvarov, K., Gusztáv, M.B., Bocherens, H., Bolus, M., Boroneanţ, A., Boyadzhiev, Y., Budnik, A., Burmaz, J., Chohadzhiev, S., Conard, N.J., Cottiaux, R., Čuka, M., Cupillard, C., Drucker, D.G., Elenski, N., Francken, M., Galabova, B., Ganetsovski, G., Gély, B., Hajdu, T., Handzhyiska, V., Harvati, K., Higham, T., Iliev, S., Janković, I., Karavanić, I., Kennett, D.J., Komšo, D., Kozak, A., Labuda, D., Lari, M., Lazar, C., Leppek, M., Leshtakov, K., Vetro, D.L., Los, D., Lozanov, I., Malina, M., Martini, F., McSweeney, K., Meller, H., Menđušić, M., Mirea, P., Moiseyev, V., Petrova, V., Price, T.D., Simalcsik, A., Sineo, L., Šlaus, M., Slavchev, V., Stanev, P., Starović, A., Szeniczey, T., Talamo, S., Teschler-Nicola, M., Thevenet, C., Valchev, I., Valentin, F., Vasilyev, S., Veljanovska, F., Venelinova, S., Veselovskaya, E., Viola, B., Virag, C., Zaninović, J., Zäuner, S., Stockhammer, P.W., Catalano, G., Krauß, R., Caramelli, D., Zariņa, G., Gaydarska, B., Lillie, M., Nikitin, A.G., Potekhina, I., Papathanasiou, A., Borić, D., Bonsall, C., Krause, J., Pinhasi, R., Reich, D., 2018. The genomic history of southeastern Europe. Nature 555, 197–203. https://doi.org/10.1038/nature25778

McVean, G., 2009. A genealogical interpretation of principal components analysis. PLoS

Genet 5, e1000686. https://doi.org/10.1371/journal.pgen.1000686

Molinaro, L., Marnetto, D., Mondal, M., Ongaro, L., Yelmen, B., Lawson, D.J., Montinaro, F., Pagani, L., 2021. A Chromosome-Painting-Based Pipeline to Infer Local Ancestry under Limited Source Availability. Genome Biol. Evol. 13, evab025. https://doi.org/10.1093/gbe/evab025

Montinaro, F., Busby, G.B.J., Pascali, V.L., Myers, S., Hellenthal, G., Capelli, C., 2015. Unravelling the hidden ancestry of American admixed populations. Nat. Commun. 6, 6596. https://doi.org/10.1038/ncomms7596

Moorjani, P., Patterson, N., Hirschhorn, J.N., Keinan, A., Hao, L., Atzmon, G., Burns, E., Ostrer, H., Price, A.L., Reich, D., 2011. The History of African Gene Flow into Southern Europeans, Levantines, and Jews. PLOS Genet. 7, e1001373. https://doi.org/10.1371/journal.pgen.1001373

Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M., Kim, A.M., Olivieri, L.M., Coppa, A., Vidale, M., Mallory, J., Moiseyev, V., Kitov, E., Monge, J., Adamski, N., Alex, N., Broomandkhoshbacht, N., Candilio, F., Callan, K., Cheronet, O., Culleton, B.J., Ferry, M., Fernandes, D., Freilich, S., Gamarra, B., Gaudio, D., Hajdinjak, M., Harney, É., Harper, T.K., Keating, D., Lawson, A.M., Mah, M., Mandl, K., Michel, M., Novak, M., Oppenheimer, J., Rai, N., Sirak, K., Slon, V., Stewardson, K., Zalzala, F., Zhang, Z., Akhatov, G., Bagashev, A.N., Bagnera, A., Baitanayev, B., Bendezu-Sarmiento, J., Bissembaev, A.A., Bonora, G.L., Chargynov, T.T., Chikisheva, T., Dashkovskiy, P.K., Derevianko, A., Dobeš, M., Douka, K., Dubova, N., Duisengali, M.N., Enshin, D., Epimakhov, A., Fribus, A.V., Fuller, D., Goryachev, A., Gromov, A., Grushin, S.P., Hanks, B., Judd, M., Kazizov, E., Khokhlov, A., Krygin, A.P., Kupriyanova, E., Kuznetsov, P., Luiselli, D., Maksudov, F., Mamedov, A.M., Mamirov, T.B., Meiklejohn, C., Merrett, D.C., Micheli, R., Mochalov, O., Mustafokulov, S., Nayak, A., Pettener, D., Potts, R., Razhev, D., Rykun, M., Sarno, S., Savenkova, T.M., Sikhymbaeva, K., Slepchenko, S.M., Soltobaev, O.A., Stepanova, N., Svyatko, S., Tabaldiev, K., Teschler-Nicola, M., Tishkin, A.A., Tkachev, V.V., Vasilyev, S., Velemínský, P., Voyakin, D., Yermolayeva, A., Zahir, M., Zubkov, V.S., Zubova, A., Shinde, V.S., Lalueza-Fox, C., Meyer, M., Anthony, D., Boivin, N., Thangaraj, K., Kennett, D.J., Frachetti, M., Pinhasi, R., Reich, D., 2019. The formation of human populations in South and Central Asia. Science 365, eaat7487. https://doi.org/10.1126/science.aat7487

Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., Altena, E., Lipson, M., Lazaridis, I., Harper, T.K., Patterson, N., Broomandkhoshbacht, N., Diekmann, Y., Faltyskova, Z., Fernandes, D., Ferry, M., Harney, E., de Knijff, P., Michel, M., Oppenheimer, J., Stewardson, K., Barclay, A., Alt, K.W., Liesau, C., Ríos, P., Blasco, C., Miguel, J.V., García, R.M., Fernández, A.A., Bánffy, E., Bernabò-Brea, M., Billoin, D., Bonsall, C., Bonsall, L., Allen, T., Büster, L., Carver, S., Navarro, L.C., Craig, O.E., Cook, G.T., Cunliffe, B., Denaire, A., Dinwiddy, K.E., Dodwell, N., Ernée, M., Evans, C., Kuchařík, M., Farré, J.F., Fowler, C., Gazenbeek, M., Pena, R.G., Haber-Uriarte, M., Haduch, E., Hey, G., Jowett, N., Knowles, T., Massy, K., Pfrengle, S., Lefranc, P., Lemercier, O., Lefebvre, A., Martínez, C.H., Olmo, V.G., Ramírez, A.B., Maurandi, J.L., Majó, T., McKinley, J.I., McSweeney, K., Mende, B.G., Modi, A., Kulcsár, G., Kiss, V., Czene, A., Patay, R., Endrődi, A., Köhler, K., Hajdu, T., Szeniczey, T., Dani, J., Bernert, Z., Hoole, M., Cheronet, O., Keating, D., Velemínský, P., Dobeš, M., Candilio, F., Brown, F., Fernández, R.F., Herrero-Corral, A.-M., Tusa, S., Carnieri, E., Lentini, L., Valenti, A., Zanini, A., Waddington, C., Delibes, G., Guerra-Doce, E., Neil, B., Brittain, M., Luke, M., Mortimer, R., Desideri, J., Besse, M., Brücken, G., Furmanek, M., Hałuszko, A., Mackiewicz, M., Rapiński, A., Leach, S., Soriano, I., Lillios, K.T., Cardoso, J.L., Pearson, M.P., Włodarczak, P., Price, T.D., Prieto, P., Rey, P.-J., Risch, R., Rojo Guerra, M.A., Schmitt, A., Serralongue, J., Silva, A.M., Smrčka, V., Vergnaud, L., Zilhão, J., Caramelli, D., Higham, T., Thomas, M.G., Kennett, D.J., Fokkens, H.,

Heyd, V., Sheridan, A., Sjögren, K.-G., Stockhammer, P.W., Krause, J., Pinhasi, R., Haak, W., Barnes, I., Lalueza-Fox, C., Reich, D., 2018. The Beaker phenomenon and the genomic transformation of northwest Europe. Nature 555, 190–196. https://doi.org/10.1038/nature25738

Ongaro, L., Scliar, M.O., Flores, R., Raveane, A., Marnetto, D., Sarno, S., Gnecchi-Ruscone, G.A., Alarcón-Riquelme, M.E., Patin, E., Wangkumhang, P., Hellenthal, G., Gonzalez-Santos, M., King, R.J., Kouvatsi, A., Balanovsky, O., Balanovska, E., Atramentova, L., Turdikulova, S., Mastana, S., Marjanovic, D., Mulahasanovic, L., Leskovac, A., Lima-Costa, M.F., Pereira, A.C., Barreto, M.L., Horta, B.L., Mabunda, N., May, C.A., Moreno-Estrada, A., Achilli, A., Olivieri, A., Semino, O., Tambets, K., Kivisild, T., Luiselli, D., Torroni, A., Capelli, C., Tarazona-Santos, E., Metspalu, M., Pagani, L., Montinaro, F., 2019. The Genomic Impact of European Colonization of the Americas. Curr. Biol. 29, 3974-3986.e4. https://doi.org/10.1016/j.cub.2019.09.076

Pankratov, V., Montinaro, F., Kushniarevich, A., Hudjashov, G., Jay, F., Saag, L., Flores, R., Marnetto, D., Seppel, M., Kals, M., Võsa, U., Taccioli, C., Möls, M., Milani, L., Aasa, A., Lawson, D.J., Esko, T., Mägi, R., Pagani, L., Metspalu, A., Metspalu, M., 2020. Differences in local population history at the finest level: the case of the Estonian population. Eur. J. Hum. Genet. 28, 1580–1591. https://doi.org/10.1038/s41431-020-0699-4

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012. Ancient admixture in human history. Genetics 192, 1065–1093. https://doi.org/10.1534/genetics.112.145037

Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. PLoS Genet. 2, e190. https://doi.org/10.1371/journal.pgen.0020190

Peter, B.M., 2022. A geometric relationship of F_2, F_3 and F_4-statistics with principal component analysis. Philos. Trans. R. Soc. B Biol. Sci. 377, 20200413. https://doi.org/10.1098/rstb.2020.0413

Posth, C., Zaro, V., Spyrou, M.A., Vai, S., Gnecchi-Ruscone, G.A., Modi, A., Peltzer, A., Mötsch, A., Nägele, K., Vågene, Å.J., Nelson, E.A., Radzevičiūtė, R., Freund, C., Bondioli, L.M., Cappuccini, L., Frenzel, H., Pacciani, E., Boschin, F., Capecchi, G., Martini, I., Moroni, A., Ricci, S., Sperduti, A., Turchetti, M.A., Riga, A., Zavattaro, M., Zifferero, A., Heyne, H.O., Fernández-Domínguez, E., Kroonen, G.J., McCormick, M., Haak, W., Lari, M., Barbujani, G., Bondioli, L., Bos, K.I., Caramelli, D., Krause, J., 2021. The origin and legacy of the Etruscans through a 2000-year archeogenomic time transect. Sci. Adv. 7, eabi7673. https://doi.org/10.1126/sciadv.abi7673

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Raveane, A., Aneli, S., Montinaro, F., Athanasiadis, G., Barlera, S., Birolo, G., Boncoraglio, G., Di Blasio, A.M., Di Gaetano, C., Pagani, L., Parolo, S., Paschou, P., Piazza, A., Stamatoyannopoulos, G., Angius, A., Brucato, N., Cucca, F., Hellenthal, G., Mulas, A., Peyret-Guzzon, M., Zoledziewska, M., Baali, A., Bycroft, C., Cherkaoui, M., Chiaroni, J., Di Cristofaro, J., Dina, C., Dugoujon, J.M., Galan, P., Giemza, J., Kivisild, T., Mazieres, S., Melhaoui, M., Metspalu, M., Myers, S., Pereira, L., Ricaut, F.X., Brisighelli, F., Cardinali, I., Grugni, V., Lancioni, H., Pascali, V.L., Torroni, A., Semino, O., Matullo, G., Achilli, A., Olivieri, A., Capelli, C., 2019. Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe. Sci. Adv. 5, eaaw3492. https://doi.org/10.1126/sciadv.aaw3492

Raveane, A., Molinaro, L., Aneli, S., Capodiferro, M.R., Gennaro, L. de, Ongaro, L., Migliore, N.R., Soffiati, S., Scarano, T., Torroni, A., Achilli, A., Ventura, M., Pagani, L., Capelli, C., Olivieri, A., Bertolini, F., Semino, O., Montinaro, F., 2022. Assessing temporal and geographic contacts across the Adriatic Sea through the analysis of genome-wide data from Southern Italy. Genomics 114, 110405. https://doi.org/10.1016/j.ygeno.2022.110405

Reitsema, L.J., Mittnik, A., Kyle, B., Catalano, G., Fabbri, P.F., Kazmi, A.C.S., Reinberger, K.L., Sineo, L., Vassallo, S., Bernardos, R., Broomandkhoshbacht, N., Callan, K.,

Candilio, F., Cheronet, O., Curtis, E., Fernandes, D., Lari, M., Lawson, A.M., Mah, M., Mallick, S., Mandl, K., Micco, A., Modi, A., Oppenheimer, J., Özdogan, K.T., Rohland, N., Stewardson, K., Vai, S., Vergata, C., Workman, J.N., Zalzala, F., Zaro, V., Achilli, A., Anagnostopoulos, A., Capelli, C., Constantinou, V., Lancioni, H., Olivieri, A., Papadopoulou, A., Psatha, N., Semino, O., Stamatoyannopoulos, J., Valliannou, I., Yannaki, E., Lazaridis, I., Patterson, N., Ringbauer, H., Caramelli, D., Pinhasi, R., Reich, D., 2022. The diverse genetic origins of a Classical period Greek army. Proc. Natl. Acad. Sci. 119, e2205272119. https://doi.org/10.1073/pnas.2205272119

Saupe, T., Montinaro, F., Scaggion, C., Carrara, N., Kivisild, T., D'Atanasio, E., Hui, R., Solnik, A., Lebrasseur, O., Larson, G., Alessandri, L., Arienzo, I., De Angelis, F., Rolfo, M.F., Skeates, R., Silvestri, L., Beckett, J., Talamo, S., Dolfini, A., Miari, M., Metspalu, M., Benazzi, S., Capelli, C., Pagani, L., Scheib, C.L., 2021. Ancient genomes reveal structural shifts after the arrival of Steppe-related ancestry in the Italian Peninsula. Curr. Biol. CB 31, 2576-2591.e12. https://doi.org/10.1016/j.cub.2021.04.022

Schlebusch, C.M., Skoglund, P., Sjödin, P., Gattepaille, L.M., Hernandez, D., Jay, F., Li, S., De Jongh, M., Singleton, A., Blum, M.G.B., Soodyall, H., Jakobsson, M., 2012. Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. Science 338, 374–379. https://doi.org/10.1126/science.1227721

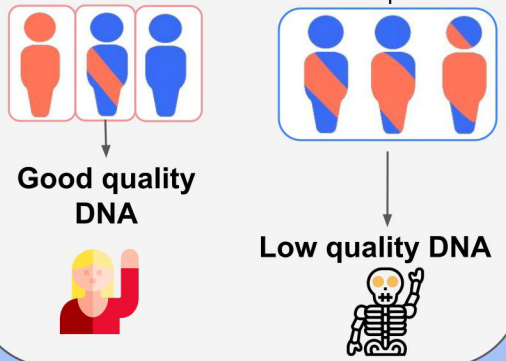
Skoglund, P., Thompson, J.C., Prendergast, M.E., Mittnik, A., Sirak, K., Hajdinjak, M., Salie, T., Rohland, N., Mallick, S., Peltzer, A., Heinze, A., Olalde, I., Ferry, M., Harney, E., Michel, M., Stewardson, K., Cerezo-Román, J.I., Chiumia, C., Crowther, A., Gomani-Chindebvu, E., Gidna, A.O., Grillo, K.M., Helenius, I.T., Hellenthal, G., Helm, R., Horton, M., López, S., Mabulla, A.Z.P., Parkington, J., Shipton, C., Thomas, M.G., Tibesasa, R., Welling, M., Hayes, V.M., Kennett, D.J., Ramesar, R., Meyer, M., Pääbo, S., Patterson, N., Morris, A.G., Boivin, N., Pinhasi, R., Krause, J., Reich, D., 2017. Reconstructing Prehistoric African Population Structure. Cell 171, 59-71.e21. https://doi.org/10.1016/j.cell.2017.08.049

van Dorp, L., Balding, D., Myers, S., Pagani, L., Tyler-Smith, C., Bekele, E., Tarekegn, A., Thomas, M.G., Bradman, N., Hellenthal, G., 2015. Evidence for a Common Origin of Blacksmiths and Cultivators in the Ethiopian Ari within the Last 4500 Years: Lessons for Clustering-Based Inference. PLOS Genet. 11, 1–49. https://doi.org/10.1371/journal.pgen.1005397

Dataset

Reference Groups and Putative Sources

Target Admixed Samples



Good quality DNA

Low quality DNA

Approaches

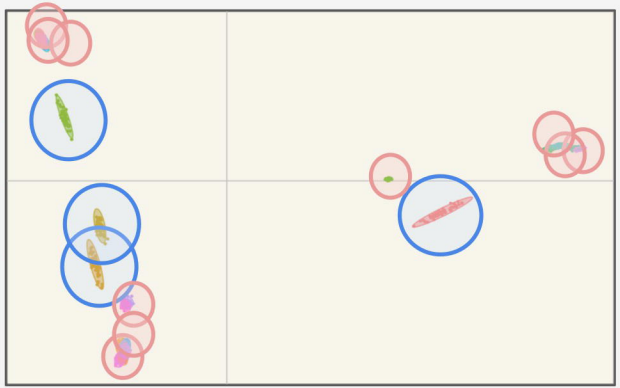
Population-based inference

Individual-based inference

Step 1: PCA

Putative Sources

Src1	PC1	PC2	PCN
Src2	PC1	PC2	PCN
Src3	PC1	PC2	PCN



Admixed Targets

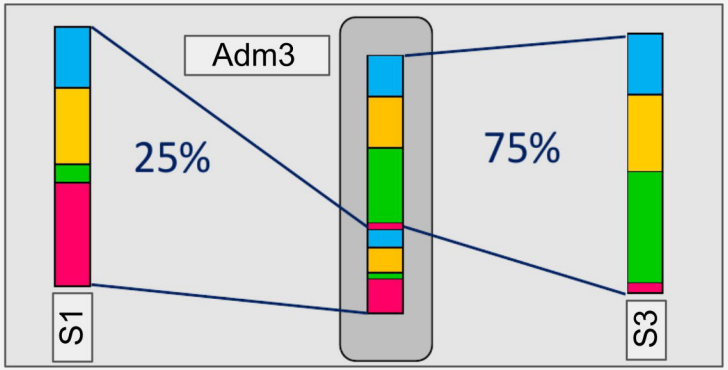
Adm1	PC1	PC2	PCN
Adm2	PC1	PC2	PCN
Adm3	PC1	PC2	PCN

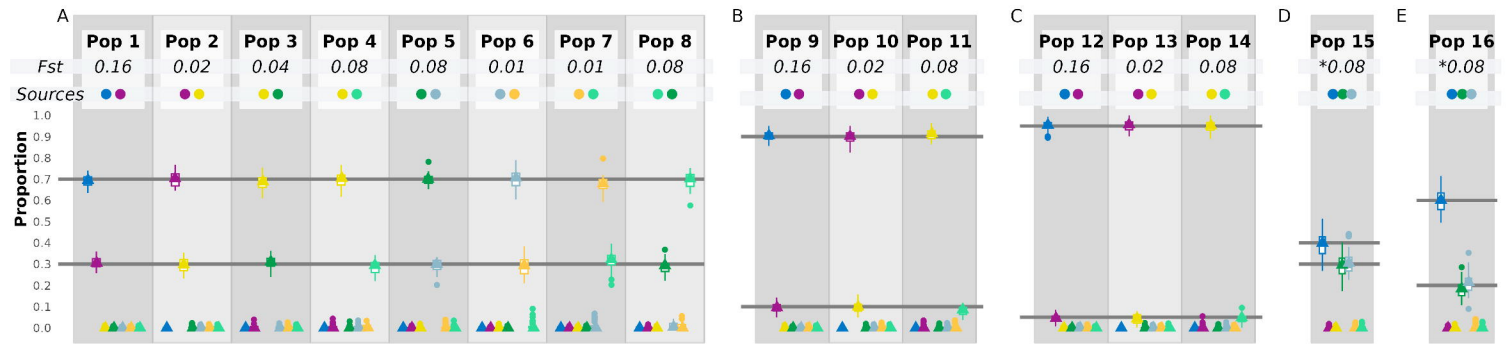
PCA Coordinates of ***all*** sources

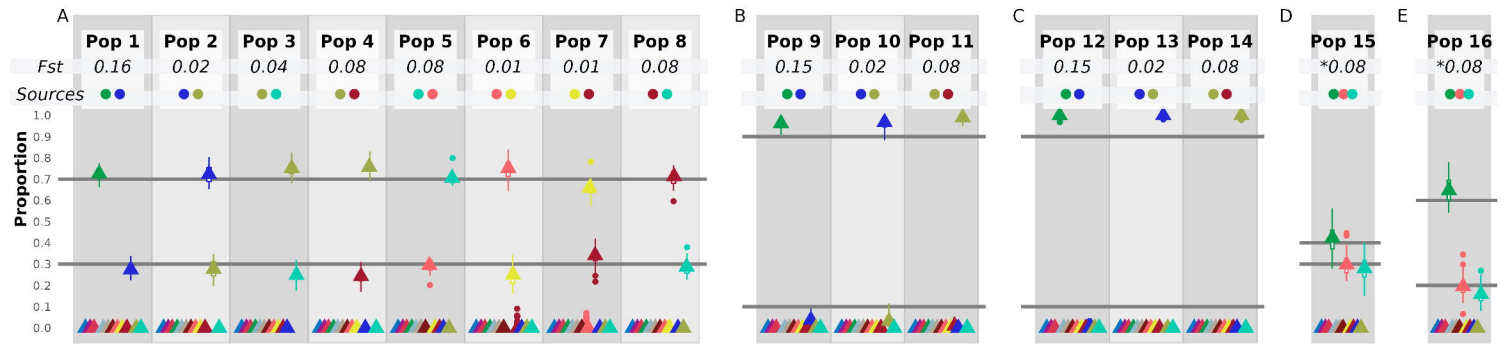
PCA Coordinates of admixed

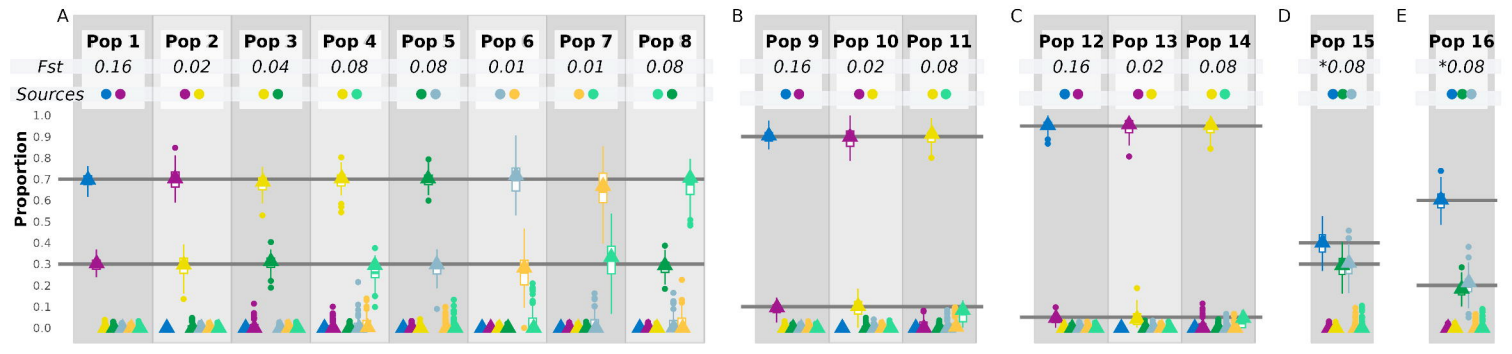
Step 2: NNLS

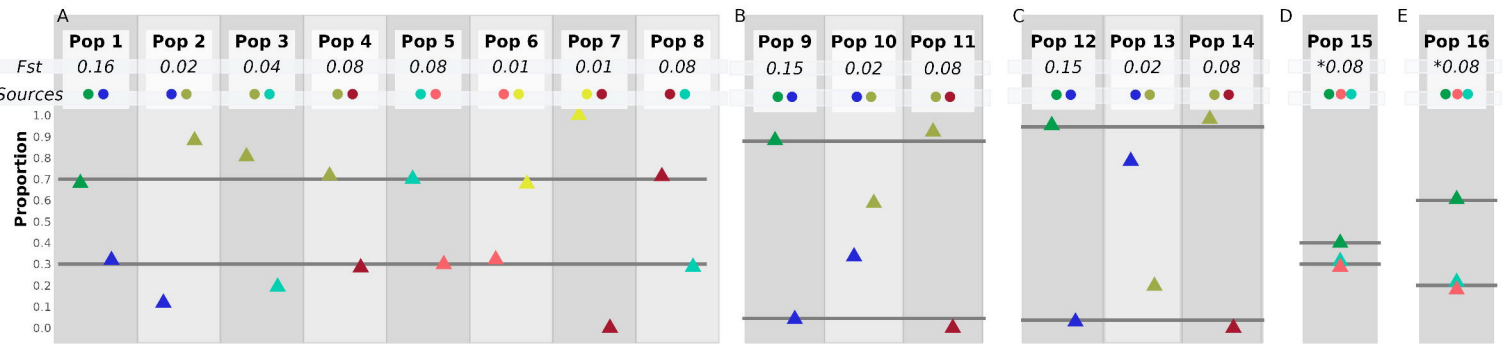
	Src1	Src2	Src3	SrcN
Adm1	0.4	0.05	0.3	0.25
Adm2	0.2	0.3	0.15	0.35
Adm3	0.7	0	0.3	0
Ind1 of Adm4	0.67	0	0.33	0

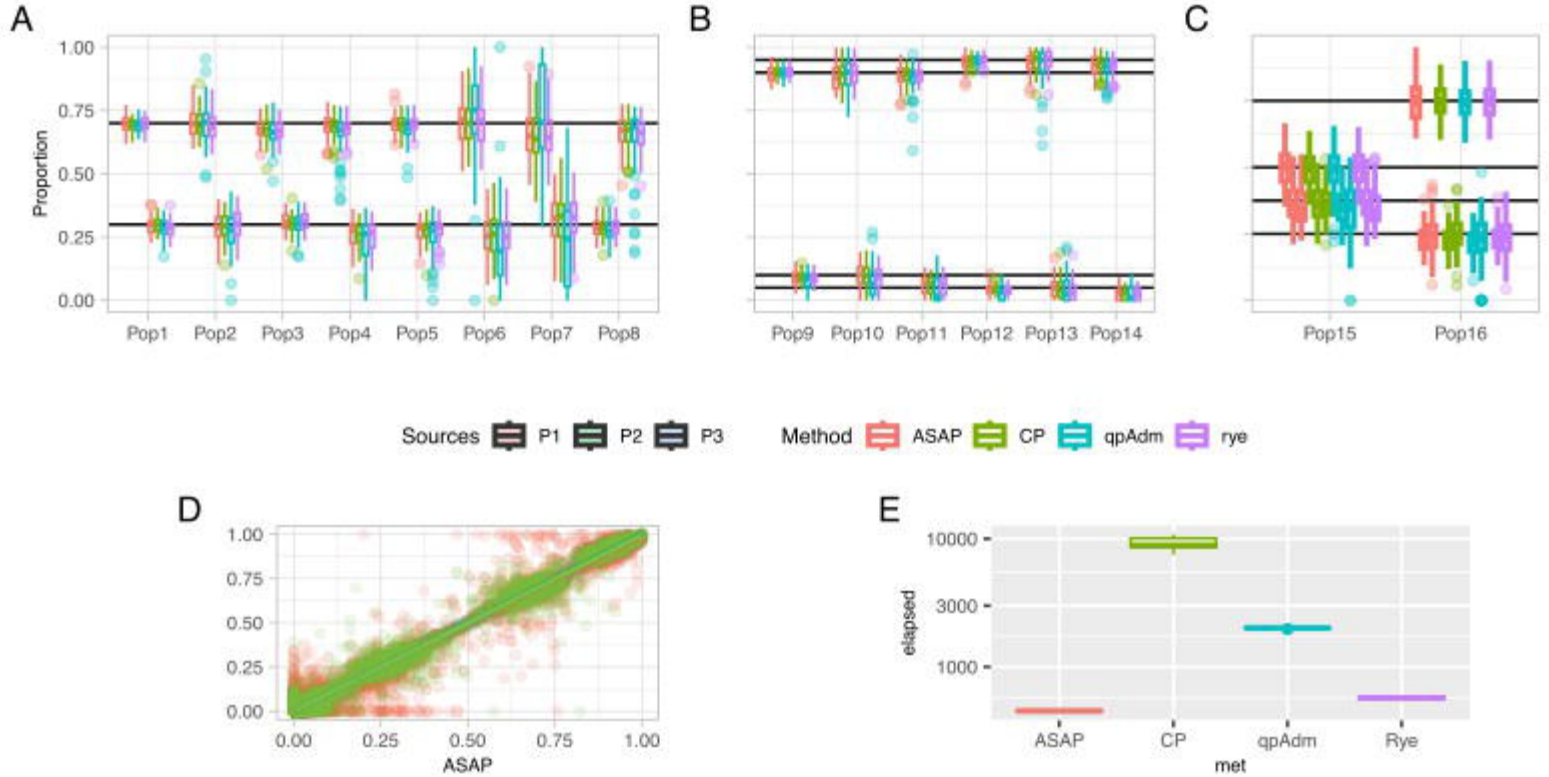


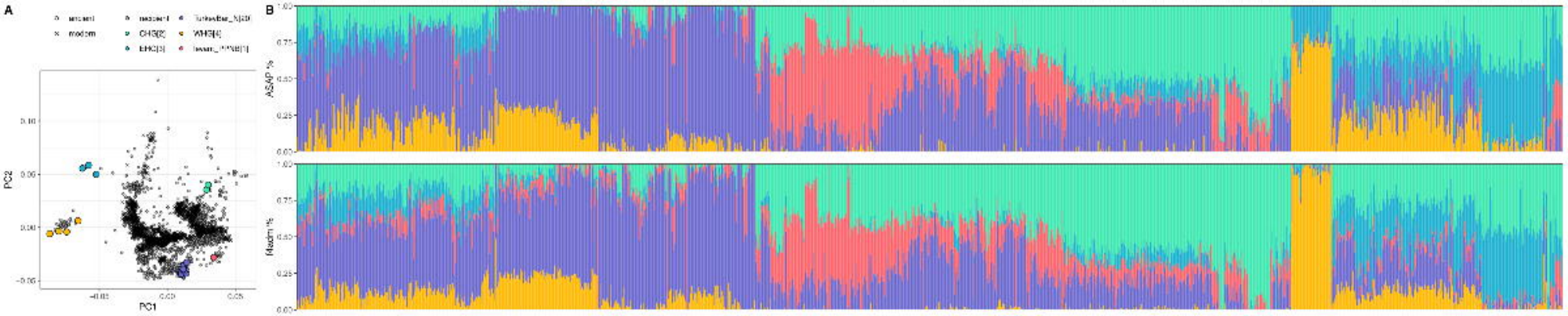

A
B
C
D
E
Pop 1
Pop 2
Pop 3
Pop 4
Pop 5
Pop 6
Pop 7
Pop 8
Pop 9
Pop 10
Pop 11
Pop 12
Pop 13
Pop 14
Pop 15
Pop 16
Fst
0.16
0.02
0.04
0.08
0.08
0.01
0.01
0.08
0.16
0.02
0.08
0.16
0.02
0.08
*0.08
*0.08
Sources
Proportion
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0


A
Pop 1
Pop 2
Pop 3
Pop 4
Pop 5
Pop 6
Pop 7
Pop 8
Fst
0.16
0.02
0.04
0.08
0.08
0.01
0.01
0.08
Sources
Proportion
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
B
Pop 9
Pop 10
Pop 11
0.15
0.02
0.08
C
Pop 12
Pop 13
Pop 14
0.15
0.02
0.08
D
Pop 15
*0.08
E
Pop 16
*0.08


A
B
C
D
E
Pop 1
Pop 2
Pop 3
Pop 4
Pop 5
Pop 6
Pop 7
Pop 8
Pop 9
Pop 10
Pop 11
Pop 12
Pop 13
Pop 14
Pop 15
Pop 16
Fst
0.16
0.02
0.04
0.08
0.08
0.01
0.01
0.08
0.16
0.02
0.08
0.16
0.02
0.08
*0.08
*0.08
Sources
Proportion
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0


A
B
C
D
E
Pop 1
Pop 2
Pop 3
Pop 4
Pop 5
Pop 6
Pop 7
Pop 8
Pop 9
Pop 10
Pop 11
Pop 12
Pop 13
Pop 14
Pop 15
Pop 16
Fst
0.16
0.02
0.04
0.08
0.08
0.01
0.01
0.08
0.15
0.02
0.08
0.15
0.02
0.08
*0.08
*0.08
Sources
Proportion
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0


A
B
C
Proportion
1.00
0.75
0.50
0.25
0.00
Pop1
Pop2
Pop3
Pop4
Pop5
Pop6
Pop7
Pop8
Pop9
Pop10
Pop11
Pop12
Pop13
Pop14
Pop15
Pop16
Sources
P1
P2
P3
Method
ASAP
CP
qpAdm
rye
D
E
ASAP
elapsed
10000
3000
1000
CP
qpAdm
Rye
met


A
ancient
modern
WHG[4]
PC1
PC2
B
ASAP %
Hadm %