False discovery rates of *qpAdm*-based screens for genetic admixture

Eren Yüncü[1,*,†], Ulaş Işıldak[1,*,†], Matthew P. Williams[2], Christian D. Huber[2], Leonid A. Vyazov[1], Piya Changmai[1], Pavel Flegontov[1,3,§]

[1] *Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czechia*

[2] *Department of Biology, Eberly College of Science, Pennsylvania State University, PA, USA*

[3] *Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA*

* the authors contributed equally

† present addresses: E.Y., Department of Biological Sciences, Middle East Technical University, Ankara, Turkey; U.I., Leibniz Institute on Aging - Fritz Lipmann Institute (FLI), Jena, Germany.

§ Correspondence to: Pavel Flegontov

(pavel.flegontov@osu.cz, pavel_flegontov@hms.harvard.edu)

Abstract

Although a broad range of methods exists for reconstructing population history from genome-wide single nucleotide polymorphism data, just a few methods gained popularity in archaeogenetics: principal component analysis (PCA); *ADMIXTURE*, an algorithm that models individuals as mixtures of multiple ancestral sources represented by actual or inferred populations; formal tests for admixture such as f_3-statistics and *D*-statistics; and *qpAdm*, a tool for fitting two-component and more complex admixture models to groups or individuals. Despite their popularity in archaeogenetics, which is explained by modest computational requirements and ability to analyse data of various types and qualities, protocols relying on *qpAdm* that screen numerous alternative models of varying complexity and find "fitting" models (often considering both estimated admixture proportions and *p*-values as a composite criterion of model fit) remain untested on complex simulated population histories in the form of admixture graphs of random topology. We analysed genotype data extracted from such simulations and tested various types of high-throughput *qpAdm* protocols ("rotating" and "non-rotating", with or without temporal stratification of target groups and proxy ancestry sources, with or without a "model competition" step). We caution that these *qpAdm* protocols may be inappropriate for exploratory analyses in poorly studied regions/periods since their false discovery rates varied between 12% and 68% depending on the details of the protocol and on the amount and quality of simulated data (i.e., >12% of fitting two-way admixture models imply gene flows that were not simulated), although our study has a number of limitations. We demonstrate that for reducing false discovery rates of *qpAdm* protocols to nearly 0% it is advisable to use large SNP sets with low missing data rates, the rotating *qpAdm* protocol with a strictly enforced rule that target groups do not pre-date their proxy sources, and an unsupervised *ADMIXTURE* analysis as a way to verify feasible *qpAdm* models.

Introduction

Although a broad range of methods exists for reconstructing population history from genome-wide autosomal single nucleotide polymorphism (SNP) data, just a few methods became the cornerstone of archaeogenetic studies: principal component analysis (PCA) (Patterson et al. 2006); an unsupervised or supervised algorithm for admixture inference in individuals, *ADMIXTURE* (Alexander et al. 2009); formal tests for admixture such as f_3-statistics (Patterson et al. 2012, Peter 2016, Soraggi and Wiuf 2019) and *D*-statistics (Green et al. 2010, Durand et al. 2011); and a tool for fitting two-component and more complex admixture models to populations, *qpAdm* (Haak et al. 2015, Harney et al. 2021). The popularity of these methods is explained by their relatively modest computational requirements and versatility since they are capable of analysing unphased biallelic genotype data of various types (pseudo-haploid or diploid), generated using either targeted enrichment on a panel of sites or shotgun sequencing technologies, and low-coverage ancient genomes with high proportions of missing data (Harney et al. 2021). However, only a few studies were devoted to testing the performance of these diverse methods on simulated genetic data (Alexander et al. 2009, Harney et al. 2021, Lazaridis et al. 2017, Martin et al. 2014, McVean 2009, Moreno-Mayar et al. 2018b, Ning et al. 2020, Soraggi and Wiuf 2019), and realistically complex population histories remain virtually unexplored in this respect.

In a typical archaeogenetic study published since the 2010s, PCA is used as a first line of analysis, providing an overview of population structure and helping to propose hypotheses about migration and admixture. Distribution of individual genomes in two- or higher dimensional spaces of principal components (PCs) does not have an unambiguous interpretation since even under ideal conditions (in the absence of missing data, batch artefacts, and selection signals) it is affected by both genetic drift and admixture (McVean 2009). Nevertheless, if context information such as geographical coordinates and dates for ancient individuals is available, PCA is routinely used for revealing “genetic clines” interpreted as signs of admixture between originally isolated groups at the ends of such clines. However, a study on simulated data by Novembre and Stephens (2008) concluded that clinal PCA patterns do not necessarily indicate historical migration events; these patterns generally emerge because of decrease in genetic similarity with distance. PC1 vs. PC2 scatterplots were

also shown to display an arch-shaped artefact, a "horseshoe" (Podani and Miklós 2002), under the homogeneous migration scenario (Novembre and Stephens 2008, Frichot et al. 2012). It was demonstrated that the distribution of individuals in the PC space depends on the expected coalescent time (McVean 2009), hence distinct demographic models with the same expected coalescence times are expected to have the same PCA projections. Additionally, imbalanced sampling across genetically divergent populations affects PCA results substantially (McVean 2009). Hence, using further methods to correlate PCA results with other lines of evidence is necessary for studying migration history (Reich et al. 2008).

Formal tests for genetic admixture such as f_3-statistics and D/f_4-statistics are often used exactly for this purpose: to prove that a certain cline spotted in PC space is a result of migration and admixture of previously isolated ancestries and does not reflect isolation by distance or recurrent bottlenecks. *D*- and f_4-statistics, which are identical except for the denominator and are not affected by genetic drift, test if an unrooted four-population tree fits the data (Reich et al. 2009, Green et al. 2010, Durand et al. 2011, Patterson et al. 2012). A statistically significant deviation of the statistic from 0 (estimated using jackknife or bootstrap resampling) means that either the assumed tree topology is wrong, or gene flow occurred between a pair of branches in the tree, assuming that recurrent mutations and SNP ascertainment bias are absent (Durand et al. 2011, Patterson et al. 2012). However, interpretation of these statistics is ambiguous since gene flow directionality remains unknown, and two pairs of branches can be responsible for a deviation of the statistic from 0 (Lipson 2020). Since gene flow may be mediated by ghost groups related only distantly to the sampled groups at the branch tips (Tricou et al. 2022), excluding one pair of branches due to geographical and temporal plausibility of gene flow is also difficult. And interpretation of deviations of *D*- and f_4-statistics from 0 becomes hardly possible if both branch pairs are connected by detectable gene flows.

"Admixture" f_3-statistics of the form f_3(target; proxy source$_1$, proxy source$_2$) constitute another formal test for admixture (Patterson et al. 2012). Significant deviation of such a statistic from 0 in the negative direction (Z-score below -3) is considered proof of admixture since allele frequencies at most sites are intermediate in the target group between those in the proxy sources (Patterson et al. 2012). However, "admixture" f_3-statistics are usually only applicable for detection of recent admixture events since they become positive when post-

admixture genetic drift on the target lineage moves allele frequencies away from these intermediate values (Patterson et al. 2012, Peter 2016).

Considering these complications, more sophisticated tests for genetic admixture are needed. The *qpAdm* method introduced by Haak *et al*. (2015) is based on matrices of f_4-statistics and does not require detailed knowledge of population phylogeny beyond a few assumptions (Lazaridis et al. 2016, Harney et al. 2021). This method tests admixture models in the form of combinations of proxy ancestral groups ("sources" or "references", Lazaridis et al. 2016) for a "target" (or "test") population, given a genetically diverse set of "outgroup" populations, and infers ancestry proportions in the target group contributed by the lineages represented by the proxy sources ("outgroups" are often termed "right" populations for convenience since they are usually not outgroups in the phylogenetic sense, and they were termed "references" by Harney *et al*. 2021). See Box 1 for definitions of various *qpAdm*-related terms.

Box 1. Terminology used in this study for describing admixture models, results of screens for admixture, and *qpAdm* protocols.

"right" (or outgroup) populations/groups	In a *qpAdm* protocol, these are reference populations needed for testing admixture models composed of a target and one or several proxy source groups.
target (or test) population/group	In an admixture model, this is a population/group whose genetic history is being modelled.
true (ancestry) source	In the context of simulated admixture graphs, this is a population directly participating in an admixture event(s) giving rise to a target group, and its ancestral (unsampled) population before its merging with other populations.
proxy (ancestry) source	In the context of simulated admixture graphs, this is a sampled population included in an admixture model as a potential source and assumed to be cladal with one of true ancestry sources.
model complexity	Number of proposed proxy ancestry sources in an admixture model "target = proxy source$_1$ + proxy source$_2$ + … proxy source$_i$".
"left" populations/groups	In a *qpAdm* model, these are proxy sources and a target.
reference populations/groups	"Right" and proxy source populations combined.
(composite) model feasibility criterion	A criterion for identifying fitting (feasible) *qpAdm* models that relies on both estimated admixture proportions and the *p*-value. The following criterion was used in this study (all the conditions listed below should be satisfied): 1) *p*-values for all models with *n–1* ancestry sources, such as one-way models "target = proxy

	source$_1$", "target = proxy source$_2$", and "proxy source$_1$ = proxy source$_2$", are all below 0.01; 2) for a model with *n* ancestry sources, such as a two-way model "target = proxy source$_1$ + proxy source$_2$", estimated admixture proportions ± 2 standard errors are between 0 and 1; 3) the *p*-value for a model with *n* ancestry sources ≥ 0.01. Other versions of this criterion are also found in the literature: with different values of the *p*-value cut-off and/or not considering standard errors of estimated admixture proportions.
feasible (or fitting) *qpAdm* model	A *qpAdm* model satisfying the feasibility criterion above.
positive/negative admixture model	An admixture model supported/not supported by one or several analytical tools such as *qpAdm*, PCA, *ADMIXTURE*.
false discovery rate (FDR)	The fraction of admixture models of a certain complexity (e.g., two-way) satisfying the *qpAdm* model feasibility criteria but classified as false considering the simulated graph topology and simulated admixture proportions. For topological criteria used for classifying two-way admixture models into true and false ones see the Results and Methods sections. The term is also applied to outcomes of more complex admixture inference pipelines composed of several methods.
false omission rate (FOR)	The fraction of feasible *qpAdm* models that are classified as true considering the simulated graph topology and simulated admixture proportions but are not supported by another method (*qpAdm* model competition, PCA, or *ADMIXTURE*) or a combination of methods.
"rotating" *qpAdm* protocol	A protocol having the following feature: a large subset of reference populations or all of them are distributed between the "right" and "left" sets according to the principle "whatever is not in the left set is in the right set", testing all possible bisections of this sort for a given rotated set and a given range of model complexities. In the most extreme case, target groups are also included in this rotation. Model testing starts from the simplest one-way models and moves on to the next complexity level if all simple models for a given target are rejected according to the composite feasibility criterion. The goal of this approach, as compared to "non-rotating" protocols, is to increase the power of the method to reject non-optimal proxy ancestry sources.
"non-rotating" *qpAdm* protocol	This protocol we also term "standard" or "basic": all models are tested with one or few fixed sets of "right" groups which usually pre-date "left" groups or are contemporaneous with them. In practice, modern populations genetically divergent from the target are often included in such a "right" set if ancient reference groups are unavailable. Model testing starts from the simplest one-way models and moves on to the next complexity level if all

	simple models for a given target are rejected according to the composite feasibility criterion.
temporal stratification of targets and proxy sources	A requirement that target groups post-date or are contemporaneous with all proxy sources in each model. In practice, this requirement is often included in rotating and non-rotating *qpAdm* protocols.
"distal" *qpAdm* protocol	A rotating or non-rotating *qpAdm* protocol with temporal stratification of targets and proxy sources.
"proximal" *qpAdm* protocol	A rotating or non-rotating *qpAdm* protocol without temporal stratification of targets and proxy sources.
model competition	A *qpAdm* protocol which starts from sets of alternative feasible *qpAdm* models of a certain complexity level for a given target, outcomes of a non-rotating protocol. Alternative proxy sources from these models form a rotated set. Rotation can be performed in two ways (groups from the rotated set are placed in the "left" set one by one, and the other groups from the rotated set are placed in the "right" set; or groups from the rotated set are placed in both "left" and "right" sets one by one), and the composite feasibility criterion is applied. The goal of this approach is to increase the power of the method to reject non-optimal proxy ancestry sources.
proximal model competition *qpAdm* protocol	As introduced by Narasimhan *et al*. (2019), this is a non-rotating *qpAdm* protocol with temporal stratification of "right" and "left" sets, with a subsequent model competition step, and with no (or very limited) temporal stratification of targets and proxy sources at both steps.

Eight years later we find *qpAdm*-based protocols routinely employed in large-scale screens of ancient human or animal populations for admixture (often between closely related sources) and used as formal tests for admixture (see Lazaridis et al. 2016, Skoglund et al. 2017, Harney et al. 2018, Mathieson et al. 2018, Antonio et al. 2019, Narasimhan et al. 2019, Fernandes et al. 2020, Marcus et al. 2020, Ning et al. 2020, Wang et al. 2020, Yang et al. 2020, Calhoff et al. 2021, Papac et al. 2021, Librado et al. 2021, Sirak et al. 2021, Wang et al. 2021, Yaka et al. 2021, Zhang et al. 2021, Allentoft et al. 2022, Bergström et al. 2022, Changmai et al. 2022a, Changmai et al. 2022b, Gnecchi-Ruscone et al. 2022, Lazaridis et al. 2022, Maróti et al. 2022, Oliveira et al. 2022, Patterson et al. 2022, Brielle et al. 2023, Lee et al. 2023, Taylor et al. 2023 for examples). *qpAdm* fits admixture models to a matrix of f_4-statistics of the form f_4("left" group$_i$, "left" group$_j$; "right" group$_i$, "right" group$_j$), which in the absence of missing data at

the group level can be reduced to a smaller matrix f_4(target group, "left" group$_j$; "right" group$_1$, "right" group$_j$), considering algebraic relationships between different f_4-statistics (Peter 2016).

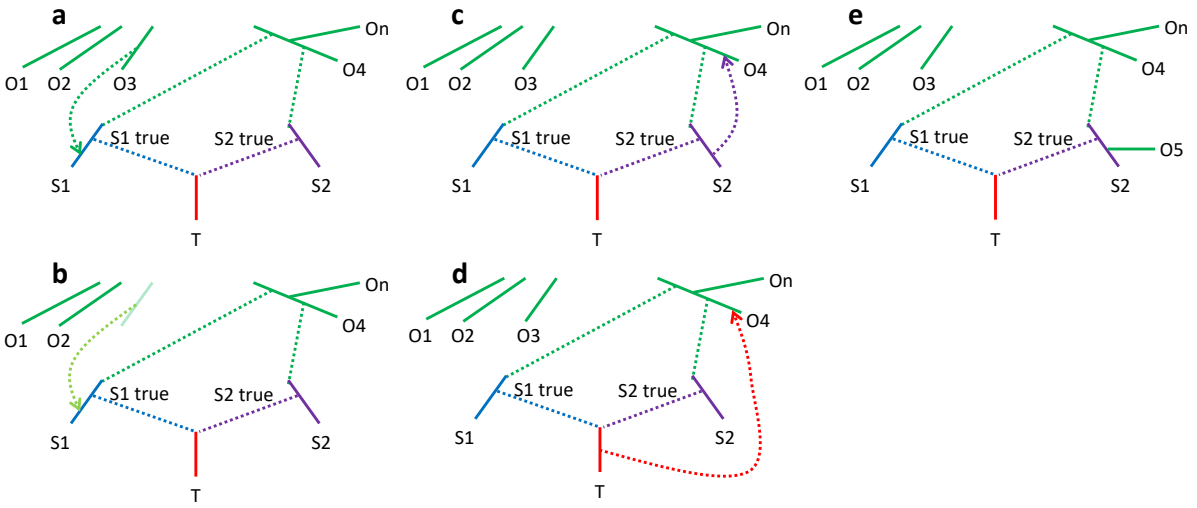


Figure 1. Admixture graphs showing an exhaustive list of assumption violations of the standard *qpAdm* protocol that may lead to rejection of the true simple model, and thus prompt the researcher to test overly complex models. (**a**) A gene flow from an outgroup O* to a proxy source after the divergence of the latter from the true source. (**b**) A gene flow from an unsampled source to a proxy source after the divergence of the latter from the true source. This case is problematic only if the outgroups are differentially related to the unsampled source. (**c**) A gene flow from a proxy source to an outgroup after the divergence of the former from the true source. (**d**) A gene flow from a target to an outgroup. (**e**) An outgroup is cladal with a proxy source.

A *qpAdm* protocol that has become the standard in archaeogenetics (Lazaridis et al. 2016) can be broken down into two parts: estimation of the number of gene flows connecting the "right" and "left" sets (this method was first published as a tool named "*qpWave*", Reich et al. 2012) and inference of admixture proportions in a target group in the "left" set (Haak et al. 2015). *qpWave* tests for the number of distinct gene flows connecting the "right" and "left" population sets, does not infer directionality of gene flows, and does not identify recipients of gene flow in the "left" or "right" population sets. Notably, the standard *qpAdm* protocol relies on the following assumptions (Lazaridis et al. 2016, Harney et al. 2021): 1) there is at least one "right" population differentially related to the proxy sources; 2) proxy sources are strictly cladal with the true ancestral admixing sources (Fig. 1a,b), 3) there are no gene flows to populations located in the "right" set from the proxy source or target lineages either after

the split of the proxy source from the true admixing source population or between the target population and the admixture event that gave rise to it (Fig. 1c-e);. In the context of our study, true sources are unsampled populations that participated in a simulated admixture event (labelled as “S1 true” and “S2 true” in Fig. 1, see also Box 1).

If the above assumptions are satisfied, it is safe to say that *qpWave/qpAdm* rejections of simpler models, and a failure to reject more complex models, are the result of a genuinely complex admixture history that connects the source and target populations rather than the false rejection of the simple model due to violations of any one of the assumptions described above. Most notably, violations of the second or third assumptions raise the probability of rejecting a simpler (true) model and prompt the researcher to test more complex (false) models (such as in Fig. 1 rejecting a two-source *qpAdm* model and exploring three-source models).

Harney *et al*. (2021) demonstrated on simulated data that, if the *qpAdm* assumptions are satisfied, it is highly favourable for statistical power of the method (for distinguishing between alternative proxy sources that are unequally distant genetically from the true ancestry source) to move at least some alternative proxy sources between the "left" and "right" sets. In other words, having "on the right" populations that do not violate the topological assumptions of *qpAdm*, but are closely related to proxy sources on the "left", increases the statistical power greatly (see also Ning et al. 2020 for another demonstration of this on simple simulated histories).

This new type of *qpAdm* protocols, termed "rotating" protocol, has been adopted in archaeogenetics widely (see, e.g., Skoglund et al. 2017, Harney et al. 2019, Narasimhan et al. 2019, Olalde et al. 2019, Calhoff et al. 2021, Fernandes et al. 2021, Librado et al. 2021, Allentoft et al. 2022, Bergström et al. 2022, Lazaridis et al. 2022, Oliveira et al. 2022, Taylor et al. 2023). The most extreme version of the "rotating" protocol simply divides a set of reference populations into all possible combinations of "right" and "proxy source" subsets of certain sizes and rotates these combinations through successive *qpAdm* analyses. Additional constraints can be placed on the rotating combinations such as restricting a set of groups (usually highly divergent from the target) to remain in the “right” set in all models. When evaluating the totality of multiple *qpAdm* tests, the simplest feasible models (e.g., one-way,

i.e., unadmixed) are favoured, and increasingly complex models are explored upon the rejection of simpler models. Model rejection for the simplest models is made according to a chosen *p*-value threshold such that *qpAdm* models are considered feasible or "fitting" the data when the *p*-value is above such a threshold (Skoglund et al. 2017, Harney et al. 2018, Narasimhan et al. 2019, Olalde et al. 2019, Yang et al. 2020, Calhoff et al. 2021, Fernandes et al. 2021, Librado et al. 2021, Zhang et al. 2021, Allentoft et al. 2022, Bergström et al. 2022, Lazaridis et al. 2022, Oliveira et al. 2022, Taylor et al. 2023). As an additional criterion of a fitting model, all inferred admixture proportions (Harney et al. 2018, Olalde et al. 2019, Yang et al. 2020, Zhang et al. 2021, Allentoft et al. 2022, Lazaridis et al. 2022, Oliveira et al. 2022), or proportions $\pm$ 2 standard errors (Narasimhan et al. 2019), may be required to lie between 0 and 1. It is important to remember that the statistical significance of the *qpAdm/qpWave* test is, strictly speaking, a function of model rejection, and thus the failure to reject a model may have underlying causes other than approximating the true situation well enough (such as lack of statistical power or a lack of suitable “right” groups that capture the divergent ancestry sources amongst the “left” group).

A less exhaustive version of the rotating *qpAdm* protocol, termed "model competition" (e.g., Narasimhan et al. 2019, Fernandes et al. 2020, Calhoff et al. 2021, Sirak et al. 2021, Zhang et al. 2021, Maróti et al. 2022, Brielle et al. 2023, Lee et al. 2023), is used even more widely than the basic rotating protocol. It involves an initial (standard non-rotating) *qpAdm* analysis on a number of source populations (see Box 1). Upon identifying a sub-list of plausible sources for a target, the algorithm re-tests feasible models for this target rotating these plausible sources between the "left" and "right" sets with the expectation of improving the power to reject models including proxy sources that are genetically distant from the true sources.

The rotating *qpAdm* protocol and model competition are increasingly used as central methods for testing admixture hypotheses proposed after inspecting distributions of individuals in PC spaces, similarity patterns in outcomes of *ADMIXTURE* analyses, and *f/D*-statistics indicative of an admixture graph rather than a simple tree relationship. Yet, the only study reporting detailed testing of *qpAdm* on simulated data (Harney et al. 2021) was performed in extremely favourable conditions: the simulated graph included just two non-nested admixture events; the sources for the principal target group diverged about 1,200

generations ago (almost 35,000 years ago in the case of humans); the proxy sources were strictly cladal with the actual ancestral groups for the target; several groups differentially related to these ancestry sources were available; the simulated data were free of ascertainment bias since sites were sampled in a random way; one million sites were used for most analyses; and only 50/50% simulated admixture proportions were tested for some analyses. This study confirmed that the method behaves as expected under these ideal conditions and offered some important guidelines on the choice and number of "right" populations for optimal specificity of the method and on model comparison strategies, and also showed that the results are robust to the presence of missing data, imbalanced group sizes, ancient DNA damage, and to a particular type of SNP ascertainment: selecting sites heterozygous in one individual from a certain population (Patterson et al. 2012). Among challenging historical scenarios, only multifurcation with subsequent continuing gene flow among several groups was explored, and it was concluded that *qpAdm* is not applicable in this case (Harney et al. 2021). Meanwhile, the false discovery rate (FDR) or the false positive rate (FPR) of the method and violations of the topological assumptions of *qpAdm* (Fig. 1) remained virtually unexplored. Thus, the method was proven to work and fail in polar (either extremely favourable or extremely unfavourable) conditions. But what happens in intermediate cases where arguably most of the history of humans and other mammals fits: a history that is not a nearly perfect tree, but that, on the other hand, cannot be represented solely by gene flows homogeneous in space and constant in time?

We are concerned that *qpAdm* performance may be compromised by the fact that the topological assumptions of the method are hard to verify in practice, especially the assumption about cladality of proxy and true ancestry sources (i.e., no gene flow to the proxy source population after its split from the true admixing source population). To explore this problem, we analyse simulated population histories in the form of complex random admixture graphs and test various types of *qpAdm* protocols common in the literature: rotating and non-rotating, with or without temporal stratification of target groups and proxy ancestry sources, with or without a model competition step. We also reproduced other aspects of a typical archaeogenetic study on simulated data: we combined various *qpAdm* protocols with PCA and an unsupervised *ADMIXTURE* analysis to explore FDR of complex admixture screening pipelines.

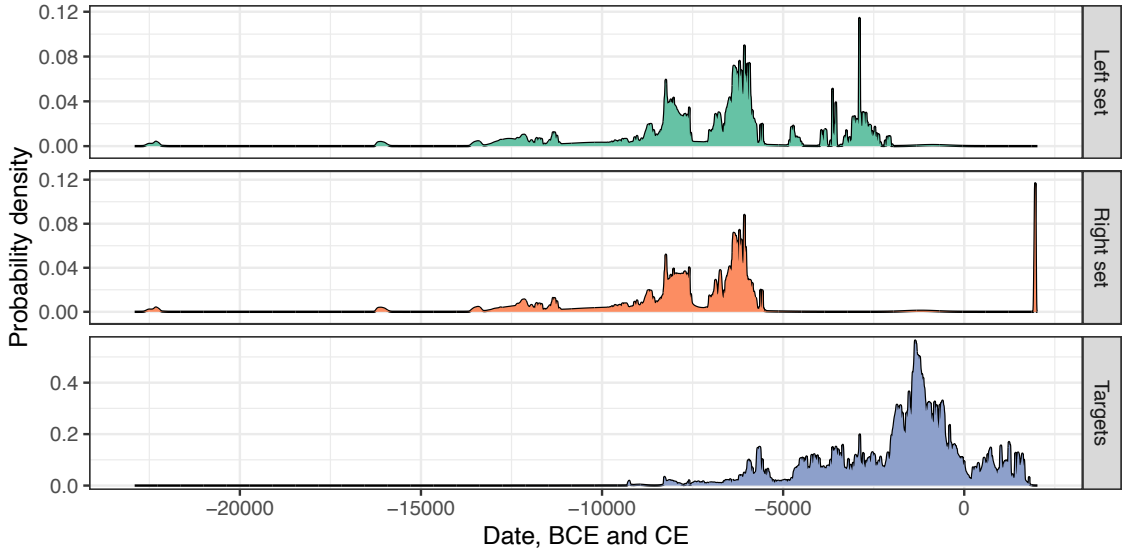
Results

An overview of published rotating and model competition qpAdm protocols

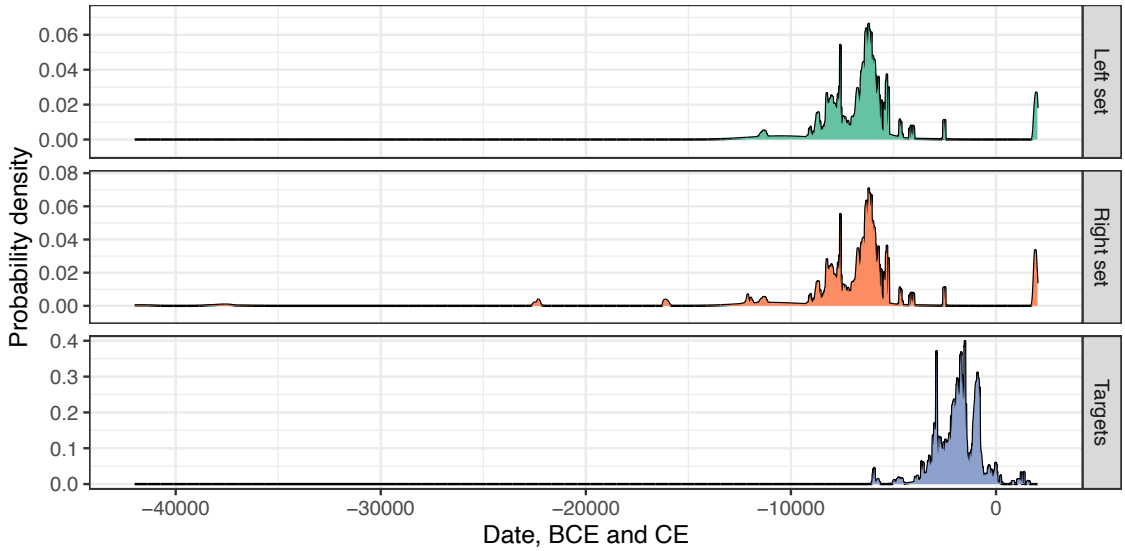
First, we outline two published rotating *qpAdm* protocols (Narasimhan et al. 2019, Lazaridis et al. 2022) that are typical for this class of protocols (see further examples in Skoglund et al. 2017, Harney et al. 2019, Olalde et al. 2019, Calhoff et al. 2021, Fernandes et al. 2021, Librado et al. 2021, Allentoft et al. 2022, Bergström et al. 2022, Oliveira et al. 2022, Taylor et al. 2023). Lazaridis *et al*. relied on the following set of 15 reference populations: 1) Mbuti (present-day Africans); 2) a Palaeolithic group from the Caucasus (CHG, Caucasian hunter-gatherers); 3) East European Mesolithic (EHG, East European hunter-gatherers); 4) Ganj Dareh (a Neolithic group from Iran); 5) Natufians (an Epipalaeolithic group from Israel); 6) a Pre-pottery Neolithic (PPN) group from the Levant; 7) Taforalt (an Epipalaeolithic group from Morocco); 8) Neolithic Mesopotamia; 9) Afontova Gora 3 (an individual from Late Upper Palaeolithic Siberia); 10) Mal'ta 1 (an individual from Late Upper Palaeolithic Siberia); 11) a Mesolithic group from the Iron Gates region (Serbia); 12) Boncuklu (a Pre-pottery Neolithic group from Central Turkey); 13) Barcın (a Neolithic group from Western Turkey); 14) Pınarbaşı (an Epipalaeolithic individual from Turkey); and 15) Mesolithic and Palaeolithic individuals from Western Europe (WHG, West European hunter-gatherers).

This reference set was divided into all possible "right" and proxy source subsets, except for the African group (Mbuti) which stayed in the "right" set in all models. Three Chalcolithic groups from Iran and an Early Bronze Age group from Russia (Yamnaya) were considered as proxy sources only and not rotated to the "right" set, and various clusters of Chalcolithic and Bronze Age individuals (most of them dated between ca. 5000 and 1000 years BCE) from the Balkans, Anatolia, Levant, Caucasus, Mesopotamia, and Iran were target groups for the *qpAdm* analyses. Thus, this protocol can be classified as a distal rotating protocol (Box 1) since most (but not all) targets do not pre-date the proxy sources (Fig. 2a). For each target, progressively more complex admixture models were tested, including from one to five proxy sources, and in most cases only the simplest feasible models were interpreted. Model feasibility criteria were as follows: estimated admixture proportions between 0 and 1, and *p*-value > 0.01. Among alternative models for the same target, those having a higher *p*-value were considered fitting the data better (Lazaridis et al. 2022).

a, Lazaridis *et al*. 2022, distal rotating protocol



b, Narasimhan *et al*. 2019, distal rotating protocol



c, Narasimhan *et al*. 2019, proximal model competition protocol

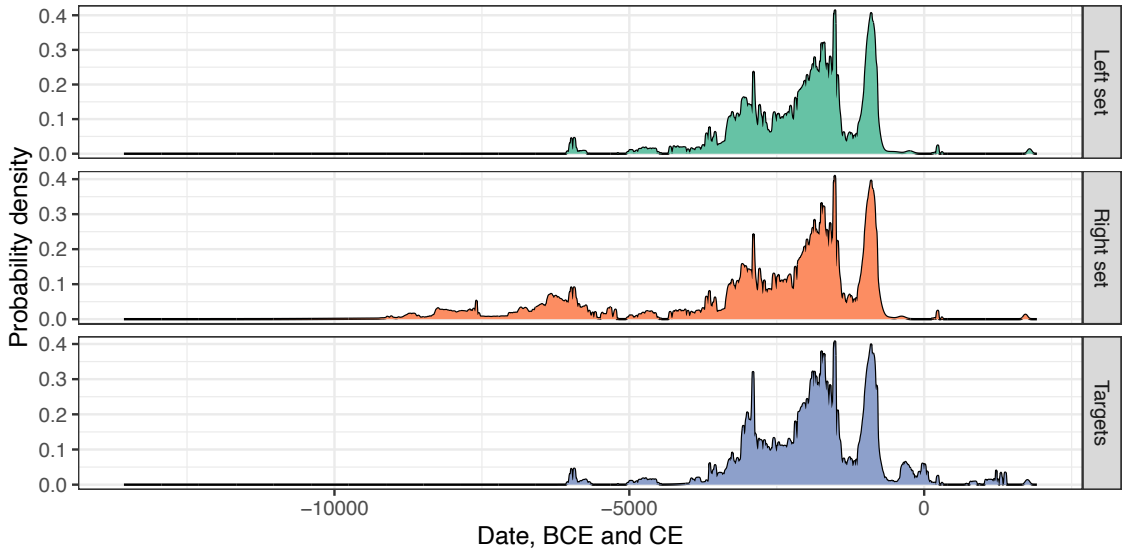


Figure 2. Distributions of radiocarbon and calendar dates for populations sets analyzed with the distal rotating *qpAdm* protocols by Lazaridis *et al*. 2022 (**a**) and Narasimhan *et al*. 2019 (**b**), and with the proximal model competition protocol from Narasimhan *et al*. 2019 (**c**). Probability density curves are shown for three sets of groups: 1) those appearing in the "right" set in at least one *qpAdm* model; 2) those appearing in the "left" set in at least one *qpAdm* model; 3) target groups. Targets in the former study were composed of large clusters of West Eurasian individuals, some of them dating back to the Palaeolithic (Lazaridis et al. 2022). For that reason, the date distribution for targets in panel **a** is very wide.

As shown in Fig. 2a, in this analytical setup there is a large temporal overlap between “left” groups (targets and, on average, earlier proxy sources) and “right” groups. For instance, such a divergent and ancient group as the Mal’ta 1 individual from the vicinity of Lake Baikal (dated to ca. 24,000 years before present, yBP, Raghavan et al. 2014) appeared “on the left” in some *qpAdm* models. Thus, “left-to-right” gene flows (that may lead to erroneous conclusions from a *qpAdm* analysis, see Fig. 1) are expected to be common in the analytical setup used by Lazaridis *et al*.

Narasimhan *et al*. (2019) used both proximal and distal *qpAdm* protocols (Box 1). The distal rotating protocol relied on the following set of 16 reference populations: 1) Mota (a 4500-years-old individual from Ethiopia); 2) Ust’-Ishim (an Upper Palaeolithic individual from West Siberia); 3) Tianyuan (an Upper Palaeolithic individual from Northeast China); 4) Late Upper Palaeolithic individuals from Siberia (Afontova Gora 3 and Mal’ta 1, collectively labelled “ANE” or “Ancient North Eurasians”); 5) a Late Upper Palaeolithic individual from Italy (Villabruna); 6) Natufians (an Epipalaeolithic group from Raqefet, Israel); 7) a Mesolithic individual from Iran (Belt Cave); 8) present-day Andamanese; 9) East European Mesolithic individuals (EEHG, East European hunter-gatherers); 10) West Siberian Mesolithic (WSHG, West Siberian hunter-gatherers); 11) a Pre-pottery Neolithic (PPN) group from the Levant; 12) a Mesolithic group from the Iron Gates region (WEHG, West European hunter-gatherers); 13) Anatolian Neolithic individuals; 14) Ganj Dareh (a Neolithic group from Iran); 15) an Early Neolithic group from the Baikal region (ESHG, East Siberian hunter-gatherers); and 16) present-day Han Chinese. This reference set was split into all possible “right” and proxy source subsets, except for the Upper Palaeolithic individuals/groups (Ust’-Ishim, Tianyuan, ANE, Villabruna) who stayed in the “right” set in all models. Diverse groups from Iran, Pakistan, Central Asia, and the Russian steppe zone (dated from the Chalcolithic to the

historical period) were used as targets for the *qpAdm* protocol. For each target, generally post-dating the proxy sources (Fig. 2b), progressively more complex admixture models were tested, from one- to five-way mixture models, and in most cases only the simplest feasible models were interpreted. Model feasibility criteria were as follows: estimated admixture proportions $\pm$ 2 standard errors are between 0 and 1, and *p*-value > 0.01 (Narasimhan et al. 2019).

In summary, this *qpAdm* protocol rotated a diverse set of groups between the "right" and "left" sets: from present-day to Mesolithic groups older than 10,000 years, and from Africans to South and East Asians (see date distributions in Fig. 2b). Another *qpAdm* protocol used by Narasimhan *et al*., termed "proximal" protocol (Box 1), relied on a smaller fixed set of groups that were kept always "on the right": 1) Mota (a 4500-years-old individual from Ethiopia); 2) East European Mesolithic individuals (EEHG, East European hunter-gatherers); 3) West Siberian Mesolithic (WSHG, West Siberian hunter-gatherers); 4) a Pre-pottery Neolithic (PPN) group from the Levant; 5) a Mesolithic group from the Iron Gates region (WEHG, West European hunter-gatherers); 6) Anatolian Neolithic individuals; 7) Ganj Dareh (a Neolithic group from Iran); 8) an Early Neolithic group from the Baikal region (ESHG, East Siberian hunter-gatherers). Thirty-one diverse Neolithic, Chalcolithic and Bronze Age groups from Eurasia were originally used as proxy sources in one- to three-way models, but if several feasible models were found for the target, the proxy sources from those models were moved one by one from the "left" to the "right" sets (i.e., model competition was performed). The set of targets and the model feasibility criteria matched those for the "distal" protocol. At the model competition step, groups very close in space and time appeared on both sides of the "left" - "right" and proxy source - target divides (Fig. 2c), making “left-to-right” gene flows (Fig. 1) highly likely. Of 45 target groups, 23 groups were also used as rotated proxy sources. For this reason, we interpreted the Narasimhan *et al*. proximal model competition protocol as follows (Box 1): its first step is a non-rotating *qpAdm* protocol with temporal stratification of “right” and “left” sets, but with no (or very limited) temporal stratification of targets and proxy sources; and the second step is a model competition protocol with no (or very limited, see Fig. 2c) temporal stratification of targets and proxy sources. We note that any such interpretation is an approximation that captures most important features of a published protocol and omits some details.

Since the sets of groups that are split into the "left" and "right" subsets in the protocols summarized above are very diverse chronologically and genetically, and since there are major overlaps in dates between the "left" and "right" subsets (Fig. 2), we argue that this approach is essentially similar to taking an admixture graph connecting populations sampled at widely different times in history, with divergence dates ranging from the Palaeolithic (up to ca. 87,000 yBP) to the "present" in the context of the dataset, and *randomly* splitting this graph into "left" and "right" population sets.

Testing qpAdm performance on complex simulation histories

Below we explore performance on simulated data (mainly FDR) of *qpAdm* protocols representing the spectrum of protocols used in the literature. The most extreme example is a protocol where all groups are rotated and all are treated alternatively as outgroups, targets, and proxy sources, i.e., there is no temporal stratification between the latter categories. We term this protocol “proximal rotating” (see Tables S1 and S2). Although such an extreme situation is, to our knowledge, rare among published *qpAdm* protocols (see Calhoff et al. 2021, Oliveira et al. 2022; in the latter study the rotating *qpAdm* strategy was used to model groups dated to 450–2600 yBP as mixtures of present-day groups), we use it to illustrate the effects of poor temporal stratification of targets and proxy sources in the case of a rotating protocol (Fig. 2). Models with targets pre-dating proxy sources are encountered in high-throughput *qpAdm* screens, but do not constitute a majority of models (Narasimhan et al. 2019, Librado et al. 2021, Allentoft et al. 2022, Bergström et al. 2022, Lazaridis et al. 2022, Taylor et al. 2023). We also explore FDR of the proximal non-rotating (Harney et al. 2018, van de Loosdrecht et al. 2018, Narasimhan et al. 2019, Prendergast et al. 2019, Wang et al. 2020, Calhoff et al. 2021, Wang et al. 2021, Zhang et al. 2021, Changmai et al. 2022a, Changmai et al. 2022b, Maróti et al. 2022, Brielle et al. 2023, Lee et al. 2023), distal rotating (Narasimhan et al. 2019, Librado et al. 2021, Allentoft et al. 2022, Bergström et al. 2022, Lazaridis et al. 2022, Taylor et al. 2023), and distal non-rotating protocols (Haak et al. 2015, Mathieson et al. 2015, Lazaridis et al. 2016, Antonio et al. 2019, Mathieson et al. 2018, Marcus et al. 2020, Yang et al. 2020, Papac et al. 2021, Yaka et al. 2021, Patterson et al. 2022) (Tables 1 and 2).

In the distal protocols, only *qpAdm* models where target group's sampling date is strictly contemporaneous with or post-dates sampling of both proxy sources were considered.

We tested performance of these *qpAdm* protocols on complex simulated genetic histories: 13 populations connected with each other by admixture graphs of random topology, including 10 pulse-like admixture events. Ten diploid individuals with no missing data were sampled from each population at "leaves" of the graph. Forty such random topologies were simulated, with an upper bound on the graph depth at 800 generations (ca. 23,000 years in the case of humans). These simulations generated sets of populations sampled at widely different dates in the past or, in other words, located at various genetic distances from the root, and matching intra-continental levels of human genetic differentiation (Fig. S1). Further details on the simulated population histories are presented in Methods and illustrated by five examples in Fig. S2. To explore the influence of data amount on *qpAdm* performance and compare it across protocols, we generated two independent sets of ten simulation replicates for each graph topology: with genomes composed of three or ten 100-Mbp-sized chromosomes (see Fig. S3 for the number of SNP loci polymorphic in the simulated datasets). These sets of simulations are referred to as "setup no. 1" and "setup no. 2" below. To explore the parameter space further, we generated single simulation replicates for two further setups: "setup no. 3", with maximal simulated history depth increased to 3,000 generations (ca. 87,000 yBP for humans), scaling all dates up proportionally; and "setup no. 4", with all terminal branches extended to the "present" of the simulation and sampled at that point. These latter simulations generated populations with median F_{ST} at the inter-continental level (no. 3) or below it (no. 4, Fig. S1).

A typical archaeogenetic dataset is composed of pseudo-haploid data with high proportions of missing sites and with widely different group sizes, including singleton groups. To generate more realistic "noisy" data, we also performed randomised subsampling of SNP datasets for simulation setups no. 1 and 2 (300- and 1,000-Mbp-sized genomes), for one simulation replicate per each setup (see Methods for details). The resulting data were pseudo-haploid, had missing data rates ranging from 5% to 95% across individuals, and had uneven group sizes ranging from 1 to 10 individuals. Ten independent subsampled datasets were generated for each simulated topology (400 replicates in total per simulation setup), including in the case of 300-Mbp-sized genomes from ca. 20,200 to 518,000 SNP loci with no missing data at the

group level and polymorphic across 13 groups (median = 89,700), and in the case of 1,000-Mbp-sized genomes from ca. 66,600 to 2,095,700 such SNPs (median = 259,400, Fig. S3).

As detailed in the preceding section, various versions of *qpAdm* protocols form a core of many recent archaeogenetic studies. These protocols are aimed at finding the simplest feasible *qpAdm* models for target groups, where feasibility is defined by setting a threshold for *qpAdm/qpWave* *p*-values and by setting plausibility criteria for admixture proportions estimated with *qpAdm*. Finding a feasible two-way or more complex admixture model for a target is often interpreted as solid evidence for gene flow, especially if the PCA and *ADMIXTURE* methods confirm the same signal. Thus, *qpAdm* protocols are used in fact as formal tests for admixture, whereas the latter two methods are not formal tests.

simulated topology no. 34:	polymorphic SNPs with no missing data	target	proxy source 1	proxy source 2	proxy source 1, inferred admixture proportion	proxy source 2, inferred admixture proportion	standard error of the admixture proportions	*qpAdm* *p*-value	*p*-value for 1-way model T=S1	*p*-value for 1-way model T=S2	*p*-value for 1-way model S1=S2	*f3* (T; S1, S2), Z-score
sampling date variable, 1,000 Mbp	1,796,083	H	A	C	63.8%	36.2%	1.4%	4.4E-26	1.8E-142	0.0E+00	0.0E+00	15.5
sampling date variable, 300 Mbp	539,274	H	A	C	62.4%	37.6%	2.3%	6.4E-06	9.8E-49	5.8E-125	0.0E+00	8.9
sampling date variable, deep divergence, 300 Mbp	1,032,128	H	A	C	60.1%	39.9%	1.1%	2.0E-62	6.3E-239	0.0E+00	0.0E+00	18.8
sampled at "present", 300 Mbp	768,376	H	A	C	68.0%	32.0%	3.2%	1.5E-02	9.6E-23	1.8E-66	1.3E-116	30.5
sampling date variable, 1,000 Mbp	1,796,083	H	A	G	57.4%	42.6%	1.6%	3.1E-18	1.8E-142	1.2E-228	0.0E+00	15.2
sampling date variable, 300 Mbp	539,274	H	A	G	60.6%	39.4%	2.5%	2.7E-05	9.8E-49	9.1E-112	3.2E-304	9.5
sampling date variable, deep divergence, 300 Mbp	1,032,128	H	A	G	53.6%	46.4%	1.7%	6.2E-34	6.3E-239	5.2E-279	0.0E+00	18.2
sampled at "present", 300 Mbp	768,376	H	A	G	61.0%	39.0%	3.1%	4.8E-01	9.6E-23	4.2E-88	3.4E-183	32.9

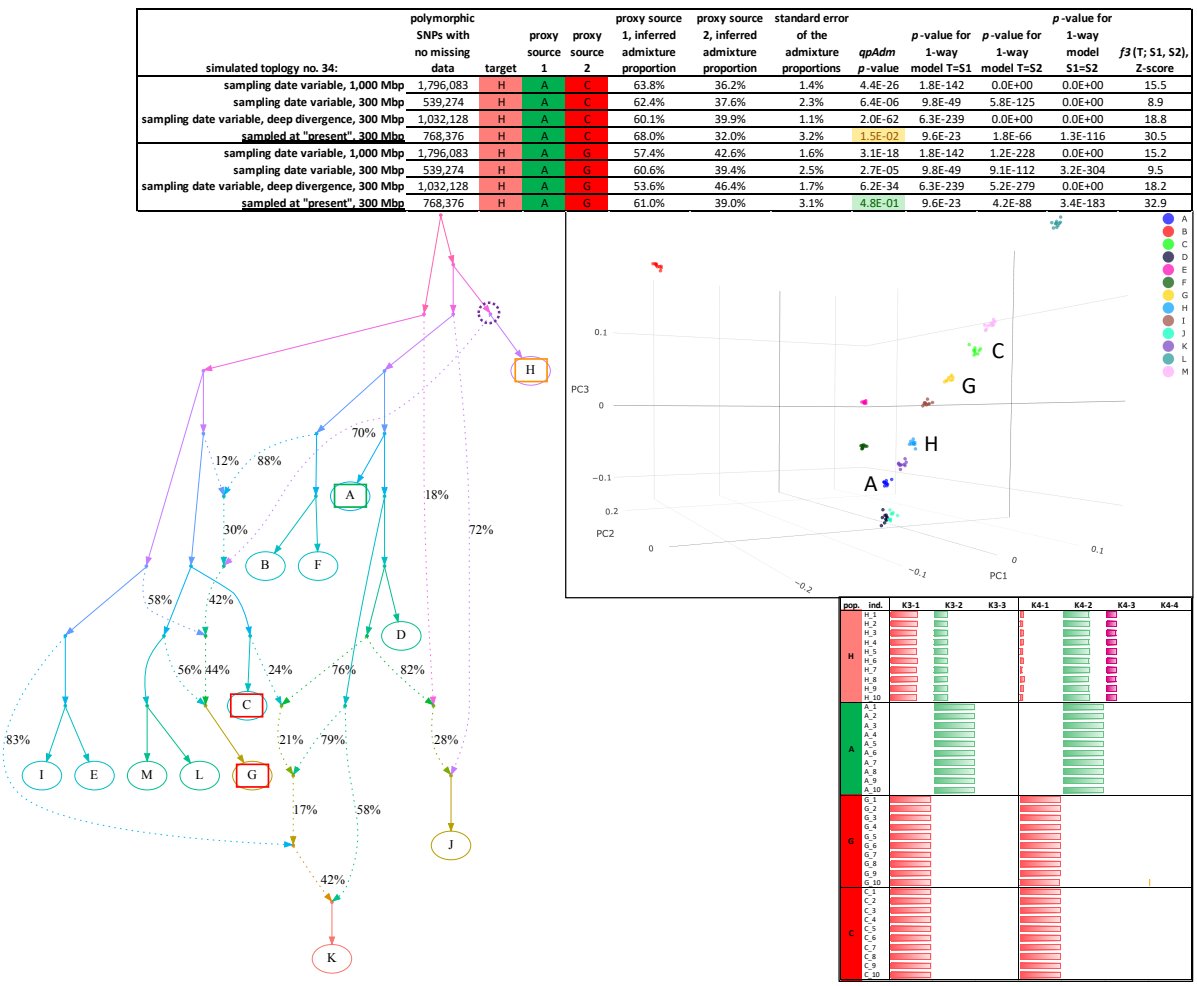


Figure 3. An example of the most common class of false positive *qpAdm* models supported by the proximal rotating protocol (accounts for 50.9% of all FP models across all the simulation and subsampling replicates in Table 2). Models of this type include at least one proxy ancestry source that is simulated as fully cladal with the target. The other proxy source may be simulated as a descendant

of the target lineage (see the model "*H = A + G*"), may belong to an unrelated lineage, or may be also cladal with the target (see the model "*H = A + C*"). Both models shown here are also fully supported by three-dimensional PCA and by an unsupervised *ADMIXTURE* analysis at one or more *K* values (under the simulation setup selected). For each model, the following information is shown: 1) simulation setups; 2) the number of polymorphic sites with no missing data at the group level; 3) admixture proportions and their standard errors estimated with *qpAdm*; 4) *p*-value of the two-way model; 5) *p*-values of the corresponding one-way models; 5) Z-score of the f_3-statistic f_3(target; proxy source$_1$, proxy source$_2$); 6) simulated admixture graph illustrating topological relationships among populations that are crucial for interpreting the models as false or true, but not dates of demographic events, sampling dates, and effective population sizes; 7) projection of a three-dimensional PCA plot with key groups labelled; 8) ancestry proportions estimated with unsupervised *ADMIXTURE* for the groups constituting the *qpAdm* models (results are shown for two selected *K* values). Items 6 to 8 are shown for the simulation setup whose name is underlined. Target groups are highlighted in orange throughout the figure; correct proxy sources are labelled in green, and incorrect ones in red. The true ancestry source for the target is marked by a dotted circle. The same selected (underlined) simulated history with dates, effective population sizes, and pairwise F_{ST} values is presented in Fig. S2a. For each simulation setup, results are shown for simulation replicate no. 1 only.

Relying on general principles, we argue that any high-throughput *qpAdm* protocol on poorly understood complex demographic relationships is questionable as a formal test for admixture since the *p*-value threshold allows to *reject*, but not to *accept* models, and it is safer to interpret those models as a certain number of gene flows connecting "left" and "right" sets in any direction, not in the form of proxy sources and admixture proportions for a target. The model feasibility criterion including both *p*-values and admixture proportions estimated with *qpAdm* is a complex construct relying on the topological assumptions outlined in Fig. 1. We expect that taking "left" and "right" sets that are not well-separated in time or contemporaneous (Fig. 2), and where relationships among groups are poorly understood (which is almost always true for exploratory studies), enriches the system for "left-to-right" gene flows, which in turn leads to frequent rejection of true simple admixture models. Since the behaviour of *qpAdm* admixture proportion estimates under different demographic scenarios is poorly understood, it is possible that a large fraction of these non-rejected complex models emerges as feasible, resulting in false signals of admixture.

The *qpAdm* protocols we applied to the simulated data were focused on the simplest models: one- and two-way admixture models (we note that histories that are more complex than two-way mixtures were common in the data, Fig. S2). The model feasibility criterion followed Narasimhan *et al*. (2019), see Box 1 for a definition. Thus, we tested all possible two-way

admixture models for 40 complex population histories (34,320 models per simulation setup and replicate).

The non-rotating *qpAdm* approach was implemented as follows: for each simulated graph six most ancient groups were selected as a fixed "right" set (ties were resolved in alphabetical order; these "right" sets remained unchanged for a given graph topology across independent simulations) and for the remaining seven groups all possible one-way and two-way admixture models were tested, applying the same composite feasibility criterion that was used for the rotating protocol.

In the context of complex and random admixture graph topologies it is hard to draw a strict boundary between true and false admixture models composed of a target and only two proxy sources. However, we classified the feasible *qpAdm* models into false and true ones relying on a set of rules. By far the most common class of false feasible *qpAdm* models (referred to as "false positive" or FP models), comprising 50.9% of all FP models generated by the proximal rotating protocol across all the simulation and subsampling replicates (setups no. 1 and 2), occurs when the target group is rejected as forming a clade with one or both proxy sources whilst they are simulated on graph topologies as clades. Interestingly, false cladality rejection accounted only for 10.1% of FP models generated by the proximal non-rotating protocol across all the simulation and subsampling replicates.

An example of this situation is shown in Fig. 3 where the clade relationship between the target (*H*) and source (*A*) is rejected due "left-to-right" gene flows violating the topological assumptions of *qpAdm*, and more complex models ("*H = A + C*" and "*H = A + G*") are evaluated as true. When a true one-way model "*H = A*" is tested, *A* represents a proxy for the only true source of ancestry in *H*, and outgroups *B*, *D*, and *F* split off the proxy branch after its divergence from the true source (this situation is shown in Fig. 1e), and ancestry in outgroups *J* and *K* is largely derived from that branch too (Fig. 1c), resulting in rejection of the one-way model with a very low *p*-value, ~10^{-23} (Fig. 3). Removal of all these outgroups (*B, D, F, J, K*) increases the *p*-value of the "*H = A*" model by many orders of magnitude, to 0.4. Models "*H = B/C/D/F*" are also rejected with *p*-values below ~10^{-34}. Interestingly, not only the one-way models, but all two-way models "*H = A/B/C/D/F + X*" were rejected according to *p*-values under at least three of four simulation setups (Fig. 3; removal of certain outgroups was not

done). The FP models shown in Fig. 3 under simulation setup no. 4 cannot be filtered out by temporal stratification of targets and sources since all groups were sampled at "present".

Other topological classes of FP models can be concisely described as follows (and are more precisely defined in Methods): 1) a proxy source included in the model is symmetrically related to all real sources of ancestry in the target (see an example of such feasible *qpAdm* models in Fig. S4a), or both proxy sources represent the same true source and are symmetrically related to all the other true sources (Fig. S4b); 2) both proxy sources represent distinct real sources, however a proxy source is a heavily admixed group sharing less than 40% of ancestry with the real source (Fig. S4c); 3) gene flow goes from the target lineage (after the last admixture event in its history) to the proxy source lineage, not in the opposite direction (Fig. S4d). Two-way admixture models for targets with population history best approximated with three-way and more complex models were considered as true positives if they included source proxies *not* satisfying the false positivity criteria listed above for at least two true sources. We also note that the class of models classified as true positive (TP) was not restricted to those including most optimal source proxies, if the models do *not* satisfy the false positivity criteria. On a random sample of 400 two-way admixture models from our 40 simulated histories, the fraction of models that were classified as appropriate (true) according to the rules described above was 17.7%. Since groups that are truly admixed are common in our simulations, we do not expect to encounter a "needle in a haystack" situation where finding true admixture models is exceedingly hard.

Violations of the topological assumptions of *qpAdm* encountered in the examples of FP model classes are described below. All the models shown below were selected among feasible models that were outcomes of the proximal rotating protocol, and, for simplicity, *qpAdm*, PCA, and *ADMIXTURE* results are presented for one simulation replicate per simulation setup. In Fig. S4a, the target group, *A*, was simulated as a two-way mixture, and we expect that models "*A = B + C/K*" would be fitting. The topological assumptions are violated when testing these models (Fig. S4a): groups *C* and *K* are cladal, with one of them appearing in the "left" set and the other one in the "right" set (see Fig. 1e); a gene flow from an outgroup branch, *D*, enters the *C/K* branch after it splits from one of the true ancestry sources (Fig. 1a). However, *p*-values of the models "*A = B + C/K*" are high (0.93, 0.97), and they were rejected due to estimated admixture proportions being negative. Removal of the outgroups *C, D,* and

K does not make these models fitting. In contrast, incorrect models "*A = B + G/J*" emerged as fitting (Fig. S4a), where the proxy sources *G* and *J* are symmetrically related to both true ancestry sources in the target.

In Fig. S4b, the target group, *A*, was simulated as a two-way mixture, and both correct (*A = F + B*) and incorrect (*A = F + M*) models are fitting, suggesting that violations of the topological assumptions play no role in the emergence of this false positive model. Failure to reject the model "*A = F + M*", where both proxy sources represent only one true ancestry source and are symmetrically related to the other, may be attributed to the lack of data, however increasing the simulated genome size from 300 Mbp to 1,000 Mbp or increasing the simulated graph depth from 800 to 3,000 generations resulted in rejections of both the correct and incorrect models with *p*-values below ~10^{-5} (Fig. S4b).

In Fig. S4c, the target group, *E*, was simulated as a two-way mixture, but no appropriate source proxy was simulated for one of the true ancestry sources: in groups *H* and *A*, 30% and 6% of their ancestry, respectively, is derived from that true source. Thus, multiple gene flows from the "right" set enter the *A* lineage after its split from the true ancestry source, and the same is true for the *H* lineage, and these are violations of the topological assumptions (Fig. 1a). The model "*E = H + L*" was rejected according to *p*-values under all four simulation setups, and the model "*E = A + L*" was rejected according to *p*-values when more data was available (the simulated genome size increased from 300 Mbp to 1,000 Mbp, or the simulated graph depth increased from 800 to 3,000 generations, see Fig. S4c). Accepting models like "*E = A + L*" at face value may lead to erroneous historical interpretations, but all thresholds for classifying models of this type into false and true are arbitrary. We chose 40% as a threshold percentage of proxy source's ancestry derived from the corresponding true source.

In Fig. S4d, the target group, *G*, was simulated as a two-way mixture, and an incorrect model "*G = C + J*" emerged as fitting. Here, *J* is a descendant of *G* (87% of its ancestry) rather than a proxy source (only ~2% of its ancestry is derived from one of the true ancestry sources). The model "*G = C + J*" was rejected according to *p*-values when more data was available (the simulated genome size increased from 300 Mbp to 1,000 Mbp, or the simulated graph depth increased from 800 to 3,000 generations, see Fig. S4d). Correct models such as "*G = A + J*" were rejected according to *p*-values (Fig. S4d) due to violations of the topological

547 assumptions: for instance, outgroups *C* and *I* are derived from the target lineage after the
548 admixture event (Fig. 1d). However, removal of various violating outgroups did not make the
549 model "*G = A + J*" and similar models "*G = L/M + J*" fitting.

550 While FP *qpAdm* models are expected for complex genetic histories, the practical usage of
551 the *qpAdm* relies on an assumption that false positives are rare. However, FDR of the four
552 *qpAdm* protocols tested here (rotating and non-rotating, proximal and distal) varied between
553 12.1% and 68.1% (across all simulation setups and replicates summarized in Table 1). Key
554 statistics in our study are false discovery rate (FDR) and false omission rate (FOR); see Box 1
555 for definitions. We estimated FDR and FOR instead of false positive and false negative rates
556 of the *qpAdm* protocols and other methods due to a technical limitation: the process of model
557 classification into true and false ones cannot be fully automated since it requires careful
558 interpretation of the topology and simulated admixture proportions. Therefore, classifying
559 all possible 34,320 two-way admixture models (858 per simulated topology) into true and
560 false was hardly feasible. We estimated false positive (FPR) and true positive rates (TPR) only
561 approximately, relying on the fractions of negatives and positives in random samples of two-
562 way admixture models (see below). FPR in our context is the probability that a two-way model
563 with an unadmixed target and/or inappropriate proxy source(s) emerges as fitting in a single
564 *qpAdm* test. TPR in our context is the probability that a two-way model with an admixed
565 target and both proxy sources being appropriate emerges as fitting in a *qpAdm* test.

566

567 ***Influence of the amount of data and temporal stratification on the performance of qpAdm***
568 ***protocols***

569 The amount of data (3.3-fold difference in simulated genome sizes) had no influence on FDR
570 of *qpAdm* protocols in the case of randomly subsampled pseudo-haploid data (no statistically
571 significant difference was found between sets of 10 subsampling replicates in the case of four
572 *qpAdm* protocols, Table 2, Fig. 4). In contrast, small but statistically significant influence of
573 the data amount on FDR (according to the two-sided Wilcoxon test) was observed for three
574 of four *qpAdm* protocols applied to high-quality data: proximal rotating and non-rotating, and
575 distal non-rotating (Table 2, Fig. 4).

simulation scheme	max. depth of the simulation, gen.	pseudohaploid noisy data	genome size, Mbp	model class	protocol	number of two-way models in the respective class (min.-max.)	fraction of models that are distal	min.-max. across 10 replicates	FDR for proximal & distal models	min.-max. across 10 replicates	FDR for distal models	min.-max. across 10 replicates	number of two-way models in the respective class	fraction of models that are distal	FDR for proximal & distal models	FDR for distal models	number of two-way models in the respective class	fraction of models that are distal	FDR for proximal & distal models	FDR for distal models
													supported by 3D PCA				supported by unsupervised *ADMIXTURE*			
date variable, no. 1 (noisy)	800	yes	300	FP	non-rotating	75 (62-95)	37.3%	30.6%-42.9%	37.4%	29.5%-42.3%	24.2%	17.6%-28.1%								
					rotating	642 (579-733)	20.9%	17.8%-24.6%	53.0%	51.6%-54.9%	25.1%	22.4%-29.0%								
				TP	non-rotating	146 (115-149)	67.0%	60.0%-76.4%												
					rotating	554 (492-613)	68.6%	66.7%-73.8%												
date variable, no. 2 (noisy)	800	yes	1000	FP	non-rotating	86 (76-102)	36.2%	28.2%-47.3%	39.3%	34.1%-41.2%	25.1%	18.6%-32.6%								
					rotating	417 (355-500)	19.4%	17.7%-23.5%	54.6%	51.7%-59.7%	24.0%	21.8%-27.9%								
				TP	non-rotating	133 (126-164)	66.8%	62.7%-72.5%												
					rotating	340 (297-405)	74.0%	70.1%-81.3%												
date variable, no. 1	800	no	300	FP	non-rotating	76 (62-92)	39.2%	34.5%-48.0%	42.6%	35.8%-45.5%	31.2%	24.5%-34.4%	20	30.0%	35.1%	18.2%	2	100.0%	4.5%	6.1%
													72	40.3%			90	36.7%		
					rotating	247 (194-268)	16.4%	10.5%-19.1%	57.4%	55.6%-62.6%	22.1%	14.4%-27.9%	129	7.8%	61.4%	14.3%	59	3.4%	49.6%	4.4%
													83	36.1%			153	24.8%		
				TP	non-rotating	105 (93-115)	65.5%	59.5%-71.2%					37	73.0%			42	73.8%		
													72	59.7%			67	58.2%		
					rotating	165 (149-202)	80.8%	78.1%-84.1%					81	74.1%			60	71.7%		
													92	89.1%			113	87.6%		
date variable, no.2	800	no	1000	FP	non-rotating	43 (32-62)	40.4%	26.8%-53.1%	38.4%	31.3%-41.5%	26.5%	16.4%-30.4%	14	35.7%	40.0%	23.8%	1	0.0%	3.2%	0.0%
													18	66.7%			31	54.8%		
					rotating	127 (106-139)	10.5%	6.0%-16.7%	62.0%	57.6%-68.1%	16.4%	12.1%-26.7%	84	3.6%	70.6%	10.3%	18	0.0%	40.9%	0.0%
													30	26.7%			96	11.5%		
				TP	non-rotating	75 (65-95)	66.0%	58.9%-74.7%					21	76.2%			30	80.0%		
													49	73.5%			40	70.0%		
					rotating	75 (65-84)	83.6%	80.0%-88.1%					35	74.3%			26	69.2%		
													49	91.8%			58	91.4%		
date variable, no. 3	3000	no	300	FP	non-rotating	36	38.9%		36.0%		24.1%		12	33.3%	34.3%	20.0%	5	20.0%	12.8%	5.3%
													24	41.7%			31	41.9%		
					rotating	61	16.4%		66.3%		25.6%		41	7.3%	77.4%	23.1%	12	8.3%	46.2%	7.7%
													20	35.0%			49	18.4%		
				TP	non-rotating	64	68.8%						23	69.6%			34	52.9%		
													41	68.3%			30	86.7%		
					rotating	31	93.5%						12	83.3%			14	85.7%		
													19	100.0%			17	100.0%		
date variable, no. 4 "present"	800	no	300	FP	non-rotating	99	40.4%		42.1%		30.3%		53	64.2%	36.1%	34.0%	18	72.2%	20.5%	19.1%
													46	13.0%			81	33.3%		
					rotating	364	17.6%		64.0%		29.8%		178	22.5%	59.5%	29.9%	59	22.0%	37.8%	15.1%
													186	12.9%			305	16.7%		
				TP	non-rotating	136	67.6%						94	70.2%			70	78.6%		
													42	61.9%			66	56.1%		
					rotating	205	73.7%						121	77.7%			97	75.3%		
													84	67.9%			108	72.2%		

Table 1. Assessing the effect of temporal stratification of targets and proxy sources on FDR of non-rotating and rotating *qpAdm* protocols and their combinations with PCA and *ADMIXTURE* analyses**.** Another kind of temporal stratification, stratification of the "right" and "left" sets, was a part of the non-rotating protocols, but not of the rotating ones. FDR values are highlighted in the respective columns and color-coded. In the right half of the table these FP and TP classes are further subdivided into those supported/not supported (highlighted in green and red, respectively) by another method (PCA or *ADMIXTURE*). For each model class and sub-class, fractions of models in the FP and TP classes that are distal are shown with magenta bars. If all groups were sampled at present (simulation setup no. 4), selection of time-stratified admixture models was performed as if they were sampled in the past (as in simulation setups no. 1-3) since simulated topologies were the same across all the setups and differed only in the amount of genetic drift on graph edges. In the case of simulation setups no. 1 and 2, 10 simulation replicates and 10 subsampling replicates derived from simulation replicate no. 1 were generated, and median, minimal, and maximal values across the replicates are shown for counts of FP and TP *qpAdm* models, for fractions of models that are distal, and for FDR.

In the case of the most extreme *qpAdm* protocol, proximal rotating, adding data increased median FDR from 57.4% to 62%, and this difference is statistically significant (Table 2). A large fraction of false positives in the case of the proximal rotating protocol emerges due to false rejections of one-way models because of violations of the topological assumptions (Fig. 1), for instance, due to "left-to-right" gene flows, and that prompts the investigator to test more complex two-way models, which often emerge as feasible. And model rejection is more efficient when more data are available, explaining the effect on FDR observed here. As we show in Suppl. text 1, it is probably impossible to decrease FDR of the proximal rotating protocol dramatically by combining it with other methods (PCA and unsupervised *ADMIXTURE*) as additional controls or by adjusting *p*-value thresholds in the *qpAdm* protocol.

two-sided Wilcoxon test, *p*-value (paired test if *qpAdm* protocols differ, but simulation parameters are the same)											
max. depth of the simulation: 800 gen.		proximal rotating	proximal non-rotating	distal rotating	distal non-rotating	proximal rotating	proximal non-rotating	distal rotating	distal non-rotating
date variable sampling		300-Mbp-sized genomes				1000-Mbp-sized genomes			
subsampling replicates on pseudohaploid data									
proximal rotating	300 Mbp	53.0%	0.002	0.002	0.002	0.166			
proximal non-rotating	300 Mbp		37.4%	0.002	0.002		0.190		
distal rotating	300 Mbp			25.1%	0.160			0.579	
distal non-rotating	300 Mbp				24.2%				0.186
proximal rotating	1000 Mbp					54.6%	0.002	0.002	0.002
proximal non-rotating	1000 Mbp						39.3%	0.002	0.002
distal rotating	1000 Mbp							24.0%	0.625
distal non-rotating	1000 Mbp								25.1%
simulation replicates on high-quality data									
proximal rotating	300 Mbp	57.4%	0.002	0.002	0.002	0.009			
proximal non-rotating	300 Mbp		42.6%	0.002	0.002		0.009		
distal rotating	300 Mbp			22.1%	0.010			0.143	
distal non-rotating	300 Mbp				31.2%				0.003
proximal rotating	1000 Mbp					62.0%	0.002	0.002	0.002
proximal non-rotating	1000 Mbp						38.4%	0.002	0.002
distal rotating	1000 Mbp							16.4%	0.002
distal non-rotating	1000 Mbp								26.5%

Table 2. Comparing FDR of four *qpAdm* protocols, proximal and distal, rotating and non-rotating, on 10 subsampling replicates derived from a single simulation replicate (low-quality data) or on 10 simulation replicates and high-quality genomes (300-Mbp-sized or 1000-Mbp-sized genomes). For details on generating replicates of low-quality data see Methods. Median FDR values for *qpAdm* protocols are shown on the diagonal. The protocols were compared using the two-sided Wilcoxon test applied to FDR values, see the cells above the diagonal.

However, in the case of both proximal and distal non-rotating protocols, adding data led to a small but statistically significant decrease in FDR: 42.6% vs. 38.4%, 31.2% vs. 26.5% (Table 2,

Fig. 4), suggesting that false model rejections due to assumption violations play a less important role here, which is expected for "right" and "left" population sets which are stratified temporally. We found no significant effect of the data amount on FDR in the case of the distal rotating protocol, which demonstrated the best median FDR values on high-quality data overall (22.1 and 16.4%, Table 2, Fig. 4). We did not compare FDR between high-quality and low-quality datasets with the Wilcoxon test since replicates in these cases were generated differently (in the latter case subsampling replicates were derived from one simulation replicate per simulation setup, while in the former case ten simulation replicates were considered per simulation setup), but we note that random subsampling of SNPs and individuals and a less robust algorithm for calculating f_4-statistics (with different statistics calculated on different SNP sets) did not lead to dramatic increases/decreases in FDR (Table 2, Fig. 4).

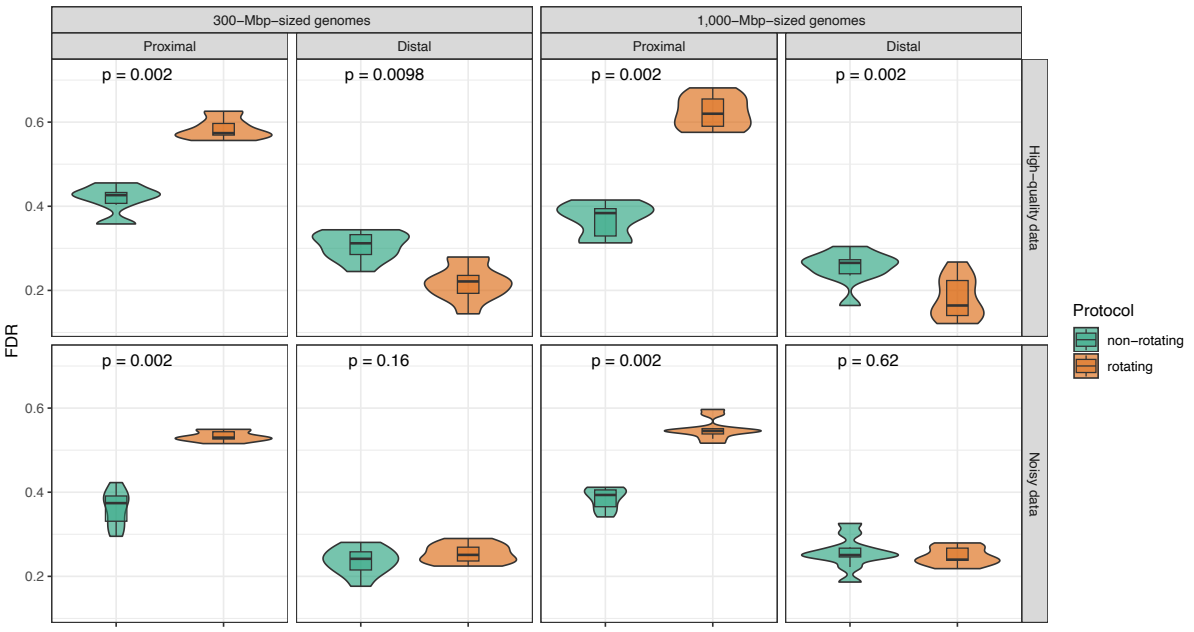


Figure 4. Distributions of FDR values across 10 simulation replicates (for high-quality data) or across 10 subsampling replicates derived from a single simulation replicate (for low-quality data). Distributions are summarized with boxplots and violin plots for four *qpAdm* protocols (proximal and distal, rotating and non-rotating) and two simulated genome sizes (300 Mbp and 1,000 Mbp). *qpAdm* protocols were compared with the paired two-sided Wilcoxon test, and *p*-values for "rotating vs. non-rotating" comparisons are shown in the panels (for the other *p*-values see Table 2).

Next, we move on to assessing the influence of *qpAdm* protocol details on FDR. Temporal stratification of targets and proxy sources (the former are not allowed to pre-date the latter) is the best way of reducing FDR, according to our analysis: from ~50%–60% to ~15%–25% for rotating protocols, and from ~40% to ~25%–30% for non-rotating protocols (Table 2, Fig. 4). All these differences are statistically significant according to the two-sided Wilcoxon test (the paired version of the test was used in this case since different *qpAdm* protocols applied to the same simulation/subsampling replicate are not totally independent experiments). Temporal stratification of "right" and "left" sets (i.e., the non-rotating protocol) is helpful in the absence of the former type of temporal stratification, of targets and proxy sources: FDR drops from ~50%–60% to ~40% (these differences are also statistically significant, Table 2, Fig. 4). However, it is not helpful (no significant difference) or even damaging to *qpAdm* performance (significantly worse) when applied to distal protocols (Table 2, Fig. 4). This result supports the conclusion by Harney et al. (2021) that rotating *qpAdm* protocols should be preferred to non-rotating ones. However, according to our analysis, this conclusion is conditional on strictly enforced temporal stratification of targets and proxy sources since the rotating protocol without such stratification ("proximal rotating") demonstrated by far the worst performance.

Another observation is that absolute FDR values are high for all *qpAdm* protocols tested (median FDR values below 16.4% were not observed), however these absolute values are expected to depend on the complexity of simulated histories and on the amount of data (Table 2), which also depends on the time depth of simulated histories (Fig. S3, see Discussion).

The fraction of two-way admixture models that are inappropriate according to our topological criteria (the fraction of negatives) in a random sample of 400 models from all simulated graphs is 82.3%, which allows us to approximately estimate not only FDR (51.6%–68.1%, see Table 1), but the FPR of the proximal rotating protocol = number of false positives per simulation replicate / number of negatives per simulation replicate [858 models × 40 graphs × fraction of negatives] = 0.4%–2.6% across simulation parameter combinations and replicates summarized in Table 2. The TPR of the proximal rotating protocol = number of true positives per simulation replicate / number of positives per simulation replicate [858 models

× 40 graphs × (1 – fraction of negatives)] = 1.1%–10.1%. Here are the same statistics for the distal non-rotating protocol: the fraction of negatives in a random sample of 400 distal non-rotating models from all simulated graphs, 74.3%; total number of distal non-rotating two-way models across all graphs, 1804; FDR across simulation parameter combinations and replicates from Table 2, 16.4%–34.4%; FPR, 0.8%–3.2%; TPR, 9.7%–24.4%. Thus, although FPR of a single *qpAdm* test is low, due to the relatively high proportion of negatives among all models, the large number of models tested in high-throughput *qpAdm* screens, and the low TPR, FDR becomes high, compromising historical interpretation of such screens for admixture.

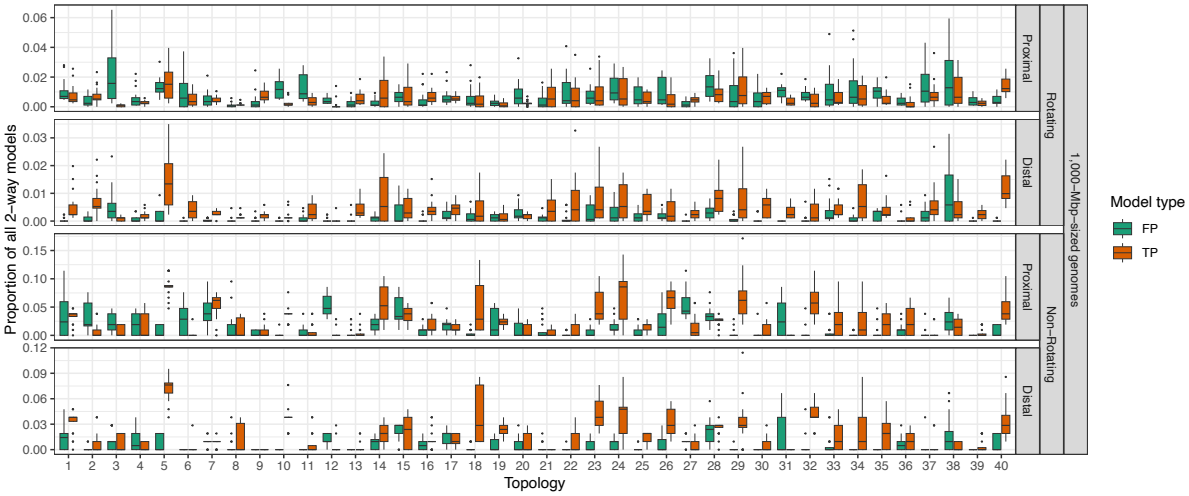


Figure 5. Proportions of all possible two-way *qpAdm* models that are false positive (FP), or true positive (TP), binned by simulated graph topology. There are 858 two-way admixture models per simulated graph including 13 groups if the rotating *qpAdm* protocol is applied, and 105 models if the non-rotating protocol is applied. For brevity, results are shown for simulation setup no. 2 (high-quality data) only. The boxplots summarize distributions of FP and TP model fractions across simulation replicates.

The fraction of feasible *qpAdm* models that are false (FP) varies a lot depending on simulated graph topology (Fig. 5), and hence it is hard to predict if a particular real genetic history is prone to generating FP signals of admixture when *qpAdm* protocols are applied. Among 80 combinations of proximal *qpAdm* protocols (rotating or non-rotating), simulation setups, and simulation/subsampling replicates we tested, in one case only a topology accounts for >20% of FP *qpAdm* models found across all the 40 simulated topologies. In contrast, in the case of distal *qpAdm* protocols, results are much more uneven across topologies: for 25 of 80

686 "protocol/simulation setup/replicate" combinations, at least one topology accounts for >20%
687 of FP *qpAdm* models found across all the 40 simulated topologies. Three topologies most
688 problematic for distal *qpAdm* protocols are illustrated in Fig. S5:

689

690 ***Admixture inference pipelines and model competition qpAdm protocols***

691 An implicit assumption of many archaeogenetic studies relying on *qpAdm* protocols is that
692 admixture models supported by clines observed in (usually two-dimensional) spaces of
693 principal components, and/or by an *ADMIXTURE* analysis, and/or by individual *D*-, f_4- or f_3-
694 statistics are especially robust. And, *vice versa*, *qpAdm* results are often interpreted as a
695 formal test of hypotheses about admixture formulated based on PCA and/or *ADMIXTURE*
696 results. We constructed "admixture inference pipelines" composed of a *qpAdm* protocol and
697 one or two further methods to test these assumptions on simulated data. We note that all
698 signals of admixture revealed by our PCA or *ADMIXTURE* analyses were not explored with
699 *qpAdm* since exhaustive lists of positive and negative two-way admixture models were not
700 compiled for each simulated graph. *Vice versa*, all feasible *qpAdm* models were checked by
701 the PCA and/or *ADMIXTURE* methods.

702 We considered a two-way admixture model to be supported by PCA if the target group was
703 located on a straight line between the two proxy source groups in the space of three PCs
704 when all 13 simulated groups were co-analysed. Deviation from the straight line was
705 acceptable to some extent as non-linear PCA clines are often observed on real data (de Barros
706 Damgaard et al. 2018, Jeong et al. 2019), and they were also common among TP two-way
707 *qpAdm* models in this study (see Methods for details and Figs. 3 and S4 for examples). This
708 situation is expected since many targets in our simulations represent three-way and more
709 complex mixtures, and since arrangement of populations in the PC space is influenced not
710 only by admixture, but also by genetic drift (McVean 2009). Our requirements for a model to
711 be declared supported by PCA were more stringent than those usually applied in the
712 literature since we considered three-dimensional PC spaces instead of two-dimensional ones.
713 Also see Methods for the rules we used to judge if an admixture model is supported by an
714 unsupervised *ADMIXTURE* analysis, and see Figs. 3 and S4 for examples.

A much more limited form of group rotation, "model competition", is used in the literature widely (Narasimhan et al. 2019, Fernandes et al. 2020, Calhoff et al. 2021, Sirak et al. 2021, Zhang et al. 2021, Maróti et al. 2022, Brielle et al. 2023, Lee et al. 2023), and we explored FDR of this method as well. A typical model competition protocol (Narasimhan et al. 2019, Maróti et al. 2022, Brielle et al. 2023) consists of two stages. First, the oldest, e.g., Palaeolithic, populations (and/or those most divergent from the target group) are used as a fixed "right" set, and populations sampled at later dates are used as proxy sources and targets. As usual, progressively more complex models are tested for targets of interest, and a composite feasibility criterion is applied.

In many publications (e.g., Haak et al. 2015, Mathieson et al. 2015, Antonio et al. 2019, Mathieson et al. 2018, Prendergast et al. 2019, Marcus et al. 2020, Papac et al. 2021, Wang et al. 2021, Yaka et al. 2021, Changmai et al. 2022a, 2022b, Patterson et al. 2022) this first non-rotating step remains the only *qpAdm* protocol used (in its distal or proximal forms). In a model competition protocol, subsequent analysis is focused on targets for whom two or more alternative *qpAdm* models emerge as feasible at the first step. For each target, alternative proxy sources are pooled and rotated between the "left" and "right" sets, testing only the models that emerged as feasible at the first step and applying a composite feasibility criterion (e.g., p-value > 0.01, estimated admixture proportions ± 2 SE are within 0 and 1).

Rotation of alternative proxy sources can be performed in various ways: "whatever is not on the left is on the right" (Brielle et al. 2023), or placing alternative sources in the "right" set one by one (Calhoff et al. 2021, Maróti et al. 2022, Brielle et al. 2023). In the latter case several "right" sets are tested for each model, and the model is considered supported by the model competition protocol only if it is not rejected under any of these "right" sets (Maróti et al. 2022). The reasoning behind this protocol is as follows: model rejection due to violations of the topological assumptions of *qpAdm* is not expected for a model composed of sources very close to the true ones since in this case branches private to the proxy sources are short, and it is unlikely that gene flows to or from the "right" set happened on these short branches. Models composed of sources closely related to the true ones are also not expected to be rejected when more distant proxy sources are placed in the "right" set (Harney et al. 2021).

For reasons detailed in the Discussion section, we explored the *qpAdm* model competition protocol and multi-method admixture inference pipelines on one replicate per simulation setup, and three (Table 3) or four (Table 1) simulation setups were involved in this analysis.

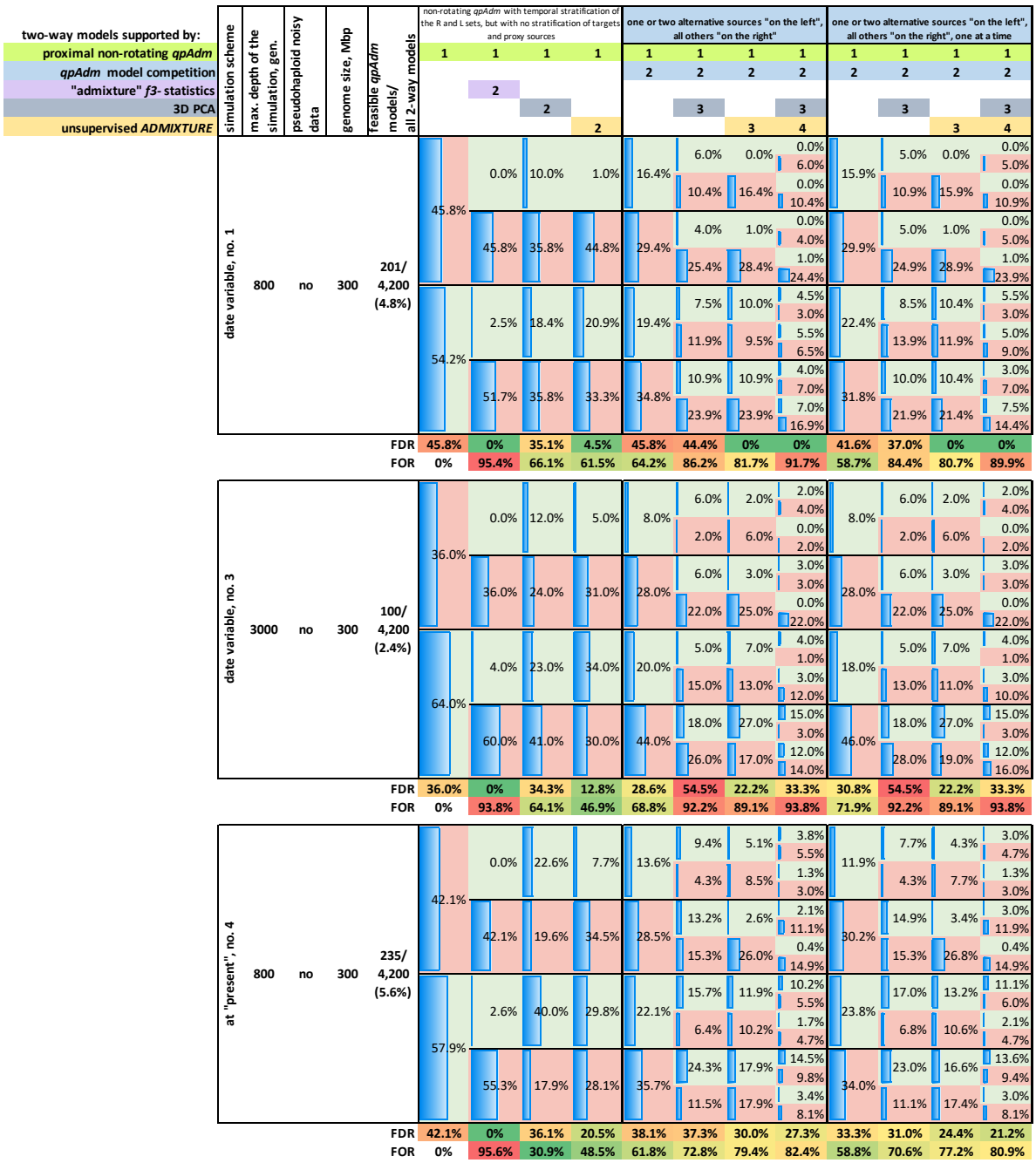


Table 3. Assessing FDR of the proximal non-rotating *qpAdm* protocol combined with model competition (Narasimhan et al. 2019). For each simulation setup, the analysis relies on simulation replicate no. 1 only. For constructing pipelines resembling those common in the archaeogenetic literature, we used five methods: proximal non-rotating *qpAdm*, *qpAdm* model competition (two alternative protocols), "admixture" f_3-statistics, 3D PCA with all individuals co-analysed, and unsupervised *ADMIXTURE* with all individuals co-analysed. For declaring a positive result, support of an admixture model by all methods in the pipeline was required, hence the order of methods is not

important except for the first method, which was the proximal non-rotating *qpAdm* in all cases. In each column, methods comprising a pipeline are color-coded and numbered by their order. All feasible two-way *qpAdm* models emerging as outcomes of the proximal non-rotating protocol were classified into FP and TP (highlighted in red and green in the leftmost column, respectively). The other columns are structured like bifurcating trees: FP *qpAdm* models supported by method no. 2; FP *qpAdm* models not supported by method no. 2; TP *qpAdm* models supported by method no. 2; TP *qpAdm* models not supported by method no. 2. The same principle is used for representing results of more complex pipelines. All model counts are normalized by the number of feasible *qpAdm* models (FP + TP), outcomes of the first method. Percentages of models supported/not supported by the last method in the pipeline are highlighted in green and red, respectively. FDR values are shown for these different pipelines. Fractions of TP *qpAdm* models that are pruned out by progressively more stringent support requirements are also shown (false omission rate or FOR).

The two alternative model competition protocols described above were applied to targets for whom more than one model was feasible given a fixed "right" set. If only one model was feasible for a target, such a model was evaluated as passing model competition (such models accounted only for 7%–11% of models feasible at the first step). The model competition protocols failed to improve *qpAdm* performance: FDR ranged from 29% to 46% (as compared to 36%–46% prior to the model competition step), and both model competition protocols demonstrated very similar results (Table 3). FOR of the model competition protocols varied from 59% to 72%. FDR also remained high for models supported by proximal non-rotating *qpAdm* & PCA or by proximal non-rotating *qpAdm* & model competition & PCA (Table 3).

Considering all simulation setups and replicates shown in Table 2, there were only 1,591 instances when a two-way admixture model was supported by both the proximal rotating and proximal non-rotating protocols on the same simulated data. In contrast, there were 5,844 and 24,046 instances when a model was supported exclusively by the proximal non-rotating and rotating protocols, respectively. Notably, FP models supported by the proximal non-rotating *qpAdm* protocol largely lacked support by an unsupervised *ADMIXTURE* analysis (Table 3), in contrast to outcomes of the proximal rotating protocol (Table S1). FDR of a pipeline composed of these two methods ranged from 5% to 21% across three simulation setups tested (Table 3). Adding a model competition step to this pipeline increased both FDR and FOR in 4 of 6 cases; and, in general, the proximal non-rotating *qpAdm* protocol combined with *ADMIXTURE* is the best-performing protocol in this analysis (Table 3) according to FDR and FOR.

789 The fact that our *ADMIXTURE* analysis supports a large fraction of FP two-way mixture models
790 emerging as outcomes of the proximal rotating *qpAdm* protocol reflects known problems in
791 modelling using *ADMIXTURE* very ancient individuals in the context of modern populations.
792 These individuals are often modelled (Raghavan et al. 2014, Haak et al. 2015, Moreno-Mayar
793 et al. 2018a) as complex mixtures of ancestry components typical for modern populations,
794 which is obviously an artefact. Sampling dates for unique targets from FP models supported
795 by both proximal rotating *qpAdm* and *ADMIXTURE* ranged from the present to 665
796 generations in the past (median = 406 generations), while the median sampling date for
797 unique targets from both FP and TP models supported by proximal rotating *qpAdm* was 215
798 generations in the past (for comparison across simulation setups, all the dates were rescaled
799 to a maximum simulation depth of 800 generations). The proximal non-rotating protocol by
800 design did not consider the oldest groups as targets for *qpAdm* and *ADMIXTURE*, thus largely
801 avoiding this problem (sampling dates for unique targets from FP models supported by both
802 proximal non-rotating *qpAdm* and *ADMIXTURE* ranged from the present to 366 generations
803 in the past; median = 44 generations).

804 Above we have discussed multi-method pipelines based on the proximal non-rotating *qpAdm*
805 protocol. Combining distal *qpAdm* protocols with PCA allows to reduce FDR of both rotating
806 and non-rotating protocols further, to 10%–24%, and distal *qpAdm* protocols combined with
807 an unsupervised *ADMIXTURE* analysis demonstrated even better FDR values ca. 0%–8%
808 (Table 1). If target and proxy source populations are sampled at approximately the same time
809 (such as those from our simulation setup no. 4 and from the proximal analysis in Narasimhan
810 et al. 2019, Fig. 2c) applying this approach is impossible. However, if our simulations with all
811 branches extended to the present are treated in the same way as their topological
812 counterparts with date-variable sampling, performance gains (decrease in FDR) of temporal
813 stratification of admixture models are similar to those mentioned above (Table 1). In this case
814 the temporal stratification procedure retains models with the latest admixture event in
815 target's history that is more recent than (or as recent as) the latest admixture events in proxy
816 sources' history. However, in the case of simulation setup no. 4 performance gains of the
817 "*qpAdm* + PCA" and "*qpAdm* + *ADMIXTURE*" method combinations were moderate (Table 1).

818

Discussion

In this study we explored performance of various *qpAdm* protocols on a collection of random complex simulated genetic histories, where admixture history of target groups may vary from the simplest (no admixture) to very complex. It is because of this research design and other limitations discussed below that our study is focused mostly on one performance metric: false discovery rate or FDR. In simple terms, we focused our analysis only on models of a chosen complexity class (two-way models) supported by a *qpAdm* protocol (feasible models), classified them manually into false and true positives according to a set of topological rules, and subjected them to further screening by PCA and/or *ADMIXTURE* methods. We did not attempt to classify rejections of two-way models by *qpAdm* or other methods into false rejections due to violations of the topological assumptions of *qpAdm* (Fig. 1) and true rejections when the true admixture history of the target does not fit a two-way model. This problem was deliberately left out since in the literature more attention is paid to interpretation of “fitting” (“feasible” or “positive”) than rejected *qpAdm* models.

Another limitation of our study is that we had to use idealized versions of *qpAdm*, PCA, and *ADMIXTURE* protocols, while in the archaeogenetic literature manual adjustment of analytical protocols is common: protocols often vary from one target group to another (see, e.g., Lazaridis et al. 2016, Zhang et al. 2021, Brielle et al. 2023, Lee et al. 2023) and from study to study. These extensive details are very hard to formalize and reproduce. In the case of *qpAdm* protocols, certain groups of populations may be placed exclusively in the “right" or in the “left" sets, with the rest rotated between these sets, and relative sizes and compositions of these three groups vary from study to study: in the case of model competition protocols, this rotated subset is small, and rotation may be restricted to a particular model complexity class, but in other cases it may encompass all or nearly all populations analysed (see, e.g., Narasimhan et al. 2019, Librado et al. 2021, Bergström et al. 2022, Lazaridis et al. 2022, Oliveira et al. 2022, Taylor et al. 2023). Reproducing all aspects of PCA and *ADMIXTURE* protocols used in the literature is also hardly possible on simulated data. For instance, PCs in archaeogenetic studies are usually calculated based on present-day populations, and ancient individuals are projected on the resulting PCs (e.g., Haak et al. 2015, Mathieson et al. 2018, Narasimhan et al. 2019, Furtwängler et al. 2020, Marcus et al. 2020, Lazaridis et al. 2022). In

849 contrast, in our study all simulated individuals were co-analysed for calculating PCs.
850 *ADMIXTURE* analyses in the literature are usually performed on worldwide or continent-wide
851 panels of populations that often overlap just partially with population sets used for *qpAdm*
852 analyses (see, for instance, Rasmussen et al. 2010, Haak et al. 2015, Harney et al. 2018,
853 Moreno-Mayar et al. 2018a, Zhang et al. 2021, Changmai et al. 2022a, Brielle et al. 2023),
854 while in our study identical population sets were used for *qpAdm*, PCA, and *ADMIXTURE*
855 analyses.

856 Another important caveat is that complexity of genetic history in the region and period of
857 interest often remains unknown and it is difficult to judge if a particular admixture graph
858 complexity is adequate for simulating the real history. However, we have not explored *qpAdm*
859 performance over a range of simulated admixture graph complexities, over a range of model
860 feasibility criteria (except for those in Table S2), for models more complex than two-way, and
861 have estimated FDR and FOR instead of false positive and false negative rates due to an
862 important technical limitation: the process of model classification into true and false ones
863 cannot be fully automated since it requires careful interpretation of the simulated topology
864 and simulated admixture proportions. For similar reasons, some comparisons of method
865 performance in this study, such as *qpAdm* vs. "*qpAdm* combined with *ADMIXTURE*", are
866 qualitative rather than quantitative: we applied the PCA and *ADMIXTURE* methods to one
867 simulation replicate only per simulation setup since automated classifiers of admixture
868 models into positive and negative ones based on 3D PCA and *ADMIXTURE* results were not
869 available. Despite these limitations, our simulations reproduce the most important aspects
870 of typical *qpAdm* protocols.

871 We demonstrated that application of the proximal rotating *qpAdm* protocol that can be
872 summarized as "whatever is not on the right is on the left" without any temporal stratification
873 of the "right" and "left" sets and of proxy sources and targets carries a risk of an FDR above
874 50% or 60%. Adding further levels of support (considering only models supported by PCA
875 and/or an *ADMIXTURE* analysis) does not help to decrease FDR drastically in this case (Table
876 S1, Suppl. text 1).

877 The proximal rotating protocol is an extreme example of *qpAdm* protocols that is rarely
878 encountered in the archaeogenetic literature (Calhoff et al. 2021, Oliveira et al. 2022) but

879 serves as a reference point in our analysis. Other protocols such as distal rotating (e.g.,
880 Narasimhan et al. 2019, Librado et al. 2021, Allentoft et al. 2022, Bergström et al. 2022,
881 Lazaridis et al. 2022, Taylor et al. 2023), distal non-rotating (e.g., Haak et al. 2015, Mathieson
882 et al. 2015, Lazaridis et al. 2016, Antonio et al. 2019, Marcus et al. 2020, Yang et al. 2020,
883 Papac et al. 2021, Yaka et al. 2021, Patterson et al. 2022), and proximal model competition
884 (e.g., Narasimhan et al. 2019, Calhoff et al. 2021, Zhang et al. 2021, Maróti et al. 2022, Brielle
885 et al. 2023, Lee et al. 2023) are often used in practice, and FDR of these three classes of
886 protocols on our simulated data ranged from 12% to 46% across simulation parameter
887 combinations and replicates (Tables 1 and 3). These FDR for best-performing standalone
888 *qpAdm* protocols are high but should not be over-interpreted since they are expected to
889 depend on the complexity of simulated histories and on the amount of data (Table 2), which
890 also depends on the time depth of simulated histories (Fig. S3). Only one graph complexity
891 level was tested, that is 13 groups and 10 admixture events; and only one time depth, 800
892 generations, was tested in a high-throughput way (Table 1). Thus, it is hard to extrapolate this
893 aspect of our results to real analyses and predict FDR levels on real data.

894 Temporal stratification tested in this study and practiced in the literature is of two sorts: 1)
895 most or all populations in the "right" set are sampled deeper in the past than those in the
896 "left" set (non-rotating protocols); 2) a target group post-dates (or is as old as) all its proxy
897 sources (distal protocols). We showed that both temporal stratification approaches helped
898 to decrease FDR of *qpAdm* admixture screens significantly, and the latter approach
899 demonstrated the best FDR among standalone *qpAdm* protocols (Table 2).

900 Although restricting analyses to distal models is often *necessary* for reducing FDR below an
901 arbitrary threshold at 10%, it is not *sufficient* for reaching this objective (Table 1) given the
902 complexity of our simulated admixture graph-shaped histories and the amounts of data we
903 generated. Respecting this threshold, only the following admixture screening protocols
904 demonstrated acceptable performance (we did not consider protocols demonstrating FOR
905 above 90% as useful in practice):

906 1) proximal non-rotating *qpAdm* with a requirement that admixture models are supported by
907 both *qpAdm* and an unsupervised *ADMIXTURE* analysis (Tables 1 and 3), under simulation

908 setups no. 1 and 2 (groups sampled at different dates in the past, maximal simulated history
909 depth = 800 generations, 300-Mbp-sized or 1000-Mbp-sized genomes simulated);

910 2) distal non-rotating or rotating *qpAdm* with a requirement that admixture models are
911 supported by both *qpAdm* and an unsupervised *ADMIXTURE* analysis (Table 1), under
912 simulation setups no. 1, 2, and 3 (groups sampled at different dates in the past, maximal
913 simulated history depth = 800 or 3000 generations, 300-Mbp-sized or 1000-Mbp-sized
914 genomes simulated).

915 FDR of these protocols was 0% – 8% (Table 1). In contrast, adding a model competition step
916 to the proximal non-rotating *qpAdm* protocol did not help to reduce FDR below 10%. The
917 performance of this type of protocols is explored in detail in Table 3. To sum up, we make the
918 following suggestions for improving robustness of admixture inference in archaeogenetics:

919 1. Our results suggest that temporal stratification of targets and proxy sources is a very
920 efficient way of reducing FDR of *qpAdm* protocols (Tables 1 and 2, Fig. 4). The distal
921 rotating and non-rotating protocols invariably demonstrated FDR significantly lower than
922 those of the proximal non-rotating and rotating protocols (Table 2). Although the
923 proximal model competition protocol (Narasimhan et al. 2019, Calhoff et al. 2021, Zhang
924 et al. 2021, Maróti et al. 2022, Brielle et al. 2023, Lee et al. 2023) was not tested on
925 multiple simulation or subsampling replicates (Table 3, we note that it demonstrated FDR
926 values higher than those of the distal non-rotating protocol (Table 1). Our results imply
927 that *qpAdm* protocols where all populations are sampled at present (similar to our setup
928 no. 4; see also Jeong et al. 2019, Changmai et al. 2022a) or where present-day groups are
929 used as proxy ancestry sources for ancient groups (e.g., Mathieson et al. 2015, van de
930 Loosdrecht et al. 2018, Narasimhan et al. 2019, Prendergast et al. 2019, Shinde et al. 2019,
931 Wang et al. 2020, Wang et al. 2021, Changmai et al. 2022b, Calhoff et al. 2021, Oliveira et
932 al. 2022) are less reliable than those where target groups are not allowed to pre-date
933 their proxy sources. While the proximal non-rotating *qpAdm* protocol demonstrated FDR
934 significantly lower than that of the proximal rotating protocol (Table 2), the distal rotating
935 protocol was in terms of FDR as good as the distal non-rotating protocol (on low-quality
936 data) or significantly better (on high-quality data, Table 2).

2. Another way of radically improving FDR of *qpAdm* protocols is combining *qpAdm* with an unsupervised *ADMIXTURE* analysis. These two approaches should probably be combined for optimal performance (Table 1).
3. Adding 3.3 times more data led to a small but significant decrease in FDR only in the case of high-quality diploid data, but not in the case of pseudo-haploid data with high missing rates (Table 2). This observation deserves further investigation.
4. It is safest to use the *qpAdm* method in controlled conditions, when relationships among populations are understood well enough to exclude violations of the topological assumptions, when radiocarbon or context dates of ancient populations are reliable and allow accurate temporal stratification, or when sets of potential proxy sources are well-constrained based on archaeological or historical scholarship: see, for instance, the earliest publications where the *qpAdm* method was employed (Haak et al. 2015, Mathieson et al. 2015) and some recent studies (e.g., Marcus et al. 2020, Papac et al. 2021, Yaka et al. 2021, Changmai et al. 2022a, 2022b, Patterson et al. 2022). Obviously, the amount of new information that the *qpAdm* method provides in these conditions is limited. However, considering that it was possible to reach FDR levels as low as 0% to 8% on our simulated data, we do not recommend avoiding *qpAdm*-based high-throughput admixture screens altogether.
5. Summing up all the results above, for reducing FDR of *qpAdm* admixture screens to nearly 0% we suggest using large SNP sets with low missing data rates, using the rotating *qpAdm* protocol with a strictly enforced rule that targets do not pre-date their proxy sources, and performing an unsupervised *ADMIXTURE* analysis to verify feasible *qpAdm* models.
6. Our study has multiple limitations and caveats discussed above, mostly related to difficulties in simulating all the details of published *qpAdm*, PCA, and *ADMIXTURE* protocols, to uncertainties about the level of admixture graph complexity that is adequate for simulating real population histories (all our simulations were of the same complexity: 13 groups and 10 pulse-like admixture events), to difficulties in interpreting topologies of random admixture graphs in an automated way for classifying even simple admixture models into true and false ones, and to difficulties in interpreting 3D PCA and *ADMIXTURE* results in an automated way. Nevertheless, our results surpass in scale previous simulation studies of *qpAdm* protocols (Lazaridis et al. 2017, Ning et al. 2020, Harney et

al. 2021) by several orders of magnitude and may serve as a guide for users of high-throughput *qpAdm* protocols.

7. Feasible *qpAdm* models are sometimes ranked by *p*-values, with a model having the highest *p*-value highlighted as the most plausible one (see, for instance, Lazaridis et al. 2022, van de Loosdrecht et al. 2018, Oliveira et al. 2022, Taylor et al. 2023). *qpWave p*-values for pairs of individuals were also used in lieu of genetic distances in the former study (Lazaridis et al. 2022). Of 1,201 instances when both false and true feasible *qpAdm* models were found for the same target group on the same data (all simulation setups, simulation/subsampling replicates, and *qpAdm* protocols), a model having the highest *p*-value was an FP in 463 (38.6%) cases, and the difference in maximal *p*-values between the TP and FP model classes was significant according to the paired two-sided Wilcoxon test (TP > FP, *p*-value = 2.2×10^{-16}). Thus, our limited analysis suggests that the approach of ranking *qpAdm* models by *p*-values is justified (see also related results in Fig. S6 and Table S2), but it generates noisy results.
8. f_3-statistic is a simple method for proving that a population is admixed, and it demonstrated FDR values much lower (6%, see Suppl. text 1) than those of standalone *qpAdm* protocols, but f_3-statistics are applicable only to recent admixture events and/or populations of large effective size since post-admixture drift on the target lineage obscures the signal. Moreover, calculating f_3-statistics for a target composed of a single pseudo-haploid individual is impossible since a heterozygosity estimate is required (Maier et al. 2022), and such singleton groups are common in archaeogenetic studies. Researchers should also be aware that f_3-statistics are defined on unrooted trees, and that may lead to rare but strong false signals of admixture (Fig. S4e).

Methods

Simulating random admixture graphs with msprime v.1.1.1

For simulating genetic data, we used *msprime v.1.1.1* which allows accurate simulation of recombination and of multi-chromosome diploid genomes relying on the Wright-Fisher model (Nelson et al. 2020, Baumdicker et al. 2022). We simulated three or ten diploid

chromosomes (each 100 Mbp long) by specifying a flat recombination rate (2×10^{-8} per nt per generation) along the chromosome and a much higher rate at the chromosome boundaries ($\log_e 2$ or ~0.693 per nt per generation, see https://tskit.dev/msprime/docs/stable/ancestry.html#multiple-chromosomes). A flat mutation rate, 1.25×10^{-8} per nt per generation (Scally & Durbin 2012), and the binary mutation model were used. To maintain the correct correlation between chromosomes, the discrete time Wright-Fischer model was used for 25 generations into the past, and deeper in the past the standard coalescent simulation algorithm was used (as recommended by Nelson et al. 2020).

Genetic histories in the form of random admixture graphs including 13 populations and 10 pulse-like admixture events were generated using the *random_admixturegraph* and *random_sim* functions from the *ADMIXTOOLS 2* package (https://uqrmaie1.github.io/admixtools/reference/random_sim.html), which produced scripts for running the *msprime v.1.1.1* simulator. Demographic events were separated by date intervals ranging randomly between 20 and 120 generations, with an upper bound on the graph depth at 800 generations (or ca. 23,000 years in the case of humans). In another set of simulations, all the dates were scaled up 3.75 times, with an upper bound on the graph depth at 3,000 generations (or 87,000 years in the case of humans). To be more precise, demographic events were not placed in time entirely randomly, but were tied to one or few other events of the same "topological depth" within the graph, as illustrated by five examples of the simulated topologies in Fig. S2. The same principle was applied to sampling dates, which were tied to other demographic events such as divergence and admixture of other populations. This was done to ensure topological consistency of random graphs.

Ten diploid individuals with no missing data were sampled from each population at "leaves" of the graph. Effective population sizes were constant along each edge and were picked randomly from the range of 1,000 – 10,000 diploid individuals. Admixture proportions for all admixture events varied randomly between 10% and 50%. This setup generates groups sampled at widely different dates in the past or, in other words, located at various genetic distances from the root. Alternatively, all terminal branches were extended to the "present" of the simulation and sampled at "present", keeping their respective effective population

sizes and topological relationships unchanged. Thus, another set of simulations was generated for the same topologies, where groups were more drifted with respect to each other (see F_{ST} distributions in Fig. S1).

In summary, four sets of independent simulations differing by the amount of data generated and by population divergence metrics were performed for a set of 40 random admixture graph topologies:

1) three 100-Mbp-sized chromosomes; groups sampled at different points in time; maximal simulated history depth at 800 generations (10 simulation replicates, median number of polymorphic sites = 669,655, see Fig. S3);

2) ten 100-Mbp-sized chromosomes; groups sampled at different points in time; maximal simulated history depth at 800 generations (10 simulation replicates, median number of polymorphic sites = 2,229,459);

3) three 100-Mbp-sized chromosomes; groups sampled at different points in time; maximal simulated history depth at 3,000 generations (one simulation replicate, median number of polymorphic sites = 1,074,336);

4) three 100-Mbp-sized chromosomes; all terminal branches extended to the "present" of the simulation and sampled at that point; maximal simulated history depth at 800 generations (one simulation replicate, median number of polymorphic sites = 838,297).

To create more realistic datasets, we performed randomised subsampling of polymorphic sites and individuals (replicates no. 1 of the first and second simulation setups were used for this, see the list above). First, we randomly sampled alleles at heterozygous sites, creating pseudo-haploid data. Then we introduced missing data by randomly selecting a missing rate between 5% and 95%, followed by randomly selecting sites according to the missing rate. This site subsampling was repeated for each individual independently. Lastly, we randomly sampled *n* (from 1 to 10) individuals from each population independently. The subsampling procedure described above was conditioned on the number of sites polymorphic in the set of 13 simulated populations and was repeated until a subsampling replicate with more than 20,000 (for 300-Mbp-sized genomes) or 66,000 such sites (for 1000-Mbp-sized genomes) was

obtained. We generated 10 independent subsampled replicates for each topology and simulation setup (800 replicates in total).

Polymorphism data in the *EIGENSTRAT* format were generated from the tree sequences using the *TreeSequence.genotype_matrix* function (https://tskit.dev/tskit/docs/stable/python-api.html#tskit.TreeSequence.genotype_matrix) and used for all subsequent analyses (*f*-statistics and *qpAdm*, PCA, *ADMIXTURE*).

For all the work on *f*-statistics and *qpAdm*, the *ADMIXTOOLS 2* software package (Maier et al. 2022) was used. For diploid SNP sets without missing data, we first calculated all possible f_2-statistics for 4-Mbp-sized genome blocks (with the "*maxmiss=0*", "*adjust_pseudohaploid=FALSE*", and "*minac2=FALSE*" settings) and then used them for calculating f_3- and f_4-statistics as linear combinations of f_2-statistics and for testing *qpAdm* models using the *qpadm* function in *ADMIXTOOLS 2* (https://uqrmaie1.github.io/admixtools/) under default settings. Inferred admixture proportions were not constrained between 0 and 1. For pseudo-haploid SNP sets with missing data and uneven group sizes, the *qpadm* function was applied directly to genotype files, with the "*allsnps=TRUE*" setting. In other words, f_4-statistics utilized by *qpAdm* and f_3-statistics were calculated off the genotype files without intermediate f_2-statistics, and removal of missing data was done for each population quadruplet or triplet separately. This setup is often used in the literature in the presence of missing data (e.g., Harney et al. 2018, Harney et al. 2019, Narasimhan et al. 2019, Lazaridis et al. 2022).

Rotating qpAdm protocols

QpWave tests were performed on sets of 13 groups divided randomly into 2 "left" and 11 "right" groups, testing all possible bisections of this form. *QpAdm* was applied to the same sets of 13 groups divided randomly into 3 "left" and 10 "right" groups, testing all possible bisections of this form for all possible target groups in "left" sets. This proximal rotating protocol was applied to all simulation setups. Subsequent work was focused only on feasible *qpAdm* models defined as follows: 1) *p*-values calculated by *qpWave* for one-way models "target = proxy source$_1$", "target = proxy source$_2$", and "proxy source$_1$ = proxy source$_2$" are all

below 0.01; 2) in the case of the two-way model "target = proxy source$_1$ + proxy source$_2$", estimated admixture proportions $\pm$ 2 standard errors are between 0 and 1; 3) the *p*-value calculated by *qpAdm* for the two-way model $\geq$ 0.01.

For exploring performance of the distal rotating protocol, feasible two-way *qpAdm* models were simply filtered according to sampling dates of target groups and proxy sources. If target group's sampling date is equal to or smaller than sampling dates of both proxy sources, such a model was considered distal.

Non-rotating and model competition qpAdm protocols

In the non-rotating protocol, for each simulated admixture graph six oldest groups were selected as a fixed "right" set (ties in sampling dates were resolved in alphabetical order; these "right" sets remained unchanged for a given topology across all independent simulations), and for the remaining seven groups all possible one-way and two-way admixture models were tested (105 models), applying the same composite feasibility criterion that was used above for the rotating protocol. This is the proximal non-rotating protocol, and alternatively we focused on distal admixture models only (distal non-rotating protocol).

In the proximal model competition protocol, subsequent analysis was focused on targets for whom two or more alternative *qpAdm* models emerged as feasible at the first step. For each target, alternative proxy sources were pooled and rotated between the "left" and "right" sets, testing only the models that emerged as feasible at the first step and applying the composite feasibility criterion (*p*-value $\geq$ 0.01, estimated admixture proportions ± 2 SE are between 0 and 1). Rotation of alternative proxy sources was performed in two alternative ways: "whatever is not on the left is on the right", or placement of alternative sources in the "right" set one by one. In the latter case several "right" sets were tested for each model, and the model was considered supported by the model competition protocol only if it was not rejected under any of these "right" sets (the latter protocol follows Maróti et al. 2022). If only one model was feasible for a target, such a model was evaluated as passing the model competition procedure. A distal model competition protocol was not tested in this study.

For testing statistical significance of differences in FDR between *qpAdm* protocols, the following approach was used. FDR was calculated either on low-quality data for 10 random site/individual subsampling replicates derived from simulation replicate no. 1 (simulation setups no. 1 and 2) or on high-quality data for 10 independent simulation replicates (simulation setups no. 1 and 2). Comparisons of four *qpAdm* protocols (rotating and non-rotating, proximal and distal) were performed independently on these four sets of replicates, using the two-sided paired Wilcoxon test (Table 2). Comparisons of the same *qpAdm* protocol on lower and higher amounts of data (300-Mbp- vs. 1,000-Mbp-sized simulated genomes) were performed using the two-sided (non-paired) Wilcoxon test since simulation replicates were independent unlike alternative *qpAdm* protocols applied to the same data (Table 2).

Classifying two-way admixture models into false and true positives

Since the simulated admixture graph topologies were complex and random, target groups modelled with *qpAdm* had very complex admixture history in some cases, being a part of gene flow networks. In this context it is hard to draw a strict boundary between true and false admixture models composed of a target and only two proxy sources. Two-way admixture models were considered false only if at least one of the following criteria was satisfied (considering only graph topologies and admixture proportions):

1. The target and at least one of the proxy sources are simulated as strictly cladal (Fig. 3). In this case the target may either be unadmixed, or it may have experienced gene flows earlier in its history that do not break its cladality with one of the proxy sources;
2. A proxy source does not represent any true source. In other words, it is symmetrically related to all true sources of ancestry in the target (Fig. S4a). Alternatively, both proxy sources represent the same true source, and are symmetrically related to all the other true sources (Fig. S4b).
3. A proxy source shares genetic drift with the corresponding true source that is not shared by the second proxy source (and the same is true for the other proxy source and another true source, i.e., condition no. 2 above is not satisfied), however less than 40% of its ancestry is derived from the true source (Fig. S4c);

4. A proxy source lineage is a recipient of gene flow from the target lineage (after the last admixture event in target's history), possibly mediated by other lineages (Fig. 3, Fig. S4d). In other words, the incorrect proxy source is a descendant of the target lineage, i.e., the expected gene flow direction is reversed.

We illustrate these topological rules with five examples of FP and feasible *qpAdm* models in Fig. 3 and Fig. S4a-e. Two-way models for targets whose population history is best approximated with three-way and more complex models were considered as true positives if they included source proxies (that do *not* satisfy the criteria above) for at least two of three or more true ancestry sources.

Principal component analysis

PCA was performed for one simulation replicate per simulation setup. Prior to the analysis, linked sites were pruned with *PLINK v.2.00a3LM* (Chang et al. 2015) using the following settings: window size, 2000 SNPs; window step, 100 SNPs; r^2 threshold = 0.5 (argument "--indep-pairwise 2000 100 0.5"). PCA was also performed using *PLINK v.2.00a3LM* under default settings, calculating 10 PCs. Interactive three-dimensional plots visualizing PC1, PC2, and PC3 were made using the *plotly* R package. A two-way admixture model was considered supported by PCA if:

1. the target group (the center of the cluster of target individuals, to be precise) lay between the clusters of proxy source individuals on a straight line in the three-dimensional PC space;
2. or if it was located at a distance of no more than three target cluster diameters from that straight line connecting the proxy source clusters.

The second pattern was more common among both TP and FP two-way admixture models: 1.5 and 1.3 times, respectively (across all non-subsampled simulated datasets). This situation is expected since many targets represent three-way and more complex mixtures, and since arrangement of populations in the PC space is influenced not only by admixture, but also by genetic drift.

ADMIXTURE analysis

ADMIXTURE analysis was performed for one simulation replicate per simulation setup. Prior to the analysis, linked sites were pruned with *PLINK v.2.00a3LM* (Chang et al. 2015) using the following settings: window size, 2000 SNPs; window step, 100 SNPs; r^2 threshold = 0.5 (argument "--indep-pairwise 2000 100 0.5"). *ADMIXTURE v.1.3* (Alexander et al. 2009) was used in the unsupervised mode under the default settings. The algorithm was run on each SNP dataset only once, with the number of hypothetical ancestral populations (*K*) ranging from 3 to 10. This range was selected since the total number of populations in each simulated history was 13. A two-way admixture model was considered supported by *ADMIXTURE* analysis if:

1. for at least one *K*, at least 5 of 10 target individuals were modelled as a mixture of at least two ancestry components, with a minor ancestry component exceeding 2%;
2. typically, ancestry component *A* in the target group was shared with at least 5 individuals in proxy source 1, but not in proxy source 2, and ancestry component *B* was shared with at least 5 individuals in proxy source 2, but not in proxy source 1 (see examples in Fig. 3 and Fig. S4); in some cases, both components *A* and *B* were found in the proxy sources, but in varying proportions;
3. if only one ancestry component in the target was shared with the two proxy sources, the model was considered unsupported;
4. ancestry components in the target that are absent in any of the sources were ignored since three-way and more complex admixture histories are common in the set of random admixture graphs explored here;
5. ancestry components in a proxy source that are absent in the target were also ignored since a proxy source may not be fully cladal with the real source.

These rules were designed to reproduce typical reasoning of an archaeogeneticist interpreting *ADMIXTURE* results. Observing a pattern of ancestry components in the target group and proxy sources compatible with the admixture model "target = proxy source$_1$ + proxy source$_2$" for one *K* value was enough for declaring that the model is supported by the *ADMIXTURE* analysis. This condition was motivated by an observation that models supported

at one *K* value only were equally common among FP and TP *qpAdm* models (10% and 13%, respectively, across four simulation setups). Models supported at four or more *K* values were more common among TP *qpAdm* models (3.3% of FP and 12.6% of TP models across four simulation setups).

Probability density curves for radiocarbon and calendar dates

Probability density curves for published radiocarbon and calendar dates were constructed in *OxCal v.4.4*. For calendar dates, we used the *C-Simulate* function in *OxCal v.4.4* for simulating normally distributed dating methods, taking the average calendar date as a median and the length of the timespan as a 95% confidence interval. For radiocarbon dates, we used calibration based on the IntCal20 calibration curve. Probability densities were summarized using the *Sum* function in *OxCal v.4.4* for each of the three groups of individuals, those included in the "left", "right", and "target" population sets in at least one of the published *qpAdm* models (Narasimhan et al. 2019, Lazaridis et al. 2022), and then plotted together.

Acknowledgements

P.F. and P.C. were supported by the Czech Science Foundation (project no. 21-27624S). P.F. was also supported by the John Templeton Foundation (grant no. 61220 to David Reich) and by a gift from Jean-Francois Clin. U.I. and L.A.V. were supported by the Czech Ministry of Education, Youth and Sports (program ERC CZ, project no. LL2103). Computational resources for this work were supplied by the projects "e-INFRA CZ" (no. 90140) and “IT4Innovations National Supercomputing Center” (no. LM2015070) supported by the Ministry of Education, Youth and Sports of the Czech Republic, and by the ELIXIR-CZ project (no. LM2022123), a part of the international ELIXIR infrastructure.

References

1. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated

individuals, *Genome Res*. **19** (2009) 1655–1664. doi: 10.1101/gr.094052.109.

2. Allentoft ME, Sikora M, Refoyo-Martínez A, *et al*. Population Genomics of Stone Age Eurasia. *bioRxiv*. (2022). doi: 10.1101/2022.05.04.490594.
3. Antonio ML, Gao Z, Moots HM, Lucci M, Candilio F, Sawyer S, Oberreiter V, Calderon D, Devitofranceschi K, Aikens RC, Aneli S, Bartoli F, Bedini A, Cheronet O, Cotter DJ, Fernandes DM, Gasperetti G, Grifoni R, Guidi A, La Pastina F, Loreti E, Manacorda D, Matullo G, Morretta S, Nava A, Fiocchi Nicolai V, Nomi F, Pavolini C, Pentiricci M, Pergola P, Piranomonte M, Schmidt R, Spinola G, Sperduti A, Rubini M, Bondioli L, Coppa A, Pinhasi R, Pritchard JK. Ancient Rome: A genetic crossroads of Europe and the Mediterranean. *Science*. **366** (2019) 708–714. doi: 10.1126/science.aay6826.
4. Baumdicker F, Bisschop G, Goldstein D, Gower G, Ragsdale AP, Tsambos G, Zhu S, Eldon B, Ellerman EC, Galloway JG, Gladstein AL, Gorjanc G, Guo B, Jeffery B, Kretzschumar WW, Lohse K, Matschiner M, Nelson D, Pope NS, Quinto-Cortés CD, Rodrigues MF, Saunack K, Sellinger T, Thornton K, van Kemenade H, Wohns AW, Wong Y, Gravel S, Kern AD, Koskela J, Ralph PL, Kelleher J. Efficient ancestry and mutation simulation with msprime 1.0, *Genetics*. **220** (2022) iyab229. doi: 10.1093/genetics/iyab229.
5. Bergström A, Stanton DWG, Taron UH, Frantz L, Sinding MS, Ersmark E, Pfrengle S, Cassatt-Johnstone M, Lebrasseur O, Girdland-Flink L, Fernandes DM, Ollivier M, Speidel L, Gopalakrishnan S, Westbury MV, Ramos-Madrigal J, Feuerborn TR, Reiter E, Gretzinger J, Münzel SC, Swali P, Conard NJ, Carøe C, Haile J, Linderholm A, Androsov S, Barnes I, Baumann C, Benecke N, Bocherens H, Brace S, Carden RF, Drucker DG, Fedorov S, Gasparik M, Germonpré M, Grigoriev S, Groves P, Hertwig ST, Ivanova VV, Janssens L, Jennings RP, Kasparov AK, Kirillova IV, Kurmaniyazov I, Kuzmin YV, Kosintsev PA, Lázničková-Galetová M, Leduc C, Nikolskiy P, Nussbaumer M, O'Drisceoil C, Orlando L, Outram A, Pavlova EY, Perri AR, Pilot M, Pitulko VV, Plotnikov VV, Protopopov AV, Rehazek A, Sablin M, Seguin-Orlando A, Storå J, Verjux C, Zaibert VF, Zazula G, Crombé P, Hansen AJ, Willerslev E, Leonard JA, Götherström A, Pinhasi R, Schuenemann VJ, Hofreiter M, Gilbert MTP, Shapiro B, Larson G, Krause J, Dalén L, Skoglund P. Grey wolf genomic history reveals a dual ancestry of dogs, *Nature*. **607** (2022) 313–320. doi: 10.1038/s41586-022-04824-9.
6. Brielle ES, Fleisher J, Wynne-Jones S, Sirak K, Broomandkhoshbacht N, Callan K, Curtis E, Iliev L, Lawson AM, Oppenheimer J, Qiu L, Stewardson K, Workman JN, Zalzala F, Ayodo G, Gidna AO, Kabiru A, Kwekason A, Mabulla AZP, Manthi FK, Ndiema E, Ogola C, Sawchuk E, Al-Gazali L, Ali BR, Ben-Salem S, Letellier T, Pierron D, Radimilahy C, Rakotoarisoa JA, Raaum RL, Culleton BJ, Mallick S, Rohland N, Patterson N, Mwenje MA, Ahmed KB, Mohamed MM, Williams SR, Monge

J, Kusimba S, Prendergast ME, Reich D, Kusimba CM. Entwined African and Asian genetic roots of medieval peoples of the Swahili coast. *Nature*. **615** (2023) 866–873. doi: 10.1038/s41586-023-05754-w.

7. Carlhoff S, Duli A, Nägele K, Nur M, Skov L, Sumantri I, Oktaviana AA, Hakim B, Burhan B, Syahdar FA, McGahan DP, Bulbeck D, Perston YL, Newman K, Saiful AM, Ririmasse M, Chia S, Hasanuddin, Pulubuhu DAT, Suryatman, Supriadi, Jeong C, Peter BM, Prüfer K, Powell A, Krause J, Posth C, Brumm A. Genome of a middle Holocene hunter-gatherer from Wallacea. *Nature*. **596** (2021) 543–547. doi: 10.1038/s41586-021-03823-6.
8. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets, *Gigascience*. **4** (2015) 7. doi: 10.1186/s13742-015-0047-8.
9. Changmai P, Jaisamut K, Kampuansai J, Kutanan W, Altınışık NE, Flegontova O, Inta A, Yüncü E, Boonthai W, Pamjav H, Reich D, Flegontov P. Indian genetic heritage in Southeast Asian populations, *PLoS Genet*. **18** (2022a) e1010036. doi: 10.1371/journal.pgen.1010036.
10. Changmai P, Pinhasi R, Pietrusewsky M, Stark MT, Ikehara-Quebral RM, Reich D, Flegontov P. Ancient DNA from Protohistoric Period Cambodia indicates that South Asians admixed with local populations as early as 1st-3rd centuries CE, *Sci. Rep*. **12** (2022b) 22507. doi: 10.1038/s41598-022-26799-3.
11. Chintalapati M, Patterson N, Moorjani P. The spatiotemporal patterns of major human admixture events during the European Holocene, *eLife*. **11** (2022) e77625. doi: 10.7554/eLife.77625.
12. de Barros Damgaard P, Martiniano R, Kamm J, Moreno-Mayar JV, Kroonen G, Peyrot M, Barjamovic G, Rasmussen S, Zacho C, Baimukhanov N, Zaibert V, Merz V, Biddanda A, Merz I, Loman V, Evdokimov V, Usmanova E, Hemphill B, Seguin-Orlando A, Yediay FE, Ullah I, Sjögren KG, Iversen KH, Choin J, de la Fuente C, Ilardo M, Schroeder H, Moiseyev V, Gromov A, Polyakov A, Omura S, Senyurt SY, Ahmad H, McKenzie C, Margaryan A, Hameed A, Samad A, Gul N, Khokhar MH, Goriunova OI, Bazaliiskii VI, Novembre J, Weber AW, Orlando L, Allentoft ME, Nielsen R, Kristiansen K, Sikora M, Outram AK, Durbin R, Willerslev E. The first horse herders and the impact of early Bronze Age steppe expansions into Asia, *Science*. **360** (2018) eaar7711. doi: 10.1126/science.aar7711.
13. Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations, *Mol. Biol. Evol*. **28** (2011) 2239–2252. doi: 10.1093/molbev/msr048.
14. Fernandes DM, Mittnik A, Olalde I, Lazaridis I, Cheronet O, Rohland N, Mallick S, Bernardos R, Broomandkhoshbacht N, Carlsson J, Culleton BJ, Ferry M, Gamarra B, Lari M, Mah M, Michel M,

Modi A, Novak M, Oppenheimer J, Sirak KA, Stewardson K, Mandl K, Schattke C, Özdoğan KT, 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 RM, Palombo MR, Ramis D, Schimmenti V, Sureda P, Teira L, Teschler-Nicola M, Kennett DJ, Lalueza-Fox C, Patterson N, Sineo L, Coppa A, Caramelli D, Pinhasi R, Reich D. The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean, *Nat. Ecol. Evol*. **4** (2020) 334–345. doi: 10.1038/s41559-020-1102-0.

15. Fernandes DM, Sirak KA, Ringbauer H, Sedig J, Rohland N, Cheronet O, Mah M, Mallick S, Olalde I, Culleton BJ, Adamski N, Bernardos R, Bravo G, Broomandkhoshbacht N, Callan K, Candilio F, Demetz L, Carlson KSD, Eccles L, Freilich S, George RJ, Lawson AM, Mandl K, Marzaioli F, McCool WC, Oppenheimer J, Özdogan KT, Schattke C, Schmidt R, Stewardson K, Terrasi F, Zalzala F, Antúnez CA, Canosa EV, Colten R, Cucina A, Genchi F, Kraan C, La Pastina F, Lucci M, Maggiolo MV, Marcheco-Teruel B, Maria CT, Martínez C, París I, Pateman M, Simms TM, Sivoli CG, Vilar M, Kennett DJ, Keegan WF, Coppa A, Lipson M, Pinhasi R, Reich D. A genetic history of the pre-contact Caribbean, *Nature*. **590** (2021) 103–110. doi: 10.1038/s41586-020-03053-2.
16. Frichot E, Schoville S, Bouchard G, François O. Correcting principal component maps for effects of spatial autocorrelation in population genetic data, *Front. Genet*. **3** (2012) 254. doi: 10.3389/fgene.2012.00254.
17. Furtwängler A, Rohrlach AB, Lamnidis TC, Papac L, Neumann GU, Siebke I, Reiter E, Steuri N, Hald J, Denaire A, Schnitzler B, Wahl J, Ramstein M, Schuenemann VJ, Stockhammer PW, Hafner A, Lösch S, Haak W, Schiffels S, Krause J. Ancient genomes reveal social and genetic structure of Late Neolithic Switzerland. *Nat. Commun*. **11** (2020) 1915. doi: 10.1038/s41467-020-15560-x.
18. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Ž, Gušic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PLF, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S. A draft sequence of the Neandertal genome, *Science*. **328** (2010) 710–722. doi: 10.1126/science.1188021.
19. Gnecchi-Ruscone GA, Szécsényi-Nagy A, Koncz I, Csiky G, Rácz Z, Rohrlach AB, Brandt G, Rohland N, Csáky V, Cheronet O, Szeifert B, Rácz TÁ, Benedek A, Bernert Z, Berta N, Czifra S, Dani J, Farkas Z, Hága T, Hajdu T, Jászberényi M, Kisjuhász V, Kolozsi B, Major P, Marcsik A, Kovacsóczy BN, Balogh C, Lezsák GM, Ódor JG, Szelekovszky M, Szeniczey T, Tárnoki J, Tóth Z, Tutkovics EK, Mende BG, Geary P, Pohl W, Vida T, Pinhasi R, Reich D, Hofmanová Z, Jeong C, Krause J. Ancient

genomes reveal origin and rapid trans-Eurasian migration of 7th century Avar elites, *Cell*. **185** (2022) 1402–1413.e21. doi: 10.1016/j.cell.2022.03.007.

20. 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 RG, Hallgren F, Khartanovich V, Khokhlov A, Kunst M, Kuznetsov P, Meller H, Mochalov O, Moiseyev V, Nicklisch N, Pichler SL, Risch R, Rojo Guerra MA, Roth C, Szécsényi-Nagy A, Wahl J, Meyer M, Krause J, Brown D, Anthony D, Cooper A, Alt KW, Reich D. Massive migration from the steppe was a source for Indo-European languages in Europe, *Nature*. **522** (2015) 207–211. doi: 10.1038/nature14317.
21. Harney É, May H, Shalem D, Rohland N, Mallick S, Lazaridis I, Sarig R, Stewardson K, Nordenfelt S, Patterson N, Hershkovitz I, Reich D. Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation, *Nat. Commun*. **9** (2018) 3336. doi: 10.1038/s41467-018-05649-9.
22. Harney É, Nayak A, Patterson N, Joglekar P, Mushrif-Tripathy V, Mallick S, Rohland N, Sedig J, Adamski N, Bernardos R, Broomandkhoshbacht N, Culleton BJ, Ferry M, Harper TK, Michel M, Oppenheimer J, Stewardson K, Zhang Z, Harashawaradhana, Bartwal MS, Kumar S, Diyundi SC, Roberts P, Boivin N, Kennett DJ, Thangaraj K, Reich D, Rai N. Ancient DNA from the skeletons of Roopkund Lake reveals Mediterranean migrants in India, *Nat. Commun*. **10** (2019) 3670. doi: 10.1038/s41467-019-11357-9.
23. Harney É, Patterson N, Reich D, Wakeley J. Assessing the performance of qpAdm: a statistical tool for studying population admixture, *Genetics*. **217** (2021) iyaa045. doi: 10.1093/GENETICS/IYAA045.
24. Jeong C, Balanovsky O, Lukianova E, Kahbatkyzy N, Flegontov P, Zaporozhchenko V, Immel A, Wang CC, Ixan O, Khussainova E, Bekmanov B, Zaibert V, Lavryashina M, Pocheshkhova E, Yusupov Y, Agdzhoyan A, Koshel S, Bukin A, Nymadawa P, Turdikulova S, Dalimova D, Churnosov M, Skhalyakho R, Daragan D, Bogunov Y, Bogunova A, Shtrunov A, Dubova N, Zhabagin M, Yepiskoposyan L, Churakov V, Pislegin N, Damba L, Saroyants L, Dibirova K, Atramentova L, Utevska O, Idrisov E, Kamenshchikova E, Evseeva I, Metspalu M, Outram AK, Robbeets M, Djansugurova L, Balanovska E, Schiffels S, Haak W, Reich D, Krause J. The genetic history of admixture across inner Eurasia, *Nat. Ecol. Evol*. **3** (2019) 966–976. doi: 10.1038/s41559-019-0878-2.
25. Lawson DJ, van Dorp L, Falush D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots, *Nat Commun*. **9** (2018) 3258. doi: 10.1038/s41467-018-05257-7.
26. Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, Fernandes D, Novak M,

Gamarra B, Sirak K, Connell S, Stewardson K, Harney E, Fu Q, Gonzalez-Fortes G, Jones ER, Roodenberg SA, Lengyel G, Bocquentin F, Gasparian B, Monge JM, Gregg M, Eshed V, Mizrahi AS, Meiklejohn C, Gerritsen F, Bejenaru L, Blüher M, Campbell A, Cavalleri G, Comas D, Froguel P, Gilbert E, Kerr SM, Kovacs P, Krause J, McGettigan D, Merrigan M, Merriwether DA, O'Reilly S, Richards MB, Semino O, Shamoon-Pour M, Stefanescu G, Stumvoll M, Tönjes A, Torroni A, Wilson JF, Yengo L, Hovhannisyan NA, Patterson N, Pinhasi R, Reich D. Genomic insights into the origin of farming in the ancient Near East, *Nature*. **536** (2016) 419–424. doi: 10.1038/nature19310.

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

28. 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 Aİ, 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, Bravo Morante G, Buttinger K, Callan K, Candilio F, Carić M, Cheronet O, Chohadzhiev S, Chovalopoulou ME, Chryssoulaki S, Ciobanu I, Čondić N, Constantinescu M, Cristiani E, Culleton BJ, Curtis E, Davis J, Demcenco TI, Dergachev V, Derin Z, Deskaj S, Devejyan S, Djordjević V, Duffett Carlson KS, Eccles LR, Elenski N, Engin A, Erdoğan N, Erir-Pazarcı S, Fernandes DM, Ferry M, Freilich S, Frînculeasa A, Galaty ML, 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 EH, Keating D, Kennett DJ, Deniz Kesici S, Khudaverdyan A, Kiss K, Kılıç S, Klostermann P, Kostak Boca Negra Valdes S, Kovačević S, Krenz-Niedbała M, Krznarić Škrivanko M, Kurti R, Kuzman P, Lawson AM, Lazar C, Leshtakov K, Levy TE, Liritzis I, Lorentz KO, 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 JM, Nekhrizov G, Nicholls R, Nikitin AG, Nikolov V, Novak M, Olalde I, Oppenheimer J, Osterholtz A, Özdemir C, Özdoğan KT, Öztürk N, Papadimitriou N, Papakonstantinou N, Papathanasiou A, Paraman L, Paskary EG, Patterson N, Petrakiev I, Petrosyan L, Petrova V, Philippa-Touchais A, Piliposyan A, Pocuca Kuzman N, Potrebica H, Preda-Bălănică B, Premužić Z, Price TD, Qiu L, Radović S, Raeuf Aziz K,

Rajić Šikanjić P, Rasheed Raheem K, 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 HM, 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 JN, Yardumian A, Yarovoy E, Yavuz AY, Yılmaz H, Zalzala F, Zettl A, Zhang Z, Çavuşoğlu R, Rohland N, Pinhasi R, Reich D, Davtyan R. The genetic history of the Southern Arc: A bridge between West Asia and Europe, *Science*. **377** (2022) eabm4247. doi: 10.1126/science.abm4247.

29. Lee J, Miller BK, Bayarsaikhan J, Johannesson E, Ventresca Miller A, Warinner C, Jeong C. Genetic population structure of the Xiongnu Empire at imperial and local scales. *Sci. Adv*. **9** (2023) eadf3904. doi: 10.1126/sciadv.adf3904.
30. Librado P, Khan N, Fages A, Kusliy MA, Suchan T, Tonasso-Calvière L, Schiavinato S, Alioglu D, Fromentier A, Perdereau A, Aury JM, Gaunitz C, Chauvey L, Seguin-Orlando A, Der Sarkissian C, Southon J, Shapiro B, Tishkin AA, Kovalev AA, Alquraishi S, Alfarhan AH, Al-Rasheid KAS, Seregély T, Klassen L, Iversen R, Bignon-Lau O, Bodu P, Olive M, Castel JC, Boudadi-Maligne M, Alvarez N, Germonpré M, Moskal-Del Hoyo M, Wilczyński J, Pospuła S, Lasota-Kuś A, Tunia K, Nowak M, Rannamäe E, Saarma U, Boeskorov G, Lõugas L, Kyselý R, Peške L, Bălăşescu A, Dumitraşcu V, Dobrescu R, Gerber D, Kiss V, Szécsényi-Nagy A, Mende BG, Gallina Z, Somogyi K, Kulcsár G, Gál E, Bendrey R, Allentoft ME, Sirbu G, Dergachev V, Shephard H, Tomadini N, Grouard S, Kasparov A, Basilyan AE, Anisimov MA, Nikolskiy PA, Pavlova EY, Pitulko V, Brem G, Wallner B, Schwall C, Keller M, Kitagawa K, Bessudnov AN, Bessudnov A, Taylor W, Magail J, Gantulga JO, Bayarsaikhan J, Erdenebaatar D, Tabaldiev K, Mijiddorj E, Boldgiv B, Tsagaan T, Pruvost M, Olsen S, Makarewicz CA, Valenzuela Lamas S, Albizuri Canadell S, Nieto Espinet A, Iborra MP, Lira Garrido J, Rodríguez González E, Celestino S, Olària C, Arsuaga JL, Kotova N, Pryor A, Crabtree P, Zhumatayev R, Toleubaev A, Morgunova NL, Kuznetsova T, Lordkipanize D, Marzullo M, Prato O, Bagnasco Gianni G, Tecchiati U, Clavel B, Lepetz S, Davoudi H, Mashkour M, Berezina NY, Stockhammer PW, Krause J, Haak W, Morales-Muñiz A, Benecke N, Hofreiter M, Ludwig A, Graphodatsky AS, Peters J, Kiryushin KY, Iderkhangai TO, Bokovenko NA, Vasiliev SK, Seregin NN, Chugunov KV, Plasteeva NA, Baryshnikov GF, Petrova E, Sablin M, Ananyevskaya E, Logvin A, Shevnina I, Logvin V, Kalieva S, Loman V, Kukushkin I, Merz I, Merz V, Sakenov S, Varfolomeyev V, Usmanova E, Zaibert V, Arbuckle B, Belinskiy AB, Kalmykov A, Reinhold S, Hansen S, Yudin AI, Vybornov AA, Epimakhov A, Berezina NS, Roslyakova N, Kosintsev PA, Kuznetsov PF, Anthony D,

Kroonen GJ, Kristiansen K, Wincker P, Outram A, Orlando L. The origins and spread of domestic horses from the Western Eurasian steppes, *Nature*. **598** (2021) 634–640. doi: 10.1038/s41586-021-04018-9.

31. Lipson M. Applying f_4-statistics and admixture graphs: Theory and examples, *Mol. Ecol. Resour*. **20** (2020) 1658–1667. doi: 10.1111/1755-0998.13230.
32. Loh PR, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, Berger B. Inferring admixture histories of human populations using linkage disequilibrium, *Genetics*. **193** (2013) 1233–1254. doi: 10.1534/genetics.112.147330.
33. Maier R, Flegontov P, Flegontova O, Işıldak U, Changmai P, Reich D. On the limits of fitting complex models of population history to *f*-statistics. *eLife*. **12** (2023) e85492. doi: 10.7554/eLife.85492.
34. Maróti Z, Neparáczki E, Schütz O, Maár K, Varga GIB, Kovács B, Kalmár T, Nyerki E, Nagy I, Latinovics D, Tihanyi B, Marcsik A, Pálfi G, Bernert Z, Gallina Z, Horváth C, Varga S, Költő L, Raskó I, Nagy PL, Balogh C, Zink A, Maixner F, Götherström A, George R, Szalontai C, Szenthe G, Gáll E, Kiss AP, Gulyás B, Kovacsóczy BN, Gál SS, Tomka P, Török T. The genetic origin of Huns, Avars, and conquering Hungarians, *Curr. Biol*. **32** (2022) 2858–2870.e7. doi: 10.1016/j.cub.2022.04.093.
35. Martin SH, Davey JW, Jiggins CD. Evaluating the use of ABBA-BABA statistics to locate introgressed loci, *Mol. Biol. Evol*. **32** (2014) 244–257. doi: 10.1093/molbev/msu269.
36. Marcus JH, Posth C, Ringbauer H, Lai L, Skeates R, Sidore C, Beckett J, Furtwängler A, Olivieri A, Chiang CWK, Al-Asadi H, Dey K, Joseph TA, Liu CC, Der Sarkissian C, Radzevičiūtė R, Michel M, Gradoli MG, Marongiu P, Rubino S, Mazzarello V, Rovina D, La Fragola A, Serra RM, Bandiera P, Bianucci R, Pompianu E, Murgia C, Guirguis M, Orquin RP, Tuross N, van Dommelen P, Haak W, Reich D, Schlessinger D, Cucca F, Krause J, Novembre J. Genetic history from the Middle Neolithic to present on the Mediterranean island of Sardinia, *Nat. Commun*. **11** (2020) 939. doi: 10.1038/s41467-020-14523-6.
37. Maróti Z, Neparáczki E, Schütz O, Maár K, Varga GIB, Kovács B, Kalmár T, Nyerki E, Nagy I, Latinovics D, Tihanyi B, Marcsik A, Pálfi G, Bernert Z, Gallina Z, Horváth C, Varga S, Költő L, Raskó I, Nagy PL, Balogh C, Zink A, Maixner F, Götherström A, George R, Szalontai C, Szenthe G, Gáll E, Kiss AP, Gulyás B, Kovacsóczy BN, Gál SS, Tomka P, Török T. The genetic origin of Huns, Avars, and conquering Hungarians, *Curr. Biol*. **32** (2022) 2858–2870.e7. doi: 10.1016/j.cub.2022.04.093.
38. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, Sirak K, Gamba C, Jones ER, Llamas B, Dryomov S, Pickrell J, Arsuaga JL, de Castro JM, Carbonell E, Gerritsen F, Khokhlov A, Kuznetsov P, Lozano M, Meller H, Mochalov O, Moiseyev V, Guerra MA, Roodenberg J, Vergès JM, Krause J, Cooper A, Alt KW,

Brown D, Anthony D, Lalueza-Fox C, Haak W, Pinhasi R, Reich D. Genome-wide patterns of selection in 230 ancient Eurasians, *Nature*. **528** (2015) 499–503. doi: 10.1038/nature16152. Epub 2015 Nov 23. PMID: 26595274; PMCID: PMC4918750.

39. 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 GG, 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 KW, Andreescu R, Antonović D, Ash A, Atanassova N, Bacvarov K, Gusztáv MB, Bocherens H, Bolus M, Boroneanţ A, Boyadzhiev Y, Budnik A, Burmaz J, Chohadzhiev S, Conard NJ, Cottiaux R, Čuka M, Cupillard C, Drucker DG, 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 DJ, Komšo D, Kozak A, Labuda D, Lari M, Lazar C, Leppek M, Leshtakov K, Vetro DL, Los D, Lozanov I, Malina M, Martini F, McSweeney K, Meller H, Menđušić M, Mirea P, Moiseyev V, Petrova V, Price TD, 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 PW, Catalano G, Krauß R, Caramelli D, Zariņa G, Gaydarska B, Lillie M, Nikitin AG, Potekhina I, Papathanasiou A, Borić D, Bonsall C, Krause J, Pinhasi R, Reich D. The genomic history of southeastern Europe, *Nature*. **555** (2018) 197–203. doi: 10.1038/nature25778.

40. McVean G. A genealogical interpretation of principal components analysis, *PLoS Genet*. **5** (2009) e1000686. doi: 10.1371/journal.pgen.1000686.

41. Moreno-Mayar JV, Potter BA, Vinner L, Steinrücken M, Rasmussen S, Terhorst J, Kamm JA, Albrechtsen A, Malaspinas AS, Sikora M, Reuther JD, Irish JD, Malhi RS, Orlando L, Song YS, Nielsen R, Meltzer DJ, Willerslev E. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature*. **553** (2018a) 203–207. doi: 10.1038/nature25173.

42. Moreno-Mayar JV, Vinner L, de Barros Damgaard P, de la Fuente C, Chan J, Spence JP, Allentoft ME, Vimala T, Racimo F, Pinotti T, Rasmussen S, Margaryan A, Iraeta Orbegozo M, Mylopotamitaki D, Wooller M, Bataille C, Becerra-Valdivia L, Chivall D, Comeskey D, Devièse T, Grayson DK, George L, Harry H, Alexandersen V, Primeau C, Erlandson J, Rodrigues-Carvalho C, Reis S, Bastos MQR, Cybulski J, Vullo C, Morello F, Vilar M, Wells S, Gregersen K, Hansen KL, Lynnerup N, Mirazón Lahr M, Kjær K, Strauss A, Alfonso-Durruty M, Salas A, Schroeder H, Higham T, Malhi RS, Rasic JT, Souza L, Santos FR, Malaspinas AS, Sikora M, Nielsen R, Song YS, Meltzer DJ, Willerslev E. Early human dispersals within the Americas, *Science*. **362** (2018b) eaav2621. doi: 10.1126/science.aav2621.

43. Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, Lazaridis I, Nakatsuka N, Olalde I, Lipson M, Kim AM, Olivieri LM, 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 BJ, Ferry M, Fernandes D, Freilich S, Gamarra B, Gaudio D, Hajdinjak M, Harney É, Harper TK, Keating D, Lawson AM, 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 AN, Bagnera A, Baitanayev B, Bendezu-Sarmiento J, Bissembaev AA, Bonora GL, Chargynov TT, Chikisheva T, Dashkovskiy PK, Derevianko A, Dobeš M, Douka K, Dubova N, Duisengali MN, Enshin D, Epimakhov A, Fribus AV, Fuller D, Goryachev A, Gromov A, Grushin SP, Hanks B, Judd M, Kazizov E, Khokhlov A, Krygin AP, Kupriyanova E, Kuznetsov P, Luiselli D, Maksudov F, Mamedov AM, Mamirov TB, Meiklejohn C, Merrett DC, Micheli R, Mochalov O, Mustafokulov S, Nayak A, Pettener D, Potts R, Razhev D, Rykun M, Sarno S, Savenkova TM, Sikhymbaeva K, Slepchenko SM, Soltobaev OA, Stepanova N, Svyatko S, Tabaldiev K, Teschler-Nicola M, Tishkin AA, Tkachev VV, Vasilyev S, Velemínský P, Voyakin D, Yermolayeva A, Zahir M, Zubkov VS, Zubova A, Shinde VS, Lalueza-Fox C, Meyer M, Anthony D, Boivin N, Thangaraj K, Kennett DJ, Frachetti M, Pinhasi R, Reich D. The formation of human populations in South and Central Asia, *Science*. **365** (2019) eaat7487. doi: 10.1126/science.aat7487.

44. Nelson D, Kelleher J, Ragsdale AP, Moreau C, McVean G, Gravel S. Accounting for long-range correlations in genome-wide simulations of large cohorts, *PLoS Genet*. **16** (2020) e1008619. doi: 10.1371/journal.pgen.1008619.

45. Ning C, Li T, Wang K, Zhang F, Li T, Wu X, Gao S, Zhang Q, Zhang H, Hudson MJ, Dong G, Wu S, Fang Y, Liu C, Feng C, Li W, Han T, Li R, Wei J, Zhu Y, Zhou Y, Wang CC, Fan S, Xiong Z, Sun Z, Ye M, Sun L, Wu X, Liang F, Cao Y, Wei X, Zhu H, Zhou H, Krause J, Robbeets M, Jeong C, Cui Y. Ancient genomes from northern China suggest links between subsistence changes and human migration, *Nat. Commun*. **11** (2020) 2700. doi: 10.1038/s41467-020-16557-2.

46. Ning C, Fernandes D, Changmai P, Flegontova O, Yüncü E, Maier R, Altınışık NE, Kassian AS, Krause J., Lalueza-Fox C, Manica A, Potter BA, Robbeets M, Sirak K, Siska V, Vajda EJ, Vyazov LA, Wang K, Wang L, Wu X, Xiao X, Zhang F, Reich D, Schiffels S, Pinhasi R, Cui Y, Flegontov P. The genomic formation of First American ancestors in East and Northeast Asia, *bioRxiv*. (2020). doi: 10.1101/2020.10.12.336628.

47. Novembre J, Stephens M. Interpreting principal component analyses of spatial population genetic variation, *Nat. Genet*. **40** (2008) 646–649. doi: 10.1038/ng.139.

48. Olalde I, Mallick S, Patterson N, Rohland N, Villalba-Mouco V, Silva M, Dulias K, Edwards CJ, Gandini F, Pala M, Soares P, Ferrando-Bernal M, Adamski N, Broomandkhoshbacht N, Cheronet

O, Culleton BJ, Fernandes D, Lawson AM, Mah M, Oppenheimer J, Stewardson K, Zhang Z, Jiménez Arenas JM, Toro Moyano IJ, Salazar-García DC, Castanyer P, Santos M, Tremoleda J, Lozano M, García Borja P, Fernández-Eraso J, Mujika-Alustiza JA, Barroso C, Bermúdez FJ, Viguera Mínguez E, Burch J, Coromina N, Vivó D, Cebrià A, Fullola JM, García-Puchol O, Morales JI, Oms FX, Majó T, Vergès JM, Díaz-Carvajal A, Ollich-Castanyer I, López-Cachero FJ, Silva AM, Alonso-Fernández C, Delibes de Castro G, Jiménez Echevarría J, Moreno-Márquez A, Pascual Berlanga G, Ramos-García P, Ramos-Muñoz J, Vijande Vila E, Aguilella Arzo G, Esparza Arroyo Á, Lillios KT, Mack J, Velasco-Vázquez J, Waterman A, Benítez de Lugo Enrich L, Benito Sánchez M, Agustí B, Codina F, de Prado G, Estalrrich A, Fernández Flores Á, Finlayson C, Finlayson G, Finlayson S, Giles-Guzmán F, Rosas A, Barciela González V, García Atiénzar G, Hernández Pérez MS, Llanos A, Carrión Marco Y, Collado Beneyto I, López-Serrano D, Sanz Tormo M, Valera AC, Blasco C, Liesau C, Ríos P, Daura J, de Pedro Michó MJ, Diez-Castillo AA, Flores Fernández R, Francès Farré J, Garrido-Pena R, Gonçalves VS, Guerra-Doce E, Herrero-Corral AM, Juan-Cabanilles J, López-Reyes D, McClure SB, Merino Pérez M, Oliver Foix A, Sanz Borràs M, Sousa AC, Vidal Encinas JM, Kennett DJ, Richards MB, Werner Alt K, Haak W, Pinhasi R, Lalueza-Fox C, Reich D. The genomic history of the Iberian Peninsula over the past 8000 years, *Science*. **363** (2019) 1230–1234. doi: 10.1126/science.aav4040.

49. Oliveira S, Nägele K, Carlhoff S, Pugach I, Koesbardiati T, Hübner A, Meyer M, Oktaviana AA, Takenaka M, Katagiri C, Murti DB, Putri RS, Mahirta, Petchey F, Higham T, Higham CFW, O'Connor S, Hawkins S, Kinaston R, Bellwood P, Ono R, Powell A, Krause J, Posth C, Stoneking M. Ancient genomes from the last three millennia support multiple human dispersals into Wallacea. *Nat. Ecol. Evol*. **6** (2022) 1024–1034. doi: 10.1038/s41559-022-01775-2.

50. Papac L, Ernée M, Dobeš M, Langová M, Rohrlach AB, Aron F, Neumann GU, Spyrou MA, Rohland N, Velemínský P, Kuna M, Brzobohatá H, Culleton B, Daněček D, Danielisová A, Dobisíková M, Hložek J, Kennett DJ, Klementová J, Kostka M, Krištuf P, Kuchařík M, Hlavová JK, Limburský P, Malyková D, Mattiello L, Pecinovská M, Petriščáková K, Průchová E, Stránská P, Smejtek L, Špaček J, Šumberová R, Švejcar O, Trefný M, Vávra M, Kolář J, Heyd V, Krause J, Pinhasi R, Reich D, Schiffels S, Haak W. Dynamic changes in genomic and social structures in third millennium BCE central Europe, *Sci. Adv*. **7** (2021) eabi6941. doi: 10.1126/sciadv.abi6941.

51. Patterson N, Price AL, Reich D. Population structure and eigenanalysis, *PLoS Genet*. **2** (2006) e190. doi: 10.1371/journal.pgen.0020190.

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

53. Patterson N, Isakov M, Booth T, Büster L, Fischer CE, Olalde I, Ringbauer H, Akbari A, Cheronet O, Bleasdale M, Adamski N, Altena E, Bernardos R, Brace S, Broomandkhoshbacht N, Callan K, Candilio F, Culleton B, Curtis E, Demetz L, Carlson KSD, Edwards CJ, Fernandes DM, Foody MGB, Freilich S, Goodchild H, Kearns A, Lawson AM, Lazaridis I, Mah M, Mallick S, Mandl K, Micco A, Michel M, Morante GB, Oppenheimer J, Özdoğan KT, Qiu L, Schattke C, Stewardson K, Workman JN, Zalzala F, Zhang Z, Agustí B, Allen T, Almássy K, Amkreutz L, Ash A, Baillif-Ducros C, Barclay A, Bartosiewicz L, Baxter K, Bernert Z, Blažek J, Bodružić M, Boissinot P, Bonsall C, Bradley P, Brittain M, Brookes A, Brown F, Brown L, Brunning R, Budd C, Burmaz J, Canet S, Carnicero-Cáceres S, Čaušević-Bully M, Chamberlain A, Chauvin S, Clough S, Čondić N, Coppa A, Craig O, Črešnar M, Cummings V, Czifra S, Danielisová A, Daniels R, Davies A, de Jersey P, Deacon J, Deminger C, Ditchfield PW, Dizdar M, Dobeš M, Dobisíková M, Domboróczki L, Drinkall G, Đukić A, Ernée M, Evans C, Evans J, Fernández-Götz M, Filipović S, Fitzpatrick A, Fokkens H, Fowler C, Fox A, Gallina Z, Gamble M, González Morales MR, González-Rabanal B, Green A, Gyenesei K, Habermehl D, Hajdu T, Hamilton D, Harris J, Hayden C, Hendriks J, Hernu B, Hey G, Horňák M, Ilon G, Istvánovits E, Jones AM, Kavur MB, Kazek K, Kenyon RA, Khreisheh A, Kiss V, Kleijne J, Knight M, Kootker LM, Kovács PF, Kozubová A, Kulcsár G, Kulcsár V, Le Pennec C, Legge M, Leivers M, Loe L, López-Costas O, Lord T, Los D, Lyall J, Marín-Arroyo AB, Mason P, Matošević D, Maxted A, McIntyre L, McKinley J, McSweeney K, Meijlink B, Mende BG, Menđušić M, Metlička M, Meyer S, Mihovilić K, Milasinovic L, Minnitt S, Moore J, Morley G, Mullan G, Musilová M, Neil B, Nicholls R, Novak M, Pala M, Papworth M, Paresys C, Patten R, Perkić D, Pesti K, Petit A, Petriščáková K, Pichon C, Pickard C, Pilling Z, Price TD, Radović S, Redfern R, Resutík B, Rhodes DT, Richards MB, Roberts A, Roefstra J, Sankot P, Šefčáková A, Sheridan A, Skae S, Šmolíková M, Somogyi K, Somogyvári Á, Stephens M, Szabó G, Szécsényi-Nagy A, Szeniczey T, Tabor J, Tankó K, Maria CT, Terry R, Teržan B, Teschler-Nicola M, Torres-Martínez JF, Trapp J, Turle R, Ujvári F, van der Heiden M, Veleminsky P, Veselka B, Vytlačil Z, Waddington C, Ware P, Wilkinson P, Wilson L, Wiseman R, Young E, Zaninović J, Žitňan A, Lalueza-Fox C, de Knijff P, Barnes I, Halkon P, Thomas MG, Kennett DJ, Cunliffe B, Lillie M, Rohland N, Pinhasi R, Armit I, Reich D. Large-scale migration into Britain during the Middle to Late Bronze Age, *Nature*. **601** (2022) 588–594. doi: 10.1038/s41586-021-04287-4.

54. Peter BM. Admixture, Population Structure, and *F*-Statistics, *Genetics*. **202** (2016) 1485–1501. doi: 10.1534/genetics.115.183913.

55. Podani J, Miklós I. Resemblance coefficients and the horseshoe effect in Principal Coordinates Analysis, *Ecology*. **83** (2002) 3331–3343. doi: 10.1890/0012-9658(2002)083[3331:RCATHE]2.0.CO;2.

56. Prendergast ME, Lipson M, Sawchuk EA, Olalde I, Ogola CA, Rohland N, Sirak KA, Adamski N, Bernardos R, Broomandkhoshbacht N, Callan K, Culleton BJ, Eccles L, Harper TK, Lawson AM, Mah M, Oppenheimer J, Stewardson K, Zalzala F, Ambrose SH, Ayodo G, Gates HL Jr, Gidna AO, Katongo M, Kwekason A, Mabulla AZP, Mudenda GS, Ndiema EK, Nelson C, Robertshaw P, Kennett DJ, Manthi FK, Reich D. Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa. *Science*. **365** (2019) eaaw6275. doi: 10.1126/science.aaw6275.

57. Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW Jr, Orlando L, Metspalu E, Karmin M, Tambets K, Rootsi S, Mägi R, Campos PF, Balanovska E, Balanovsky O, Khusnutdinova E, Litvinov S, Osipova LP, Fedorova SA, Voevoda MI, DeGiorgio M, Sicheritz-Ponten T, Brunak S, Demeshchenko S, Kivisild T, Villems R, Nielsen R, Jakobsson M, Willerslev E. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*. **505** (2014) 87–91. doi: 10.1038/nature12736.

58. Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert MT, Wang Y, Raghavan M, Campos PF, Kamp HM, Wilson AS, Gledhill A, Tridico S, Bunce M, Lorenzen ED, Binladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre TL, Grønnow B, Meldgaard M, Andreasen C, Fedorova SA, Osipova LP, Higham TF, Ramsey CB, Hansen TV, Nielsen FC, Crawford MH, Brunak S, Sicheritz-Pontén T, Villems R, Nielsen R, Krogh A, Wang J, Willerslev E. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*. **463** (2010) 757–762. doi: 10.1038/nature08835.

59. Reich D, Price AL, Patterson N. Principal component analysis of genetic data, *Nat. Genet*. **40** (2008) 491–492.

60. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history, *Nature*. **461** (2009) 489–494. doi: 10.1038/nature08365.

61. Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, García LF, Triana O, Blair S, Maestre A, Dib JC, Bravi CM, Bailliet G, Corach D, Hünemeier T, Bortolini MC, Salzano FM, Petzl-Erler ML, Acuña-Alonzo V, Aguilar-Salinas C, Canizales-Quinteros S, Tusié-Luna T, Riba L, Rodríguez-Cruz M, Lopez-Alarcón M, Coral-Vazquez R, Canto-Cetina T, Silva-Zolezzi I, Fernandez-Lopez JC, Contreras AV, Jimenez-Sanchez G, Gómez-Vázquez MJ, Molina J, Carracedo A, Salas A, Gallo C, Poletti G, Witonsky DB, Alkorta-Aranburu G, Sukernik RI, Osipova L, Fedorova SA, Vasquez R, Villena M, Moreau C, Barrantes R, Pauls D, Excoffier L, Bedoya G, Rothhammer F, Dugoujon JM, Larrouy G, Klitz W, Labuda D, Kidd J, Kidd K, Di Rienzo A, Freimer NB, Price AL, Ruiz-Linares A. Reconstructing Native American population history, *Nature*. **488** (2012) 370–374. doi: 10.1038/nature11258.

62. Ringbauer H, Coop G, Barton NH. Inferring Recent Demography from Isolation by Distance of Long Shared Sequence Blocks, *Genetics*. **205** (2017) 1335–1351. doi: 10.1534/genetics.116.196220.

63. Scally A, Durbin R. Revising the human mutation rate: implications for understanding human evolution, *Nat. Rev. Genet*. **13** (2012) 745–753. doi: 10.1038/nrg3295.

64. Sirak KA, Fernandes DM, Lipson M, Mallick S, Mah M, Olalde I, Ringbauer H, Rohland N, Hadden CS, Harney É, Adamski N, Bernardos R, Broomandkhoshbacht N, Callan K, Ferry M, Lawson AM, Michel M, Oppenheimer J, Stewardson K, Zalzala F, Patterson N, Pinhasi R, Thompson JC, Van Gerven D, Reich D. Social stratification without genetic differentiation at the site of Kulubnarti in Christian Period Nubia, *Nat. Commun*. **12** (2021) 7283. doi: 10.1038/s41467-021-27356-8.

65. Skoglund P, Thompson JC, Prendergast ME, 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 JI, Chiumia C, Crowther A, Gomani-Chindebvu E, Gidna AO, Grillo KM, Helenius IT, Hellenthal G, Helm R, Horton M, López S, Mabulla AZP, Parkington J, Shipton C, Thomas MG, Tibesasa R, Welling M, Hayes VM, Kennett DJ, Ramesar R, Meyer M, Pääbo S, Patterson N, Morris AG, Boivin N, Pinhasi R, Krause J, Reich D. Reconstructing Prehistoric African Population Structure, *Cell*. **171** (2017) 59–71.e21. doi: 10.1016/j.cell.2017.08.049.

66. Soraggi S, Wiuf C. General theory for stochastic admixture graphs and *F*-statistics, *Theor. Popul. Biol*. **125** (2019) 56–66. doi: 10.1016/j.tpb.2018.12.002.

67. Taylor WTT, Librado P, American Horse CJ, Shield Chief Gover C, Arterberry J, Afraid of Bear-Cook AL, Left Heron H, Yellow Hair RM, Gonzalez M, Means B, High Crane S, Yellow Bull WW, Dull Knife B, Afraid of Bear A, Tecumseh Collin C, Ward C, Pasqual TA, Chauvey L, Tonasso-Calviere L, Schiavinato S, Seguin-Orlando A, Fages A, Khan N, Der Sarkissian C, Liu X, Wagner S, Leonard BG, Manzano BL, O'Malley N, Leonard JA, Bernáldez-Sánchez E, Barrey E, Charliquart L, Robbe E, Denoblet T, Gregersen K, Vershinina AO, Weinstock J, Rajić Šikanjić P, Mashkour M, Shingiray I, Aury JM, Perdereau A, Alquraishi S, Alfarhan AH, Al-Rasheid KAS, Trbojević Vukičević T, Buric M, Sauer E, Lucas M, Brenner-Coltrain J, Bozell JR, Thornhill CA, Monagle V, Perri A, Newton C, Hall WE, Conver JL, Le Roux P, Buckser SG, Gabe C, Belardi JB, Barrón-Ortiz CI, Hart IA, Ryder C, Sponheimer M, Shapiro B, Southon J, Hibbs J, Faulkner C, Outram A, Patterson Rosa L, Palermo K, Solé M, William A, McCrory W, Lindgren G, Brooks S, Eché C, Donnadieu C, Bouchez O, Wincker P, Hodgins G, Trabert S, Bethke B, Roberts P, Jones EL, Running Horse Collin Y, Orlando L. Early dispersal of domestic horses into the Great Plains and northern Rockies. *Science*. **379** (2023) 1316–1323. doi: 10.1126/science.adc9691.

68. Tricou T, Tannier E, de Vienne DM. Ghost Lineages Highly Influence the Interpretation of

Introgression Tests, *Syst. Biol*. **71** (2022) 1147–1158. doi: 10.1093/sysbio/syac011.

69. van de Loosdrecht M, Bouzouggar A, Humphrey L, Posth C, Barton N, Aximu-Petri A, Nickel B, Nagel S, Talbi EH, El Hajraoui MA, Amzazi S, Hublin JJ, Pääbo S, Schiffels S, Meyer M, Haak W, Jeong C, Krause J. Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. *Science*. **360** (2018) 548–552. doi: 10.1126/science.aar8380.
70. Wang K, Goldstein S, Bleasdale M, Clist B, Bostoen K, Bakwa-Lufu P, Buck LT, Crowther A, Dème A, McIntosh RJ, Mercader J, Ogola C, Power RC, Sawchuk E, Robertshaw P, Wilmsen EN, Petraglia M, Ndiema E, Manthi FK, Krause J, Roberts P, Boivin N, Schiffels S. Ancient genomes reveal complex patterns of population movement, interaction, and replacement in sub-Saharan Africa, *Sci. Adv*. **6** (2020) eaaz0183. doi: 10.1126/sciadv.aaz0183.
71. Wang CC, Yeh HY, Popov AN, Zhang HQ, Matsumura H, Sirak K, Cheronet O, Kovalev A, Rohland N, Kim AM, Mallick S, Bernardos R, Tumen D, Zhao J, Liu YC, Liu JY, Mah M, Wang K, Zhang Z, Adamski N, Broomandkhoshbacht N, Callan K, Candilio F, Carlson KSD, Culleton BJ, Eccles L, Freilich S, Keating D, Lawson AM, Mandl K, Michel M, Oppenheimer J, Özdoğan KT, Stewardson K, Wen S, Yan S, Zalzala F, Chuang R, Huang CJ, Looh H, Shiung CC, Nikitin YG, Tabarev AV, Tishkin AA, Lin S, Sun ZY, Wu XM, Yang TL, Hu X, Chen L, Du H, Bayarsaikhan J, Mijiddorj E, Erdenebaatar D, Iderkhangai TO, Myagmar E, Kanzawa-Kiriyama H, Nishino M, Shinoda KI, Shubina OA, Guo J, Cai W, Deng Q, Kang L, Li D, Li D, Lin R, Nini, Shrestha R, Wang LX, Wei L, Xie G, Yao H, Zhang M, He G, Yang X, Hu R, Robbeets M, Schiffels S, Kennett DJ, Jin L, Li H, Krause J, Pinhasi R, Reich D. Genomic insights into the formation of human populations in East Asia, *Nature*. **591** (2021) 413–419. doi: 10.1038/s41586-021-03336-2.
72. Yaka R, Mapelli I, Kaptan D, Doğu A, Chyleński M, Erdal ÖD, Koptekin D, Vural KB, Bayliss A, Mazzucato C, Fer E, Çokoğlu SS, Lagerholm VK, Krzewińska M, Karamurat C, Gemici HC, Sevkar A, Dağtaş ND, Kılınç GM, Adams D, Munters AR, Sağlıcan E, Milella M, Schotsmans EMJ, Yurtman E, Çetin M, Yorulmaz S, Altınışık NE, Ghalichi A, Juras A, Bilgin CC, Günther T, Storå J, Jakobsson M, de Kleijn M, Mustafaoğlu G, Fairbairn A, Pearson J, Togan İ, Kayacan N, Marciniak A, Larsen CS, Hodder I, Atakuman Ç, Pilloud M, Sürer E, Gerritsen F, Özbal R, Baird D, Erdal YS, Duru G, Özbaşaran M, Haddow SD, Knüsel CJ, Götherström A, Özer F, Somel M. Variable kinship patterns in Neolithic Anatolia revealed by ancient genomes. *Curr. Biol*. **31** (2021) 2455–2468.e18. doi: 10.1016/j.cub.2021.03.050.
73. Yang MA, Fan X, Sun B, Chen C, Lang J, Ko YC, Tsang CH, Chiu H, Wang T, Bao Q, Wu X, Hajdinjak M, Ko AM, Ding M, Cao P, Yang R, Liu F, Nickel B, Dai Q, Feng X, Zhang L, Sun C, Ning C, Zeng W, Zhao Y, Zhang M, Gao X, Cui Y, Reich D, Stoneking M, Fu Q. Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science*. **369** (2020) 282–288.

1703 doi: 10.1126/science.aba0909.

1704 74. Zhang F, Ning C, Scott A, Fu Q, Bjørn R, Li W, Wei D, Wang W, Fan L, Abuduresule I, Hu X, Ruan
1705 Q, Niyazi A, Dong G, Cao P, Liu F, Dai Q, Feng X, Yang R, Tang Z, Ma P, Li C, Gao S, Xu Y, Wu S,
1706 Wen S, Zhu H, Zhou H, Robbeets M, Kumar V, Krause J, Warinner C, Jeong C, Cui Y. The genomic
1707 origins of the Bronze Age Tarim Basin mummies. *Nature*. **599** (2021) 256–261. doi:
1708 10.1038/s41586-021-04052-7.