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

Eren Yüncü[1,*,†], Ulaş Işıldak[1,*,†], Matthew P. Williams[2], Christian D. Huber[2], Olga Flegontova[1,3], Leonid A. Vyazov[1], Piya Changmai[1], Pavel Flegontov[1,4,§]

[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] *Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, České Budějovice, Czechia*

[4] *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@fas.harvard.edu, pflegontov@gmail.com)

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/f_4-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 analyze 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 analyzed 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, and with or without a "model competition" step). We caution that high-throughput *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). 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. Our study has a number of limitations: for instance, these recommendations depend on the assumption that the underlying genetic history is a complex admixture graph and not a stepping-stone model.

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, Peter 2022) 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: both those approximated best by admixture graphs and isolation by distance (stepping-stone models, see Novembre and Stephens 2008, Duforet-Frebourg and Slatkin 2016).

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 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” hypothesized as signs of admixture between originally isolated groups at the ends of such clines. However, multiple studies on simulated data showed that clinal PCA patterns do not necessarily

indicate historical "admixture clines": Lissajous curves such as horseshoe-shaped clines appear in PC spaces whenever smooth gradients of various origin exist in data (Podani and Miklós 2002, Diakonis et al. 2008, Novembre and Stephens 2008, Frichot et al. 2012, Duforet-Frebourg and Slatkin 2016, Estavoyer and Francois 2022). Thus, similar clinal patterns in PCAs based on genetic data can arise for many reasons: admixture between previously isolated groups (i.e., "admixture clines"), isolation by distance (Novembre and Stephens 2008), isolation by time in a panmictic population (Duforet-Frebourg and Slatkin 2016), and serial founder models without any gene flow (Estavoyer and Francois 2022). 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) and may obscure smooth Lissajous curves by artefactual sharp angles (Novembre and Stephens 2008). Considering this complexity, 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 (Durand et al. 2011, Patterson et al. 2012) and SNP ascertainment bias (Flegontov et al. 2023) are absent. 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

107 of deviations of *D*- and f_4-statistics from 0 becomes hardly possible if both branch pairs are
108 connected by detectable gene flows.

109 "Admixture" f_3-statistics of the form f_3(target; proxy source$_1$, proxy source$_2$) constitute
110 another formal test for admixture (Patterson et al. 2012). Significant deviation of such a
111 statistic from 0 in the negative direction (Z-score below -3) is considered proof of admixture
112 since allele frequencies at most sites are intermediate in the target group between those in
113 the proxy sources (Patterson et al. 2012). However, "admixture" f_3-statistics are usually only
114 applicable for detection of recent admixture events since they become positive when post-
115 admixture genetic drift on the target lineage moves allele frequencies away from these
116 intermediate values (Patterson et al. 2012, Peter 2016, Peter 2022).

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

128 **Box 1.** Terminology used in this study for describing admixture models, results of screens for
129 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-graph-shaped histories, 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 deeper in time.

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 *p*-values. 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 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 $\pm$ 2 standard errors are between 0 and 1; 3) the *p*-value for a model with *n* ancestry sources $\geq$ 0.01. Other versions of this criterion are also found in the literature: with different values of the *p*-value cut-off (usually 0.05 or 0.01) and/or imposing limits on estimated admixture proportions, but not considering their standard errors.
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 FDR 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).

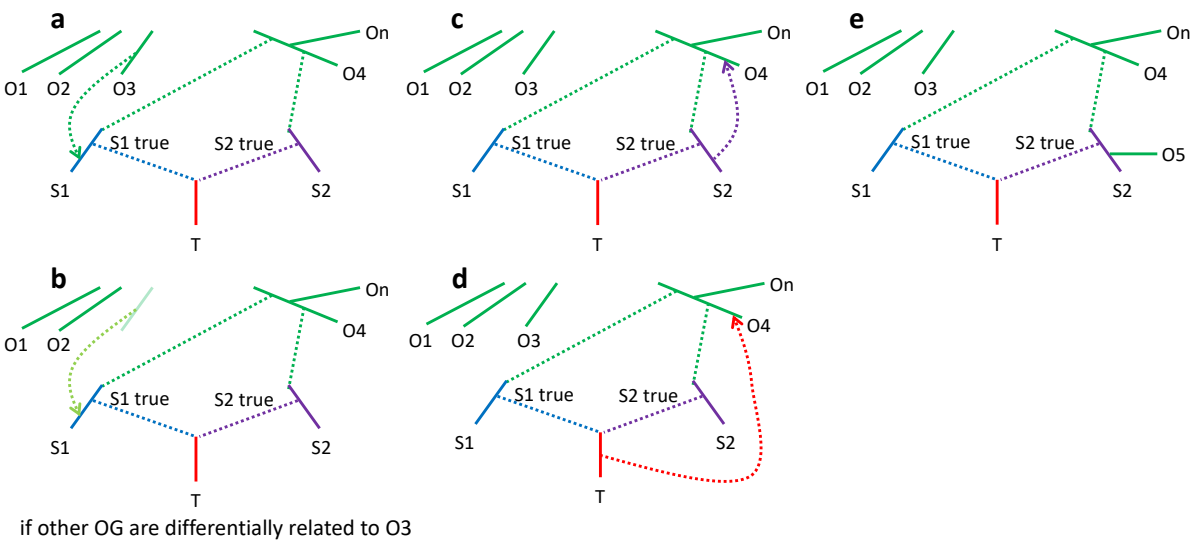


Figure 1. Admixture graphs showing an exhaustive list of assumption violations of the standard *qpAdm* protocol that may lead to rejection of the true simple model, and thus prompt the researcher to test overly complex models. (**a**) A gene flow from an outgroup O* to a proxy source after the divergence 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.

155 A *qpAdm* protocol that has become the standard in archaeogenetics (Lazaridis et al. 2016)
156 can be broken down into two parts: estimation of the number of gene flows connecting the
157 “right” and “left” sets (this method was first published as a tool named “*qpWave*”, Reich et
158 al. 2012) and inference of admixture proportions in a target group in the “left” set (Haak et
159 al. 2015). *qpWave* tests for the number of distinct gene flows connecting the "right" and "left"
160 population sets, does not infer directionality of gene flows, and does not identify recipients
161 of gene flow in the "left" or "right" population sets. Notably, the standard *qpAdm* protocol
162 relies on the following assumptions (Lazaridis et al. 2016, Harney et al. 2021): 1) there is at
163 least one "right" population differentially related to the proxy sources; 2) proxy sources are
164 strictly cladal with the true ancestral admixing sources (**Fig. 1a, b**), 3) there are no gene flows
165 to populations located in the “right” set from the proxy source or target lineages either after
166 the split of the proxy source from the true admixing source population or between the target
167 population and the admixture event that gave rise to it (**Fig. 1c-e**). In the context of our study,
168 true sources are unsampled populations that participated in a simulated admixture event
169 (labelled as “S1 true” and “S2 true” in **Fig. 1**, see also **Box 1**). A more concise definition of
170 these assumptions exists: there should be no genetic drift shared exclusively by proxy sources
171 and “right” populations (but not by the target) and shared exclusively by the target and
172 “right” populations (but not by proxy sources).

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

181 Harney *et al*. (2021) demonstrated on simulated data that, if the *qpAdm* assumptions are
182 satisfied, it is highly favourable for statistical power of the method (for distinguishing
183 between alternative proxy sources that are unequally distant genetically from the true
184 ancestry source) to move at least some alternative proxy sources between the "left" and
185 "right" sets. In other words, having "on the right" populations that do not violate the

186 topological assumptions of *qpAdm*, but are closely related to proxy sources on the "left",
187 increases the statistical power greatly (see also Ning et al. 2020 for another demonstration
188 of this on simple simulated histories).

189 This new type of *qpAdm* protocols, termed "rotating" protocol, has been adopted in
190 archaeogenetics widely (see, e.g., Skoglund et al. 2017, Harney et al. 2019, Narasimhan et al.
191 2019, Olalde et al. 2019, Calhoff et al. 2021, Fernandes et al. 2021, Librado et al. 2021,
192 Allentoft et al. 2022, Bergström et al. 2022, Lazaridis et al. 2022, Oliveira et al. 2022, Taylor
193 et al. 2023). The most extreme version of the "rotating" protocol simply divides a set of
194 reference populations into all possible combinations of "right" and "proxy source" subsets of
195 certain sizes and rotates these combinations through successive *qpAdm* analyses. Additional
196 constraints can be placed on the rotating combinations such as restricting a set of groups
197 (usually highly divergent from the target) to remain in the “right” set in all models. When
198 evaluating the totality of multiple *qpAdm* tests, the simplest feasible models (e.g., one-way,
199 i.e., unadmixed) are favoured, and increasingly complex models are explored upon the
200 rejection of simpler models. Model rejection for the simplest models is made according to a
201 chosen *p*-value threshold such that *qpAdm* models are considered feasible or "fitting" the
202 data when the *p*-value is above such a threshold (Skoglund et al. 2017, Harney et al. 2018,
203 Narasimhan et al. 2019, Olalde et al. 2019, Yang et al. 2020, Calhoff et al. 2021, Fernandes et
204 al. 2021, Librado et al. 2021, Zhang et al. 2021, Allentoft et al. 2022, Bergström et al. 2022,
205 Lazaridis et al. 2022, Oliveira et al. 2022, Taylor et al. 2023). As an additional criterion of a
206 fitting model, all inferred admixture proportions (Harney et al. 2018, Olalde et al. 2019, Yang
207 et al. 2020, Zhang et al. 2021, Allentoft et al. 2022, Lazaridis et al. 2022, Oliveira et al. 2022),
208 or proportions $\pm$ 2 standard errors (Narasimhan et al. 2019), may be required to lie between
209 0 and 1. It is important to remember that the statistical significance of the *qpAdm/qpWave*
210 test is, strictly speaking, a function of model rejection, and thus the failure to reject a model
211 may have underlying causes other than approximating the true situation well enough (such
212 as lack of statistical power or a lack of suitable “right” groups that capture the divergent
213 ancestry sources amongst the “left” set of populations).

214 A less exhaustive version of the rotating *qpAdm* protocol, termed "model competition" (e.g.,
215 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 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, 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 favorable 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 and a one-dimensional stepping-stone model with uniform and continuous gene flow among several groups was explored, and it was concluded that *qpAdm* is not applicable in this case (Harney et al. 2021). Meanwhile, false discovery rate (FDR) or 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

247 a nearly perfect tree, but that, on the other hand, cannot be approximated solely as isolation
248 by distance?

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

260

261 **Results**

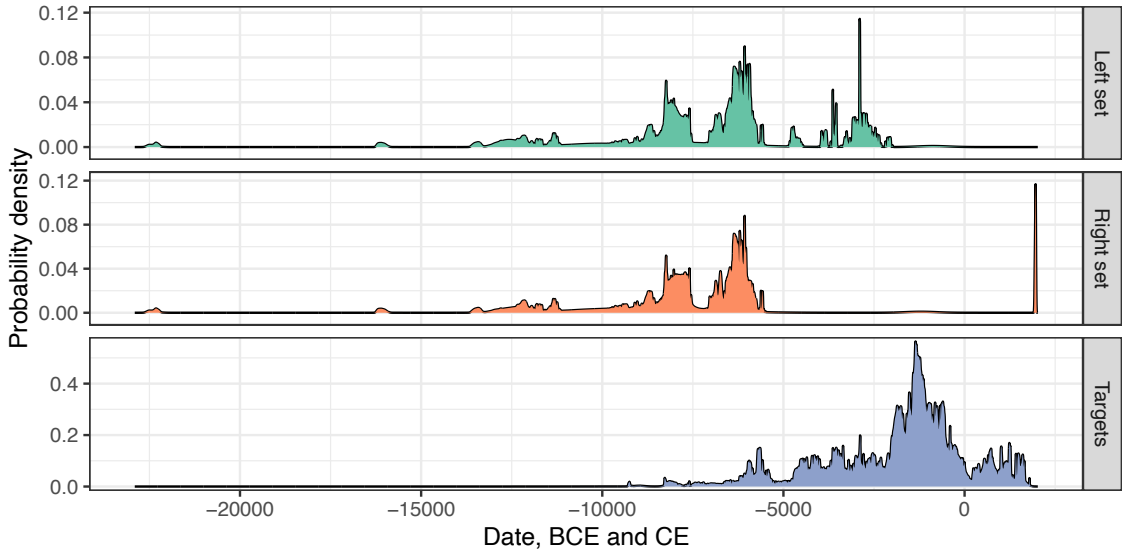
262 ***An overview of published rotating and model competition qpAdm protocols***

263 First, we outline two published rotating *qpAdm* protocols (Narasimhan et al. 2019, Lazaridis
264 et al. 2022) that are typical for this class of protocols (see further examples in Skoglund et al.
265 2017, Harney et al. 2019, Olalde et al. 2019, Calhoff et al. 2021, Fernandes et al. 2021, Librado
266 et al. 2021, Allentoft et al. 2022, Bergström et al. 2022, Oliveira et al. 2022, Taylor et al. 2023).
267 Lazaridis *et al*. relied on the following set of 15 reference populations: 1) Mbuti (present-day
268 Africans); 2) a Palaeolithic group from the Caucasus (CHG, Caucasian hunter-gatherers); 3)
269 East European Mesolithic (EHG, East European hunter-gatherers); 4) Ganj Dareh (a Neolithic
270 group from Iran); 5) Natufians (an Epipalaeolithic group from Israel); 6) a Pre-pottery
271 Neolithic (PPN) group from the Levant; 7) Taforalt (an Epipalaeolithic group from Morocco);
272 8) Neolithic Mesopotamia; 9) Afontova Gora 3 (an individual from Late Upper Palaeolithic
273 Siberia); 10) Mal'ta 1 (an individual from Late Upper Palaeolithic Siberia); 11) a Mesolithic
274 group from the Iron Gates region (Serbia); 12) Boncuklu (a Pre-pottery Neolithic group from
275 Central Turkey); 13) Barcın (a Neolithic group from Western Turkey); 14) Pınarbaşı (an

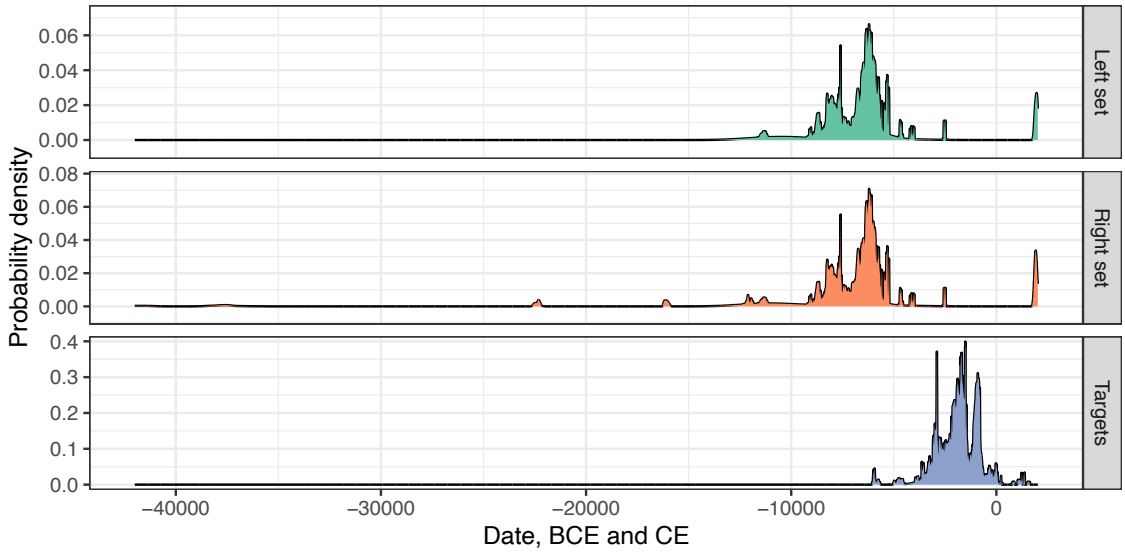
Epipalaeolithic individual from Turkey); and 15) Mesolithic and Palaeolithic individuals from Western Europe (WHG, West European hunter-gatherers).

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

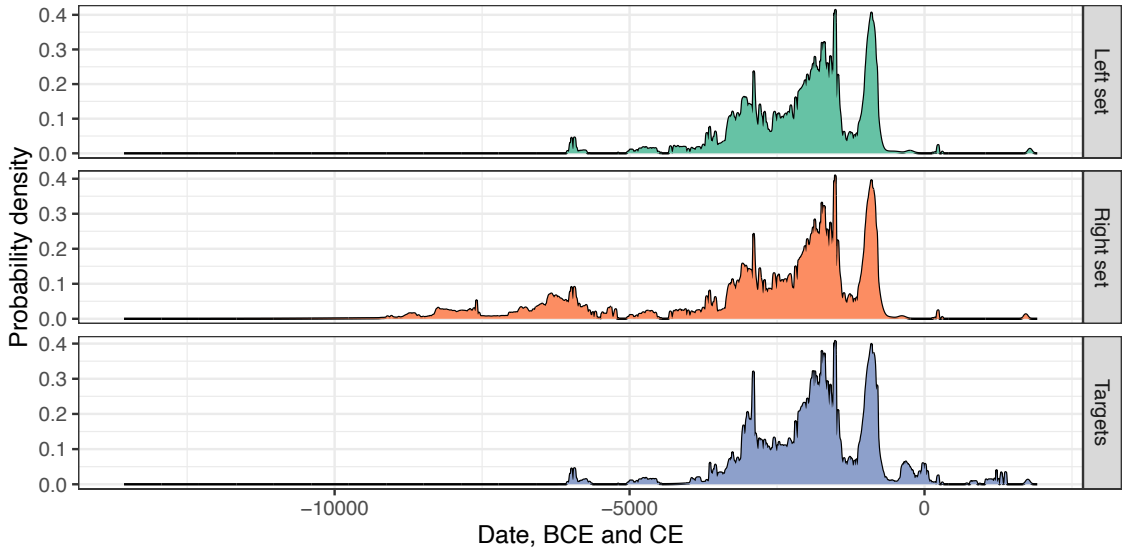


Figure 2. Distributions of radiocarbon and calendar dates for populations sets analyzed with the distal rotating *qpAdm* protocols by Lazaridis *et al*. 2022 (**a**) and Narasimhan *et al*. 2019 (**b**), and with the proximal model competition protocol from Narasimhan *et al*. 2019 (**c**). Probability 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

311 to ca. 24,000 years before present, yBP, Raghavan et al. 2014) appeared “on the left” in some
312 *qpAdm* models. Thus, “left-to-right” gene flows (that may lead to erroneous conclusions from
313 a *qpAdm* analysis, see **Fig. 1**) are expected to be common in the analytical setup used by
314 Lazaridis *et al*.

315 Narasimhan *et al*. (2019) used both proximal and distal *qpAdm* protocols (**Box 1**). The distal
316 rotating protocol relied on the following set of 16 reference populations: 1) Mota (a 4500-
317 years-old individual from Ethiopia); 2) Ust’-Ishim (an Upper Palaeolithic individual from West
318 Siberia); 3) Tianyuan (an Upper Palaeolithic individual from Northeast China); 4) Late Upper
319 Palaeolithic individuals from Siberia (Afontova Gora 3 and Mal’ta 1, collectively labelled
320 “ANE” or “Ancient North Eurasians”); 5) a Late Upper Palaeolithic individual from Italy
321 (Villabruna); 6) Natufians (an Epipalaeolithic group from Raqefet, Israel); 7) a Mesolithic
322 individual from Iran (Belt Cave); 8) present-day Andamanese; 9) East European Mesolithic
323 individuals (EEHG, East European hunter-gatherers); 10) West Siberian Mesolithic (WSHG,
324 West Siberian hunter-gatherers); 11) a Pre-pottery Neolithic (PPN) group from the Levant;
325 12) a Mesolithic group from the Iron Gates region (WEHG, West European hunter-gatherers);
326 13) Anatolian Neolithic individuals; 14) Ganj Dareh (a Neolithic group from Iran); 15) an Early
327 Neolithic group from the Baikal region (ESHG, East Siberian hunter-gatherers); and 16)
328 present-day Han Chinese. This reference set was split into all possible “right” and proxy
329 source subsets, except for the Upper Palaeolithic individuals/groups (Ust’-Ishim, Tianyuan,
330 ANE, Villabruna) who stayed in the “right” set in all models. Diverse groups from Iran,
331 Pakistan, Central Asia, and the Russian steppe zone (dated from the Chalcolithic to the
332 historical period) were used as targets for the *qpAdm* protocol. For each target, generally
333 post-dating the proxy sources (**Fig. 2b**), progressively more complex admixture models were
334 tested, from one- to five-way mixture models, and in most cases only the simplest feasible
335 models were interpreted. Model feasibility criteria were as follows: estimated admixture
336 proportions $\pm$ 2 standard errors are between 0 and 1, and *p*-value > 0.01 (Narasimhan et al.
337 2019).

338 In summary, this *qpAdm* protocol rotated a diverse set of groups between the "right" and
339 "left" sets: from present-day to Mesolithic groups older than 10,000 years, and from Africans
340 to South and East Asians (see date distributions in **Fig. 2b**). Another *qpAdm* protocol used by

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

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

370

371 ***Testing qpAdm performance on complex simulation histories***

372 Below we explore performance on simulated data (mainly FDR) of *qpAdm* protocols
373 representing the spectrum of protocols used in the literature (for an overview of our
374 workflow see **Fig. 3**). The most extreme example is a protocol where all groups are rotated
375 and all are treated alternatively as outgroups, targets, and proxy sources, i.e., there is no
376 temporal stratification between the latter categories. We term this protocol "proximal
377 rotating" (see **Tables S1 and S2**). Although such an extreme situation is, to our knowledge,
378 rare among published *qpAdm* protocols (see Calhoff et al. 2021, Oliveira et al. 2022; in the
379 latter study the rotating *qpAdm* strategy was used to model groups dated to 450–2600 yBP
380 as mixtures of present-day groups), we use it to illustrate the effects of poor temporal
381 stratification of targets and proxy sources in rotating protocols (**Fig. 2**). Models with targets
382 pre-dating proxy sources are encountered in high-throughput *qpAdm* screens, but do not
383 constitute a majority of models (Narasimhan et al. 2019, Librado et al. 2021, Allentoft et al.
384 2022, Bergström et al. 2022, Lazaridis et al. 2022, Taylor et al. 2023). We also explore FDR of
385 the proximal non-rotating (Harney et al. 2018, van de Loosdrecht et al. 2018, Narasimhan et
386 al. 2019, Prendergast et al. 2019, Wang et al. 2020, Calhoff et al. 2021, Wang et al. 2021,
387 Zhang et al. 2021, Changmai et al. 2022a, Changmai et al. 2022b, Maróti et al. 2022, Brielle
388 et al. 2023, Lee et al. 2023), distal rotating (Narasimhan et al. 2019, Librado et al. 2021,
389 Allentoft et al. 2022, Bergström et al. 2022, Lazaridis et al. 2022, Taylor et al. 2023), and distal
390 non-rotating protocols (Haak et al. 2015, Mathieson et al. 2015, Lazaridis et al. 2016, Antonio
391 et al. 2019, Mathieson et al. 2018, Marcus et al. 2020, Yang et al. 2020, Papac et al. 2021,
392 Yaka et al. 2021, Patterson et al. 2022) (**Tables 1 and 2**). In the distal protocols, only *qpAdm*
393 models where target group's sampling date is strictly contemporaneous with or post-dates
394 sampling of both proxy sources were considered.

395 We tested performance of these *qpAdm* protocols on complex simulated genetic histories in
396 the shape of admixture graphs: 13 populations connected with each other by admixture
397 graphs of random topology, including 10 pulse-like admixture events. Ten diploid individuals
398 with no missing data were sampled from each population at "leaves" of the graph. Forty such
399 random topologies were simulated, with an upper bound on the graph depth at 800
400 generations (ca. 23,000 years in the case of humans). These simulations generated sets of
401 populations sampled at widely different dates in the past or, in other words, located at

402 various genetic distances from the root, and matching intra-continental levels of human
403 genetic differentiation (median F_{ST} per simulated topology ranged from 0.022 to 0.051, see
404 also **Fig. S1**). Further details on the simulated population histories are presented in Methods
405 and illustrated by 14 examples in **Fig. S2**. To explore the influence of data amount on *qpAdm*
406 performance and compare it across protocols, we generated two independent sets of ten
407 simulation replicates for each graph topology: with genomes composed of three or ten 100-
408 Mbp-sized chromosomes (see **Fig. S3** for the number of SNP loci polymorphic in the simulated
409 datasets). These sets of simulations are referred to as "setup no. 1" and "setup no. 2" below.
410 To explore the parameter space further, we generated single simulation replicates for two
411 further setups: "setup no. 3", with maximal simulated history depth increased to 3,000
412 generations (ca. 87,000 yBP for humans), scaling all dates up proportionally; and "setup no.
413 4", with all terminal branches extended to the "present" of the simulation and sampled at
414 that point (**Fig. 3**). These latter simulations generated populations with median F_{ST} at the
415 inter-continental level (no. 3) or below it (no. 4, **Fig. S1**).

416 A typical archaeogenetic dataset is composed of pseudo-haploid data with high proportions
417 of missing sites and with widely different group sizes, including singleton groups. To generate
418 more realistic "noisy" data, we also performed randomized subsampling of SNP datasets for
419 simulation setups no. 1 and 2 (300- and 1,000-Mbp-sized genomes), for one simulation
420 replicate per each setup (**Fig. 3**, see Methods for details). The resulting data were pseudo-
421 haploid, had missing data rates ranging from 5% to 95% across individuals, and had uneven
422 group sizes ranging from 1 to 10 individuals. Ten independent subsampled datasets were
423 generated for each simulated topology (400 replicates in total per simulation setup),
424 including in the case of 300-Mbp-sized genomes from ca. 20,200 to 518,000 SNP loci with no
425 missing data at the group level and polymorphic across 13 groups (median = 89,700), and in
426 the case of 1,000-Mbp-sized genomes from ca. 66,600 to 2,095,700 such SNPs (median =
427 259,400, **Fig. S3**). Ranking of key simulation setups by these median site counts is as follows:
428 300-Mbp-sized genomes, low-quality data; 1,000-Mbp-sized genomes, low-quality data; 300-
429 Mbp-sized genomes, high-quality data; 1,000-Mbp-sized genomes, high-quality data (**Fig. S3**).

430

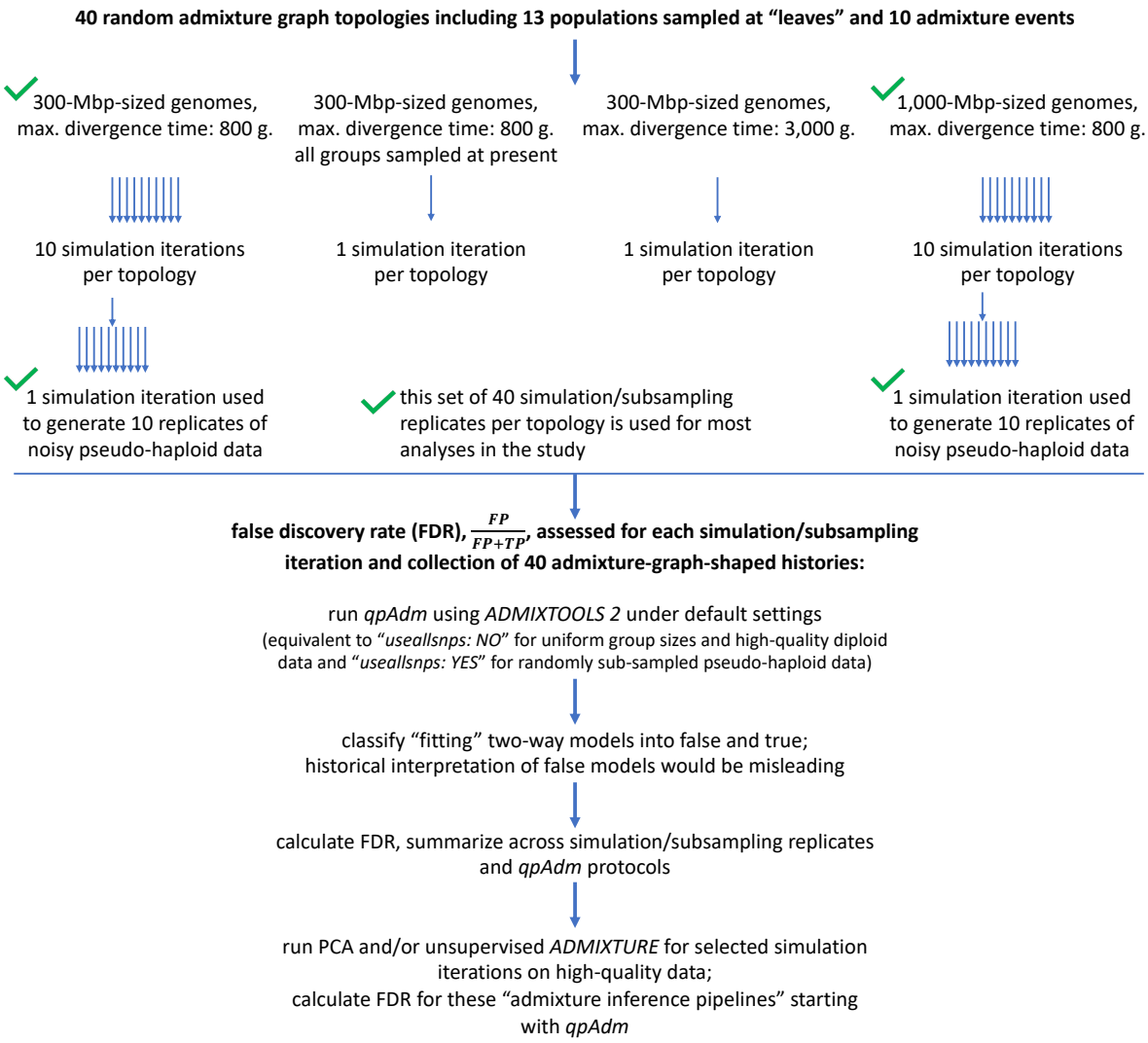


Figure 3. A diagram illustrating our analytical workflow.

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 by *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 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.

442 Relying on general principles, we argue that any high-throughput *qpAdm* protocol on poorly
443 understood complex demographic relationships approximated as admixture graphs and/or
444 stepping-stone models is questionable as a formal test for admixture since the *p*-value
445 threshold allows to *reject*, but not to *accept* models, and it is safer to interpret those models
446 as a certain number of gene flows connecting "left" and "right" sets in any direction, not in
447 the form of proxy sources and admixture proportions for a target. The model feasibility
448 criterion including both *p*-values and admixture proportions estimated with *qpAdm* is a
449 complex construct relying on the topological assumptions outlined in **Fig. 1**. We expect that
450 taking "left" and "right" sets that are not well-separated in time or contemporaneous (**Fig. 2**),
451 and where relationships among groups are poorly understood (which is almost always true
452 for exploratory studies), enriches the system for “left-to-right” gene flows, which in turn leads
453 to frequent rejection of true simple admixture models. Since the behavior of *qpAdm*
454 admixture proportion estimates under diverse demographic scenarios is poorly understood,
455 it is possible that a large fraction of these non-rejected complex models emerges as feasible,
456 resulting in false signals of admixture. Another approach probably leading to the same
457 outcome (enriching the system for “left-to-right” gene flows and other violations of the
458 topological assumptions) is selection of pairs of “right” populations maximizing deviations of
459 statistics $f_4(\text{left}_i, \text{left}_j, \text{right}_i, \text{right}_j)$ from zero since such deviations due to tree-like topologies
460 or “left-to-right” gene flows are impossible to tell apart. This approach is sometimes used for
461 increasing the power of *qpAdm* to discriminate between alternative proxy sources.

462 The *qpAdm* protocols we applied to the simulated data were focused on the simplest models:
463 one- and two-way admixture models (we note that histories that are more complex than two-
464 way mixtures were common in the data, **Fig. S2**). The model feasibility criterion followed
465 Narasimhan *et al*. (2019), see **Box 1** for a definition. Thus, we tested all possible two-way
466 admixture models for 40 complex population histories (34,320 proximal and distal rotating
467 models per simulation setup and replicate).

468 The non-rotating *qpAdm* approach was implemented as follows: for each simulated graph six
469 most ancient groups were selected as a fixed “right” set (ties were resolved in alphabetical
470 order; these “right” sets remained unchanged for a given graph topology across independent
471 simulations) and for the remaining seven groups all possible one-way and two-way admixture

472 models were tested, applying the same composite feasibility criterion that was used for the
473 rotating protocol (4,200 proximal and distal models per simulation setup and replicate).

474

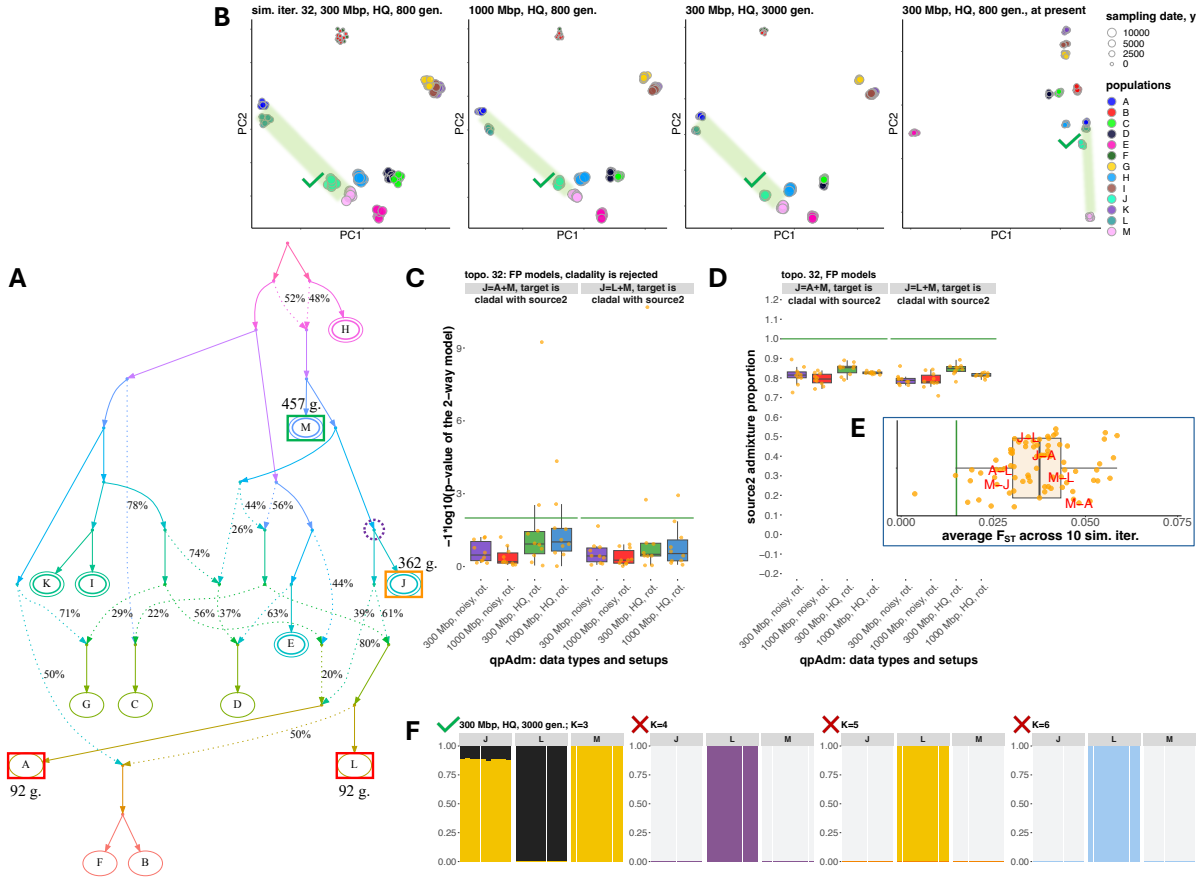
475 ***Case studies illustrating false and true feasible qpAdm models***

476 *False positive models*

477 In the context of complex and random admixture graph topologies it is hard to draw a strict
478 boundary between true and false admixture models composed of a target and only two proxy
479 sources. However, we classified the feasible *qpAdm* models into false and true ones relying
480 on a set of rules. By far the most common class of false feasible *qpAdm* models (referred to
481 as "false positive" or FP models), comprising 50.9% of all FP models generated by the
482 proximal rotating protocol across all the simulation and subsampling replicates (setups no. 1
483 and 2), occurs when the target group is falsely rejected as forming a clade with one or both
484 proxy sources whilst they are simulated as clades. In other words, these are false rejections
485 of single-source admixture models. Interestingly, under proximal non-rotating protocol
486 across all the simulation and subsampling replicates false rejection of cladality accounted
487 only for 10.1% of FP models, revealing that model violations become 'visible' under the
488 rotating protocol.

489 An example of the situation whereby the target group is falsely rejected as forming a clade is
490 shown in **Fig. 4** where the clade relationship between the target (*J*) and source (*M*) is rejected
491 by the rotating protocol due "left-to-right" gene flows violating the topological assumptions
492 of *qpAdm*, and more complex models ("*J = A + M*" and "*J = L + M*") are evaluated as true.
493 When a true one-way model "*J = M*" is tested, *M* represents a proxy for the only true source
494 of ancestry in *J*, and outgroups *A*, *C*, *D*, and *L* are partially derived from the target lineage,
495 resulting in rejection of the one-way model with *p*-values from ~10^{-32} to ~10^{-50} (in the case of
496 1000-Mbp-sized genomes and high-quality data). Rejections of two-way models "*J = A + M*"
497 and "*J = L + M*" according to *p*-values are rare (**Fig. 4c**), and nearly all estimated ancestry
498 proportions ±2 SE are between 0 and 1 even on low-quality pseudo-haploid data (**Fig. 4d**),
499 that is why these models are classified as fitting in nearly all cases (in 34 and 36 of 40
500 simulation/subsampling replicates, respectively). Groups *A* and *L* are 270 generations

501 younger than *J*, and >50% of their ancestry was simulated as derived from the *J* lineage (**Fig.**
502 **4a**). These FP models illustrate not only cladality rejection, but also incorrect gene flow
503 direction suggested by the *qpAdm* results.



504

505 **Figure 4.** A case study illustrating the most common class of false positive *qpAdm* models supported
506 by the proximal rotating protocol (accounts for 50.9% of all FP models across all the simulation and
507 subsampling replicates in **Table 2** and **Suppl. dataset 1**). Models of this type include at least one proxy
508 ancestry source that is simulated as fully cladal with the target. The other proxy source may be
509 simulated as a descendant of the target lineage (as shown here), may belong to an unrelated lineage
510 (**Fig. S4a**), or may be also cladal with the target. Both models shown here, "*J = A + M*" and "*J = L + M*",
511 are also fully supported by three-dimensional PCA and by an unsupervised *ADMIXTURE* analysis at
512 one or more *K* values. Below each panel of the figure is described in detail. **A.** Simulated history in the
513 form of an admixture graph (only topology is shown here, for divergence/admixture dates and
514 effective population sizes see the respective simulated history shown in **Fig. S2**). Sampled populations
515 are marked by letters. The target population from the *qpAdm* models illustrated here is enclosed in
516 an orange rectangle; correct proxy sources – in green rectangles, correct proxy sources with violations
517 of the topological assumptions of *qpAdm* – in green-yellow rectangles, and inappropriate proxy
518 sources, if any – in red rectangles. Sampling dates for these groups (in generations before present)
519 are shown beside the rectangles (the dates are from the "shallow" simulations up to 800 generations
520 deep). True simulated ancestry source(s) for the target are enclosed in dashed circles. Six populations
521 used as outgroups for the non-rotating *qpAdm* protocol are enclosed in double ovals. **B.** Two-
522 dimensional PCA plots for four simulated high-quality datasets indicated in plot titles. Simulated
523 populations are colored according to the legend on the right, and larger points correspond to older
524 sampling dates. The space between the proxy sources is shaded in green. If admixture model(s) are

525 supported by three-dimensional (sic!) PCA (PC1 vs. PC2 vs. PC3) according to the criteria listed in
526 Methods, a green tick mark is placed beside the target group in the plot, and a red cross mark is placed
527 otherwise. **C.** Boxplots summarizing *p*-values of two-way *qpAdm* models (indicated in plot titles)
528 across simulation or subsampling replicates, grouped by simulation setups and *qpAdm* protocols as
529 indicated on the x-axis. Green lines show the *p*-value threshold of 0.01 used in this study for model
530 rejection. *P*-values for one-way models are not shown for brevity. The *qpAdm* models shown here
531 were not tested using the non-rotating protocol since two "left" populations belong to the set of six
532 "right" populations. **D.** Boxplots summarizing estimated admixture proportions across simulation or
533 subsampling replicates, grouped by simulation setups and *qpAdm* protocols. The proportions are
534 shown either for the first or second proxy source, as indicated on the y-axis; green lines show the
535 simulated admixture proportion. **E.** All F_{ST} values for population pairs from the simulated history.
536 Average F_{ST} values are shown across 10 simulation replicates for 1000-Mbp-sized genomes and high-
537 quality data. Population pairs formed by components of the illustrated two-way *qpAdm* model(s) are
538 labelled in red. The green line shows median F_{ST} (0.015) for the Bronze Age West Asian populations
539 analyzed by Lazaridis et al. (2016) as an example of human genetic diversity at the sub-continental
540 scale. **F.** Ancestry proportions estimated with unsupervised *ADMIXTURE* for the groups constituting
541 the *qpAdm* model(s). For brevity, results are shown for one or two *qpAdm* models, for one simulated
542 dataset indicated in the plot title (usually simulation setup no. 3: 300-Mbp-sized genomes, max.
543 divergence date of 3,000 generations, high-quality data), and for four selected *K* values. If admixture
544 model(s) are supported by this analysis at a given *K* value, a green tick mark is placed beside the target
545 group in the plot, and a red cross mark is placed otherwise.

546

547 These results illustrate risks of applying rotating *qpAdm* protocols to sets including ancient
548 groups "on the left" and much younger groups "on the right" (Calhoff et al. 2021, Oliveira et
549 al. 2022), since those may be descended partially from the "left" groups (**Fig. 1c, d, e**) and
550 thus lead to false positive findings of gene flow. Three-dimensional PCA on all four datasets
551 we analyzed (**Fig. 4b**) and unsupervised *ADMIXTURE* results at some *K* values (**Fig. 4f**) support
552 these FP models. Thus, a researcher co-analyzing ancient and modern groups due to lack of
553 ancient samples from certain regions may encounter this situation in practice and, given
554 support by all methods in the standard archaeogenetic toolkit (*qpAdm*, PCA, and
555 *ADMIXTURE*), is likely to conclude that FP models are true. We illustrate another similar
556 situation often leading to rejection of cladality and acceptance of two-way or even more
557 complex models (in 33-37 of 40 simulation/subsampling replicates, depending on model and
558 *qpAdm* protocol) in **Fig. S4a**.

559 Other topological classes of FP models can be concisely described as follows (and are more
560 precisely defined in Methods): 1) gene flow goes from the target lineage (after the last
561 admixture event in its history) to a proxy source lineage (**Fig. 4, Fig. S4a, b**); 2) a proxy source
562 included in the model is symmetrically related to all real sources of ancestry in the target (see

563 examples of such feasible *qpAdm* models in **Fig. S4c, d**), or both proxy sources represent the
564 same true source and are symmetrically related to all the other true sources (**Fig. S4e**); 3)
565 both proxy sources represent distinct real sources, however a proxy source is a heavily
566 admixed group sharing less than 40% of ancestry with the real source (**Fig. S4a, d**); 4) the
567 target gets a gene flow from a deep-branching source not represented by any sampled
568 population, and an inappropriate proxy source is included in the fitting model (**Fig. S4f, g**).
569 Violations of the topological assumptions of *qpAdm* encountered in the examples of FP model
570 classes are described below. All the models shown below were selected among feasible
571 models that were outcomes of the proximal or distal rotating protocols.

572 In **Fig. S4b**, the target group, *G*, was simulated as a two-way mixture, and an incorrect
573 proximal model "*G = C + J*" emerged as fitting in 18 of 40 simulation/subsampling replicates.
574 Here, *C* is a group genetically similar to *G* (F_{ST} < 0.015 in **Fig. S4b**) and is largely its descendant
575 (87% of ancestry in *C* is derived from the *G/I* lineage, see also **Fig. S2c**) rather than a proxy
576 source (only 1.7% of its ancestry is derived from one of the true ancestry sources for *G*).
577 Despite the small F_{ST}, the cladality of groups *G* and *C* was rejected in 31 of 40
578 simulation/subsampling replicates due to simulated gene flows from outgroup (*L* and *M*)
579 lineages into *C* (**Fig. S4b**; this situation is schematically depicted in **Fig. 1a**). The model "*G = C*
580 *+ J*" was rejected according to *p*-values only when a lot of data was available (1,000-Mbp-
581 sized genomes and high-quality data, **Fig. S4b**). This rejection of the incorrect two-way model
582 does not mean that the outcome of *qpAdm* analysis for the target is acceptable since even
583 more complex three-way false models may end up being fitting, although we did not fit three-
584 way models in this study. A correct model we picked as an example, "*G = A + J*", was
585 universally rejected according to *p*-values (**Fig. S4b**) due to multiple violations of the
586 topological assumptions: for instance, outgroups *C* and *I* are derived from the target lineage
587 after the admixture event (**Fig. 1d**). However, removal of various violating outgroups did not
588 make the model "*G = A + J*" and similar models "*G = L/M + J*" fitting. On all datasets the correct
589 model "*G = A + J*" was not supported by PCA and *ADMIXTURE*, whereas the FP model "*G = C*
590 *+ J*" was supported by these methods in most cases (**Fig. S4b**). Moreover, no true two-way
591 model for group *G* emerged as feasible across the 40 simulation/subsampling replicates we
592 explored in this analysis (**Suppl. dataset 1**): other FP models were "*G = B + C*" (also supported
593 by PCA, see **Fig. S4b**) and "*G = I + X*". This case study is another illustration of incorrect

594 inference of gene flow direction by rotating *qpAdm* and of the risk associated with application
595 of rotating protocols in the situation when proxy sources post-date putative target groups
596 (group *C* is much younger than *G* and is a descendant of a population closely related to it, see
597 **Fig. S2c**). This case study also shows that certain combinations of simulated admixture graph
598 topologies and population sampling patterns create very unfavorable conditions for rotating
599 *qpAdm* protocols and for certain target groups lead to false findings exclusively; and these
600 false findings are supported by other methods such as PCA and *ADMIXTURE*.

601 In **Fig. S4c**, the target group, *A*, was simulated as a two-way mixture, and we expect that
602 proximal models "*A = B + C/K*" would be fitting. The topological assumptions are violated
603 when testing these models with the rotating protocol (**Fig. S4c**): groups *C* and *K* are cladal,
604 with one of them appearing in the "left" set and the other one in the "right" set (this situation
605 is schematically illustrated in **Fig. 1e**); a gene flow from an outgroup branch, *D*, enters the *C/K*
606 branch after it splits from one of the true ancestry sources (**Fig. 1a**). However, *p*-values of the
607 models "*A = B + C/K*" were high, but they were universally considered unfeasible due to
608 estimated admixture proportions being negative ("*A = B + K*" is shown as an example in **Fig.**
609 **S4c**). Removal of the outgroups *C, D,* and *K* does not make these models fitting. In contrast,
610 incorrect proximal models "*A = B + G/J*" often emerged as fitting. The proxy sources *G* and *J*
611 are symmetrically related to both true ancestry sources in the target. In this case proxy
612 sources *B* and *G/J* do not have equal standing: while group *B* is indeed related to one of the
613 two real sources for group *A*, *G* and *J* are groups completely "external" with respect to the
614 admixture history of *A*. The model "*A = B + G*" shown as an example in **Fig. S4c** was rejected
615 in some replicates according to *p*-values, but only if a large amount of data was available
616 (1000-Mbp-sized genomes, high-quality data). Overall, it was fitting in 18 of 40
617 simulation/subsampling replicates (**Suppl. dataset 1**). Both the false ("*A = B + G*") and true
618 positive ("*A = B + K*") models discussed as examples here were universally supported by three-
619 dimensional PCA and unsupervised *ADMIXTURE* (at some *K* values) on all datasets analyzed
620 (**Fig. S4c**). Of 72 feasible two-way models found for group *A* as a target across the 40
621 simulation/subsampling replicates (**Suppl. dataset 1**), all were classified as FP, and all
622 followed the pattern "*A = B + X*", with *G*, *J*, *D*, *E*, and *I* being the most common second sources.
623 Another example of an inappropriate proxy source symmetrically related to both true
624 ancestry sources for a target is shown in **Fig. S4d**: that is distal model "*E = K + L*" which was

625 usually rejected by the rotating protocol if enough data is available (feasible in 6 of 40
626 simulation/subsampling replicates) but was feasible in the case of the non-rotating protocol
627 (in 34 of 40 simulation/subsampling replicates). This model was also supported by PCA on all
628 four datasets used for this analysis.

629 In **Fig. S4e**, the target group, *A*, was simulated as a two-way mixture, and both correct and
630 incorrect two-way models were fitting approximately equally often for this target: across the
631 40 simulation/subsampling replicates, 38 true positive (TP) models and 47 FP models were
632 found (**Suppl. dataset 1**). We selected two models as examples: both correct ("*A = B + F*") and
633 incorrect ("*A = F + M*") proximal models fitted approximately equally often (in 28 and 23 of
634 40 replicates, respectively), suggesting that violations of the topological assumptions of
635 *qpAdm* play no role in the emergence of this FP model. FDR for target *A* varied between 56%
636 and 61% in the case of low-quality data and/or 300-Mbp-sized genomes; in contrast, for
637 1,000-Mbp-sized genomes and high-quality data FDR dropped to 14%. Failure to reject the
638 model "*A = F + M*", where both proxy sources represent only one true ancestry source and
639 are symmetrically related to the other, may be attributed to low F_{ST} between all the
640 populations involved in both models (**Fig. S4e**) and to the lack of data: increasing the
641 simulated genome size to 1,000 Mbp and using high-quality data resulted in rejection
642 (according to *p*-values) of the correct model "*A = B + F*" in 4 of 10 replicates, and in rejection
643 of the incorrect model "*A = F + M*" in 9 of 10 replicates (**Fig. S4e; Suppl. dataset 1**). Both
644 models shown here as examples had universal support by PCA and support by unsupervised
645 *ADMIXTURE* (at some *K* values; **Fig. S4e**). In this case study, the two-way model "*A = F + M*"
646 would lead to misleading historical interpretations: although the target is indeed a two-way
647 mixture, one of the true sources is completely missing in the model.

648 In **Fig. S4d**, the target group, *E*, was simulated as a two-way mixture, but no appropriate
649 source proxy was sampled for one of the true ancestry sources: in groups *H* and *A*, 30% and
650 6% of their ancestry, respectively, is derived from that true source. Thus, multiple gene flows
651 from the "right" set enter the *A* lineage after its split from the true ancestry source, and the
652 same is true for the *H* lineage, and these are violations of the topological assumptions of
653 *qpAdm* (**Fig. 1a**). The other true source for group *E* has a closely related proxy sampled (group
654 *L*, see **Fig. S2g**), and F_{ST} between groups *E* and *L* is low, < 0.015 (**Fig. S4d**). Despite this low F_{ST},
655 the (inaccurate) one-way model "*E = L*" was rejected by both the rotating and non-rotating

656 *qpAdm* protocols in 39 of 40 simulation/subsampling replicates (**Suppl. dataset 1**). The two-
657 way model "*E = H + L*" was rejected according to *p*-values in 37 of 40 replicates; in contrast,
658 the proximal model "*E = A + L*" was rejected according to *p*-values only in the case of 1,000-
659 Mbp-sized genomes and high-quality data (**Fig. S4d**). A similar proximal model, "*E = A + M*",
660 was usually rejected by the rotating protocol (feasible in 6 of 40 replicates; group *L* is closer
661 to the respective real source than group M), but both models "*E = A + L/M*" were feasible if
662 the non-rotating protocol was used (in 36 or 38 of 40 replicates). The ancestry proportion
663 contributed by the proxy sources *L* or *M* was consistently overestimated by all the models
664 and protocols we investigated (**Fig. S4d**); this appears to be related to the lack of appropriate
665 proxies for the other true ancestry source. Accepting models like "*E = A + L/M*" at face value
666 may lead to erroneous historical interpretations as the "present-day" group *A* is genetically
667 far from the real source, having diverged from it 390 generations ago and having received
668 94% of its ancestry from other sources (**Fig. S2g**). But all thresholds for classifying models of
669 this type into false and true are arbitrary. We chose 40% as a threshold percentage of proxy
670 source's ancestry derived from the corresponding true source. All 212 two-way feasible
671 models for target *E* (across 40 replicates and both rotating and non-rotating protocols) were
672 classified as false (**Suppl. dataset 1**). Since no true model was feasible for target *E*, rejection
673 of an incorrect two-way model given ample data is not an acceptable result, and even more
674 complex false models (which we did not test) may emerge as feasible. Models "*E = A + L/M*"
675 were supported by PCA on all four datasets we explored and by unsupervised *ADMIXTURE* on
676 two of these datasets (**Fig. S4d**). This case study illustrates again that applying *qpAdm*
677 protocols to poorly sampled regions of the world (where no close proxies are available for
678 some key ancestry sources) and without temporal stratification of "right" and "left" sets may
679 be risky.

680 In **Fig. S4f**, the target group, *H*, was simulated as a two-way mixture, with one source being
681 the deepest lineage in the graph, and the other source close to a sampled group, *M*. This
682 situation is similar to the basal Eurasian hypothesis (Lazaridis et al. 2016). A group of FP
683 models was found for this target. One way-models "*H = M*" were almost always rejected (in
684 39 of 40 simulation/subsampling replicates; **Suppl. dataset 1**), despite low F_{ST} for this
685 population pair (**Fig. S4f**), and two-way distal models "*H = F/I/L + M*" were often fitting,
686 especially the model "*H = F + M*" which was fitting in 35 of 40 simulation/subsampling

687 replicates including those with the highest amount of data. Group *F* is the second-deepest
688 lineage in this simulated graph after the true ancestry source for *H*, but it is distinct from the
689 latter. The distal model "*H = F + M*" was universally supported by three-dimensional PCA and
690 negative f_3-statistics f_3(target; proxy source$_1$, proxy source$_2$) (**Fig. S4f, Suppl. text 1**), as well
691 as strongly supported by the unsupervised *ADMIXTURE* analysis (**Fig. S4f**) in the case of
692 simulation setup no. 3 (at 7 of 8 *K* values tested). Given the simulated history and population
693 sampling pattern, a researcher relying on the standard archaeogenetic toolkit would
694 conclude that group *H*, sampled 44 generations ago, is derived from a lineage closely related
695 to *M*, sampled 74 generation ago, and an extinct lineage related to a much more ancient
696 group *F* sampled 604 generation ago. While the conclusion about deeply-divergent ancestry
697 in *H* is correct, and the proportion of that ancestry estimated on low- or high-quality data is
698 precise (**Fig. S4f**), its true source is a separate unsampled lineage not cladal with *F*.
699 Interpretation of such a *qpAdm* model in terms of population history would be misleading if
700 the true divergence point of the deep ancestry is not revealed by an admixture graph analysis
701 similar to that supporting the basal Eurasian hypothesis (Lazaridis et al. 2016). Importantly,
702 all 113 two-way feasible models for target *H* (across 40 replicates and both rotating and non-
703 rotating protocols) were classified as false (**Suppl. dataset 1**).

704 Another situation where deeply-divergent ancestry in a target is misrepresented by an
705 inappropriate proxy source is model "*F = A + E*" in **Fig. S4g**. In contrast to the situation in **Fig.**
706 **S4f**, the deeply-divergent ancestry in *F* is not unique to this group, but also contributed to
707 several other sampled groups. The distal model "*F = A + E*" was feasible in 24 of 40
708 simulation/subsampling replicates (for both rotating and non-rotating *qpAdm* protocols;
709 **Suppl. dataset 1**). It was weakly rejected when a lot of data was available (**Fig. S4g**), and on
710 the lowest-quality data we tested (300-Mbp-sized genomes, noisy data) it was often rejected
711 (e.g., in 5 of 10 replicates analyzed with the rotating protocol) due to negative admixture
712 proportion estimates and/or high *p*-values for one-way models. In contrast to most models
713 discussed above, this model was not supported by PCA and unsupervised *ADMIXTURE* (**Fig.**
714 **S4g**). Of 70 two-way feasible models for target *F* (across 40 replicates and both rotating and
715 non-rotating protocols), 69 were classified as false. Thus, this situation would likely be
716 ambiguous for a researcher encountering it in practice.

717

718

719 *True positive models*

720 Above we have discussed several examples of FP models that in practice would result in
721 misleading conclusions about population history, and now we turn to feasible models that
722 were interpreted as true in our analysis. Two-way admixture models for targets with
723 population history best approximated with three-way and more complex models were
724 considered as true positives if they included at least two source proxies *not* satisfying the
725 topological criteria of false models listed above. We also note that TP models were not
726 restricted to those including most optimal source proxies, if the models do *not* satisfy the
727 false positivity criteria. On a random sample of 400 two-way admixture models from our 40
728 simulated histories, the fraction of models that were classified as appropriate (true)
729 according to the rules described above was 17.7%. Since groups that are truly admixed are
730 common in our simulations, we do not expect to encounter a "needle in a haystack" situation
731 where finding true admixture models is exceedingly hard.

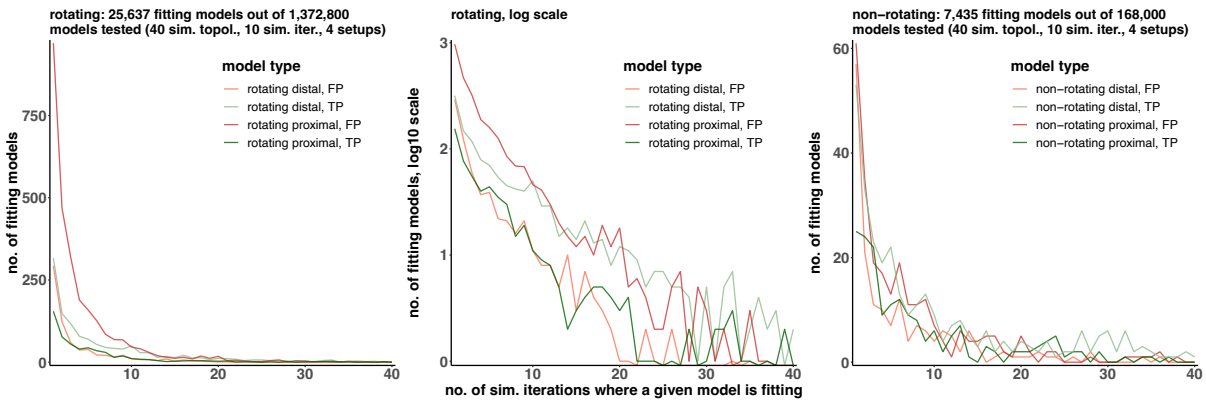
732 In **Fig. S4h** (and **Fig. S2a**), a set of ideal distal TP models is presented. The target, group *K*, was
733 simulated as a two way-mixture, and close (*H, I*) and more distant proxies (*G, M*) for both true
734 sources were sampled. All four alternative proxy sources display no violations of the
735 topological assumptions of *qpAdm* (**Fig. 1**), and outgroups such as *L* and *C* related
736 differentially to the true and proxy sources are available (**Fig. S4h**). An ideal model including
737 the closest proxy sources sampled, "*K = H + I*", and a nearly ideal model (with one proxy being
738 more distant from the true source), "*K = G + I*", were nearly universally fitting when the
739 rotating protocol was applied (in 38 and 35 of 40 simulation/subsampling iterations,
740 respectively) and yielded precise estimates of admixture proportions (**Fig. S4h**). However, the
741 non-rotating protocol demonstrated low power for this target on noisy pseudo-haploid data,
742 resulting in models being unfeasible due to high *p*-values for one-way models and/or negative
743 admixture proportion estimates: the "*K = H + I*" and "*K = G + I*" models were feasible in just 2
744 or 4 of 20 subsampling iterations, respectively. The two-way models were not rejected
745 according to *p*-values in this case, and had usually non-negative, but imprecise, admixture
746 proportion estimates (**Fig. S4h**). This observation illustrates the power of rotating protocols
747 (Harney et al. 2021) and the motivation behind their adoption. In the case of high-quality data

748 though, both models were unfeasible in just 1 of 20 simulation iterations analyzed by the
749 non-rotating protocol. Non-optimal models "*K = H + M*" and "*K = G + M*" were rejected
750 according to *p*-values by the rotating protocol (but not the non-rotating one) if enough data
751 was available (**Fig. S4h**). All 112 two-way feasible models for target *K* (across 40 replicates
752 analyzed by the rotating protocol) were classified as true, but of 142 feasible models that
753 were outcomes of the non-rotating protocol, just 86 were classified as true (**Suppl. dataset**
754 **1**). These observations demonstrate again the advantage of the rotating protocol (in rejecting
755 models with non-optimal or inappropriate proxy sources) that exist when there are no
756 violations of the topological assumptions of *qpAdm*. All four models for *K* had universal
757 support by three-dimensional PCA and unsupervised *ADMIXTURE* analyses (at least at some
758 *K* values, see **Fig. S4h**).

759 Another series of examples (**Fig. S4i-j, Fig. S2c, e**) illustrates models that very often emerge
760 as feasible despite imperfect cladality of proxy and true ancestry sources (see schematic
761 representations in **Fig. 1a, b**). The targets in these cases were simulated as two-way mixtures.
762 Despite non-cladality of one proxy source with the corresponding true source, distal model
763 "*C = I + K*" (**Fig. S4i**) was usually feasible when both the rotating and non-rotating protocols
764 were applied (in 37 and 30 of 40 simulation/subsampling replicates, respectively) and yielded
765 precise estimates of admixture proportions (**Fig. S4i**; **Suppl. dataset 1**). The proximal models
766 illustrated in **Fig. S4j**, "*K =C + D*" and "*K = D + L*", demonstrated violations of the topological
767 assumptions at one or two proxy sources, were almost universally rejected by the rotating
768 (but not the non-rotating) protocol, and produced estimates of admixture proportions that
769 were always biased (**Fig. S4j**). Given that the true model in this case is two-way, rejection of
770 a model due to non-cladality of true and proxy sources may lead to erroneous interpretations
771 if more complex (and false) three-way models emerge as feasible. However, as stated above,
772 three-way models were not explored in this study. All true (or false) two-way models for
773 target *K* were rejected by the rotating protocol on high-quality data (300- or 1000-Mbp-sized
774 genomes). Thus, this case study (**Fig. S4j**) demonstrates a disadvantage of the rotating
775 protocol when all proxies sampled for a certain true ancestry source demonstrate violations
776 of the topological assumptions of *qpAdm*.

777 Finally, we consider examples of targets that were simulated as three-way mixtures, but for
778 whom two-way *qpAdm* models were often feasible (**Fig. S4i, k-m**). In this case rejection of

such a two-way model is considered a desired outcome and was indeed observed very often when the rotating protocol was applied to high-quality data. In **Fig. S4i** (target *E*), all three true sources were represented by sampled proxies, however two of them demonstrated non-cladality with the true sources. A two-way model, "*E = C +K*", was rejected by the rotating (but not the non-rotating) protocol only when a large amount of data was available and, surprisingly, yielded precise estimates of admixture proportions (**Fig. S4i**). This incomplete (two-way) model for target *E* was also not supported by PCA and *ADMIXTURE* (**Fig. S4i**). In **Fig. S4k**, two-way models for a three-way admixed target were almost always rejected by the rotating protocol (but not the non-rotating protocol) according to *p*-values. In one case ("*I = D + J*") rejection was driven also by negative estimates of admixture proportions (**Fig. S4k**). The two-way models for target *I* were overwhelmingly supported by PCA and *ADMIXTURE* (**Fig. S4k**). That is, the third ancestry component in group *I* is largely invisible to these methods. A similar situation is illustrated in **Fig. S4l**, where both the rotating and non-rotating protocols tend to reject all three two-way models shown (according to *p*-values), but only if a lot of data is available (1000-Mbp-sized genomes and high-quality data). The two-way models for target *H* were also overwhelmingly supported by PCA and *ADMIXTURE* (**Fig. S4l**). In **Fig. S4m**, an opposite situation is shown. The model "*D = F + I*" was feasible in nearly all cases (35 of 40 replicates when the rotating protocol was applied and 38 of 40 replicates when the non-rotating protocol was applied; **Suppl. dataset 1**) since proxy source *I* represents two true sources at the same time (**Fig. S4m**), however it lacked support by PCA and *ADMIXTURE* (**Fig. S4m**). Ancestry proportions were estimated with a small bias since proxy source *I* does not have the correct ratio of ancestries. We stress that whenever two-way models were feasible for targets modelled as more complex mixtures, they were nevertheless considered as TP in our subsequent analyses.



804 **Figure 5.** Plots illustrating that for any *qpAdm* protocol or model type (false or true), the vast majority
805 of models only fit the data sporadically. Model feasibility was assessed across the 40 simulation or
806 subsampling replicates analyzed by the rotating and non-rotating *qpAdm* protocols. Models are
807 binned by the number of simulation/subsampling replicates where a model is feasible (fitting) (on the
808 x-axis). Separate decay curves are shown for TP and FP, proximal and distal two-way admixture
809 models. In the case of the rotating protocol, two plots illustrate the same data: a linear-scale plot on
810 the left and a log-scale plot in the center.

811

812 The case studies discussed above illustrate advantages of the rotating *qpAdm* protocol over
813 the non-rotating one, but also its problems, especially in the case of proximal models. The
814 admixture models discussed as examples above were feasible in large fractions of the 40
815 simulation or subsampling replicates (**Fig. 3**) analyzed with at least one *qpAdm* protocol
816 (rotating or non-rotating), however there is a long tail composed of thousands of models that
817 were just sporadically feasible in our analysis (**Fig. 5**). This exponential decay pattern is
818 common to FP and TP, proximal and distal models.

819 Below we move from case studies on selected target groups and selected simulated histories
820 to exploring performance of *qpAdm* protocols on a collection of random complex genetic
821 histories through the lens of FDR. While FP *qpAdm* models are expected for complex genetic
822 histories, the practical usage of *qpAdm* protocols relies on an assumption that false positives
823 are rare. However, FDR of the four *qpAdm* protocols tested here (rotating and non-rotating,
824 proximal and distal) varied between 12.1% and 68.1% (across all simulation setups and
825 replicates summarized in **Table 1**). Key statistics in our study are false discovery rate (FDR)
826 and false omission rate (FOR); see **Box 1** for definitions. We estimated FDR and FOR instead
827 of false positive and false negative rates of the *qpAdm* protocols and other methods due to a
828 technical limitation: the process of model classification into true and false ones cannot be
829 fully automated since it requires careful interpretation of the topology and simulated
830 admixture proportions. Therefore, classifying all possible 34,320 two-way admixture models
831 (858 per simulated topology) into true and false was hardly feasible. We estimated false
832 positive (FPR) and true positive rates (TPR) only approximately, relying on the fractions of
833 negatives and positives in random samples of two-way admixture models (see below). FPR in
834 our context is the probability that a two-way model with an unadmixed target and/or

835 inappropriate proxy source(s) emerges as fitting in a single *qpAdm* test. TPR in our context is
836 the probability that a two-way model with an admixed target and both proxy sources being
837 appropriate emerges as fitting in a *qpAdm* test.

838

839 ***Influence of the amount of data and temporal stratification on the performance of qpAdm***
840 ***protocols***

841 The amount of data (3.3-fold difference in simulated genome sizes) had no influence on FDR
842 of any *qpAdm* protocols in the case of randomly subsampled pseudo-haploid data (no
843 statistically significant difference was found between sets of 10 subsampling replicates in the
844 case of four *qpAdm* protocols; **Table 2, Fig. 6**). In contrast, small but statistically significant
845 influence of the data amount on FDR (according to the two-sided Wilcoxon test) was
846 observed for three of four *qpAdm* protocols when applied to high-quality data: proximal
847 rotating and non-rotating, and distal non-rotating (**Table 2, Fig. 6**). For a complete list
848 including 33,072 feasible two-way *qpAdm* models on which **Table 2** and **Fig. 6** are based (and
849 for corresponding one-way *qpAdm* models) see **Suppl. dataset 1**.

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

861 However, in the case of both proximal and distal non-rotating protocols, adding data led to a
862 small but statistically significant decrease in FDR: 42.6% vs. 38.4%, 31.2% vs. 26.5% (**Table 2,**
863 **Fig. 6**), 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. Notably, 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 overall (as low as 16.4%; **Table 2, Fig. 6**). 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. 6**).

simulation scheme	max. depth of the simulation, gen.	pseudohaploid noisy data	genome size, Mbp	two-way feasible models binned by: model type	protocol	bin size: median (min.-max.)	fraction of models that are distal: median	fraction of models that are distal: min.-max.	FDR for proximal & distal models: median	FDR for proximal & distal models: min.-max.	FDR for distal models: median	FDR for distal models: min.-max.	supported by 3D PCA: bin size:	fraction of models that are distal:	FDR for proximal & distal models:	FDR for distal models:	supported by unsupervised *ADMIXTURE*: bin size:	fraction of models that are distal:	FDR for proximal & distal models:	FDR for distal models:
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%		
at "present", no. 4	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**.**

878 Another kind of temporal stratification, stratification of the "right" and "left" sets, was a part of the
879 non-rotating protocols, but not of the rotating ones. In the left half of the table, feasible two-way
880 models are binned by model type (false or true positive) and *qpAdm* protocol (rotating or non-
881 rotating), and for each bin its size (number of models) and fraction of models that are distal (magenta
882 bars) are shown. FDR values are visualized by cell background color. In the right half of the table these
883 bins are further subdivided into models supported/not supported (highlighted in green and red,
884 respectively) by another method (PCA or *ADMIXTURE*). In the case of simulation setups no. 1 and 2,
885 10 simulation replicates and 10 subsampling replicates derived from simulation replicate no. 1 were
886 generated, and median, minimal, and maximal values across the replicates are shown for sizes of the
887 bins, for fractions of models that are distal, and for FDR.

888

889 The fact that the decrease in FDR with increasing amount of data is small is possibly explained
890 in the following way. *P*-values of individual two-way models (FP or "non-ideal" TP models) in
891 most cases decrease exponentially with the amount of data (**Fig. S4**). However, model
892 rejection is not always "beneficial" for protocol's performance assessed through FDR since
893 both models with inappropriate proxy sources (false models) and true models with violations
894 of the topological assumptions (**Fig. 1**) are rejected more often when more data is added.

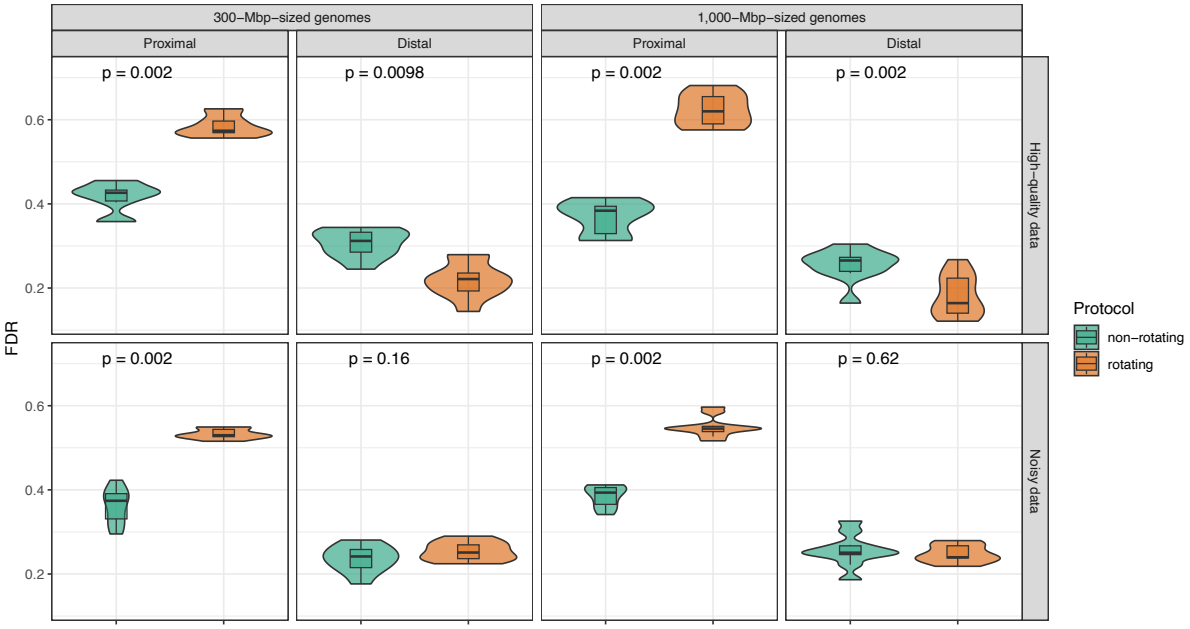
two-sided Wilcoxon test, *p*-value (paired test if *qpAdm* protocols differ, but simulation parameters are the same)									
max. depth of the simulation: 800 gen.		proximal rotating	proximal non-rotating	distal rotating	distal non-rotating	proximal rotating	proximal non-rotating	distal rotating	distal 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%

895

896 **Table 2.** Comparing FDR of four *qpAdm* protocols, proximal and distal, rotating and non-rotating, on
897 10 subsampling replicates derived from a single simulation replicate (low-quality data: top box) or on
898 10 simulation replicates and high-quality genomes (300-Mbp-sized or 1000-Mbp-sized genomes:
899 bottom box). For details on generating replicates of low-quality data see Methods. Median FDR values
900 for *qpAdm* protocols are shown on the diagonal and color-coded from largest (red) to smallest (green).

901 Significance of FDR differences between each *qpAdm* analysis protocol was tested using the two-sided
902 Wilcoxon test and the results are shown in the cells above the diagonal FDR values. For significance
903 of FDR differences between 300-Mbp and 1000-Mbp-sized genomes see the top right cell diagonal
904 under 1000-Mbp-sized genomes.

905



906

907 **Figure 6.** Distributions of FDR values across 10 simulation replicates (for high-quality data) or across
908 10 subsampling replicates derived from a single simulation replicate (for low-quality data).
909 Distributions are summarized with boxplots and violin plots for four *qpAdm* protocols (proximal and
910 distal, rotating and non-rotating) and two simulated genome sizes (300 Mbp and 1,000 Mbp). *qpAdm*
911 protocols were compared with the paired two-sided Wilcoxon test, and *p*-values for "rotating vs. non-
912 rotating" comparisons are shown in the panels (for the other *p*-values see **Table 2**).

913

914 Next, we assessed the influence of *qpAdm* protocol details on FDR. Temporal stratification of
915 targets and proxy sources (the former are not allowed to pre-date the latter) is the best way
916 of reducing FDR, according to our analysis: from ~50%–60% to ~15%–25% for rotating
917 protocols, and from ~40% to ~25%–30% for non-rotating protocols (**Table 2, Fig. 6**). All these
918 differences are statistically significant according to the two-sided Wilcoxon test (the paired
919 version of the test was used in this case since different *qpAdm* protocols applied to the same
920 simulation/subsampling replicate are not totally independent experiments). Temporal
921 stratification of "right" and "left" sets (i.e., the non-rotating protocol) is helpful in the absence
922 of the former type of temporal stratification, of targets and proxy sources: FDR drops from

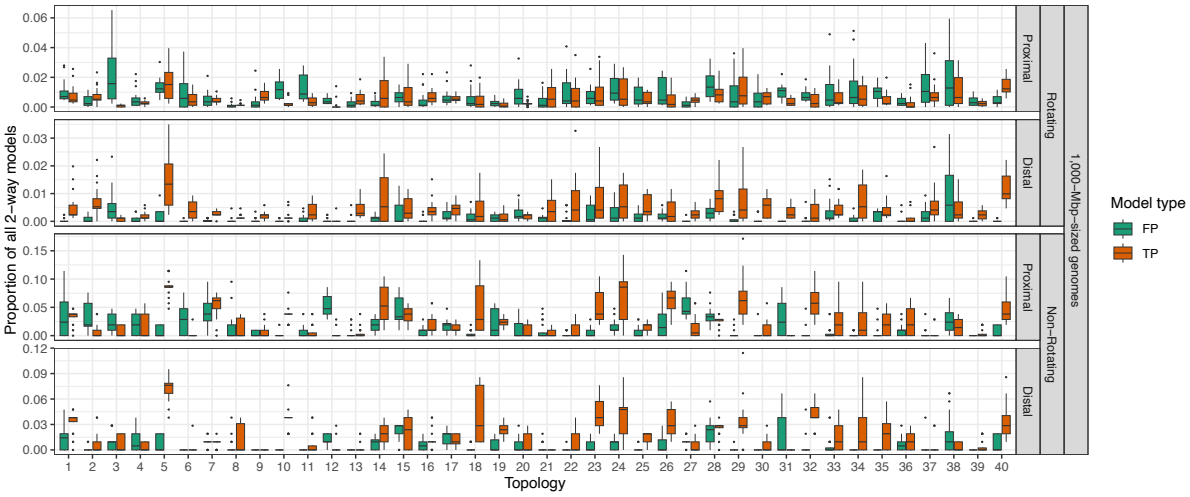
923 ~50%–60% to ~40% (these differences are also statistically significant; **Table 2, Fig. 6**).
924 However, it is not helpful (no significant difference) or even damaging to *qpAdm* performance
925 (significantly worse) when applied to distal protocols (**Table 2, Fig. 6**). This result supports the
926 conclusion by Harney et al. (2021) (also illustrated by some examples in the previous section)
927 that rotating *qpAdm* protocols should be preferred to non-rotating ones. However, according
928 to our analysis, this conclusion is conditional on strictly enforced temporal stratification of
929 targets and proxy sources since the rotating protocol without such stratification ("proximal
930 rotating") demonstrated by far the worst performance in our analysis. There is a possibility
931 that this conclusion is also conditional on the underlying history being approximated as a
932 complex graph, but not as a stepping-stone model.

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

938 The fraction of two-way admixture models that are inappropriate according to our
939 topological criteria (the fraction of negatives) in a random sample of 400 models from all
940 simulated graph-shaped genetic histories is 82.3%, which allows us to approximately estimate
941 not only FDR (51.6%–68.1%, see **Table 1**), but the FPR of the proximal rotating protocol =
942 number of false positives per simulation replicate / number of negatives per simulation
943 replicate [858 models × 40 graphs × fraction of negatives] = 0.4%–2.6% across simulation
944 parameter combinations and replicates summarized in **Table 2**. The TPR of the proximal
945 rotating protocol = number of true positives per simulation replicate / number of positives
946 per simulation replicate [858 models × 40 graphs × (1 – fraction of negatives)] = 1.1%–10.1%.
947 Here are the same statistics for the distal non-rotating protocol: the fraction of negatives in
948 a random sample of 400 distal non-rotating models from all simulated graphs, 74.3%; total
949 number of distal non-rotating two-way models across all graph topologies, 1804; FDR across
950 simulation parameter combinations and replicates from **Table 2**, 16.4%–34.4%; FPR, 0.8%–
951 3.2%; TPR, 9.7%–24.4%. Thus, although FPR of a single *qpAdm* test applied to a random graph-
952 shaped genetic history is low, due to the high proportion of negatives among all models (**Fig.**

953 **5**), the large number of models tested in high-throughput *qpAdm* screens, and the low TPR,
954 FDR becomes high, compromising historical interpretation of such screens for admixture.

955 The fraction of feasible *qpAdm* models that are false (FP) varies substantially depending on
956 topology of simulated graph-shaped genetic history (**Fig. 7**), and hence it is hard to predict if
957 a particular real genetic history is prone to generating FP signals of admixture when *qpAdm*
958 protocols are applied. Among 80 combinations of proximal *qpAdm* protocols (rotating or non-
959 rotating), simulation setups, and simulation/subsampling replicates we tested, in one
960 combination only a topology accounts for >20% of FP *qpAdm* models found across all the 40
961 simulated topologies. In contrast, in the case of distal *qpAdm* protocols, results are much
962 more uneven across topologies: for 25 of 80 "protocol/simulation setup/replicate"
963 combinations, at least one topology accounts for >20% of FP *qpAdm* models found across all
964 the 40 simulated topologies. For an illustration of the three topologies most problematic for
965 distal *qpAdm* protocols see **Fig. S5**. These differences in *qpAdm* performance across
966 topologies are probably driven at least in part by the fact that some simulated graph-shaped
967 histories have much "denser" violations of the *qpAdm* topological assumptions than others,
968 but this frequency of assumption violations was not quantified in our study since it requires
969 another layer of manual topological analysis.



970

971 **Figure 7.** Proportions of all possible two-way *qpAdm* models that are false positive (FP), or true
972 positive (TP), binned by simulated graph topology. There are 858 two-way admixture models per
973 simulated topology including 13 groups if the rotating *qpAdm* protocol is applied, and 105 models if
974 the non-rotating protocol with 6 outgroups is applied. For brevity, results are shown for 1000-Mbp-
975 sized genomes and high-quality data only. The boxplots summarize distributions of FP and TP model
976 fractions across simulation replicates.

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

978 An implicit assumption of many archaeogenetic studies relying on *qpAdm* protocols is that
979 admixture models supported by clines observed in (usually two-dimensional) spaces of
980 principal components, and/or by an *ADMIXTURE* analysis, and/or by individual *D*-, f_4- or f_3-
981 statistics are especially robust. And, *vice versa*, *qpAdm* results are often interpreted as a
982 formal test of hypotheses about admixture formulated based on PCA and/or *ADMIXTURE*
983 results. We constructed "admixture inference pipelines" composed of a *qpAdm* protocol and
984 one or two further methods to test these assumptions on simulated data. We note that all
985 patterns in our PCA or *ADMIXTURE* analyses consistent with admixture were not explored
986 with *qpAdm*. *Vice versa*, all feasible *qpAdm* models were checked by the PCA and/or
987 *ADMIXTURE* methods.

988 We considered a two-way admixture model to be supported by PCA if the target group was
989 located on a straight line between the two proxy source groups in the space of first three PCs
990 when all 13 simulated groups were co-analyzed on high-quality data. Deviation from the
991 straight line due to post-admixture genetic drift was acceptable to a certain extent (see
992 Methods) considering theoretical results by Peter (2022). Non-linear PCA clines are often
993 observed on real data (de Barros Damgaard et al. 2018, Jeong et al. 2019), and they were also
994 common among TP two-way *qpAdm* models in this study (see Methods for details and **Fig. 4**
995 and **Fig. S4** for examples). This situation is expected since many target groups in our
996 simulations represent three-way and more complex mixtures (see examples in **Fig. S4k-m**),
997 and since arrangement of populations in PC spaces is influenced not only by admixture of
998 previously isolated groups or continuous gene flows decaying with distance (2016 Duforet-
999 Frebourg and Slatkin 2016), but also by genetic drift (McVean 2009, Peter 2022). Our
1000 requirements for a model to be declared supported by PCA were more stringent than those
1001 usually applied in the literature since we considered three-dimensional PC spaces instead of
1002 two-dimensional ones (however, 2D PCAs are shown in **Fig. S4** for clarity). Also see Methods
1003 for the rules we used to judge if an admixture model is supported by an unsupervised
1004 *ADMIXTURE* analysis and see **Fig. 4** and **Fig. S4** for examples.

1005 A much more limited form of group rotation, "model competition", is used in the literature
1006 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.

two-way models supported by:	simulation scheme	max. depth of the simulation, gen.	pseudohaploid noisy data	genome size, Mbp	feasible *qpAdm* models/ all 2-way models	non-rotating *qpAdm* with temporal stratification of the R and L sets, but with no stratification of targets and proxy sources				one or two alternative sources "on the left", all others "on the right"				one or two alternative sources "on the left", all others "on the right", one at a time			
proximal non-rotating *qpAdm*						1	1	1	1	1	1	1	1	1	1	1	1
qpAdm model competition										2	2	2	2	2	2	2	2
"admixture" *f3*- statistics							2										
3D PCA								2			3		3		3		3
unsupervised *ADMIXTURE*									2			3	4			3	4
date variable, no. 1	800	no	300		201/ 4,200 (4.8%)	45.8%	0.0%	10.0%	1.0%	16.4%	6.0% / 10.4%	0.0% / 16.4%	0.0% / 6.0% / 0.0% / 10.4%	15.9%	5.0% / 10.9%	0.0% / 15.9%	0.0% / 5.0% / 0.0% / 10.9%
							45.8%	35.8%	44.8%	29.4%	4.0% / 25.4%	1.0% / 28.4%	0.0% / 4.0% / 1.0% / 24.4%	29.9%	5.0% / 24.9%	1.0% / 28.9%	0.0% / 5.0% / 1.0% / 23.9%
						54.2%	2.5%	18.4%	20.9%	19.4%	7.5% / 11.9%	10.0% / 9.5%	4.5% / 3.0% / 5.5% / 6.5%	22.4%	8.5% / 13.9%	10.4% / 11.9%	5.5% / 3.0% / 5.0% / 9.0%
							51.7%	35.8%	33.3%	34.8%	10.9% / 23.9%	10.9% / 23.9%	4.0% / 7.0% / 7.0% / 16.9%	31.8%	10.0% / 21.9%	10.4% / 21.4%	3.0% / 7.0% / 7.5% / 14.4%
					FDR	45.8%	0%	35.1%	4.5%	45.8%	44.4%	0%	0%	41.6%	37.0%	0%	0%
					FOR	0%	95.4%	66.1%	61.5%	64.2%	86.2%	81.7%	91.7%	58.7%	84.4%	80.7%	89.9%
date variable, no. 3	3000	no	300		100/ 4,200 (2.4%)	36.0%	0.0%	12.0%	5.0%	8.0%	6.0% / 2.0%	2.0% / 6.0%	2.0% / 4.0% / 0.0% / 2.0%	8.0%	6.0% / 2.0%	2.0% / 6.0%	2.0% / 4.0% / 0.0% / 2.0%
							36.0%	24.0%	31.0%	28.0%	6.0% / 22.0%	3.0% / 25.0%	3.0% / 3.0% / 0.0% / 22.0%	28.0%	6.0% / 22.0%	3.0% / 25.0%	3.0% / 3.0% / 0.0% / 22.0%
						64.0%	4.0%	23.0%	34.0%	20.0%	5.0% / 15.0%	7.0% / 13.0%	4.0% / 1.0% / 3.0% / 12.0%	18.0%	5.0% / 13.0%	7.0% / 11.0%	4.0% / 1.0% / 3.0% / 10.0%
							60.0%	41.0%	30.0%	44.0%	18.0% / 26.0%	27.0% / 17.0%	15.0% / 3.0% / 12.0% / 14.0%	46.0%	18.0% / 28.0%	27.0% / 19.0%	15.0% / 3.0% / 12.0% / 16.0%
					FDR	36.0%	0%	34.3%	12.8%	28.6%	54.5%	22.2%	33.3%	30.8%	54.5%	22.2%	33.3%
					FOR	0%	93.8%	64.1%	46.9%	68.8%	92.2%	89.1%	93.8%	71.9%	92.2%	89.1%	93.8%
at "present", no. 4	800	no	300		235/ 4,200 (5.6%)	42.1%	0.0%	22.6%	7.7%	13.6%	9.4% / 4.3%	5.1% / 8.5%	3.8% / 5.5% / 1.3% / 3.0%	11.9%	7.7% / 4.3%	4.3% / 7.7%	3.0% / 4.7% / 1.3% / 3.0%
							42.1%	19.6%	34.5%	28.5%	13.2% / 15.3%	2.6% / 26.0%	2.1% / 11.1% / 0.4% / 14.9%	30.2%	14.9% / 15.3%	3.4% / 26.8%	3.0% / 11.9% / 0.4% / 14.9%
						57.9%	2.6%	40.0%	29.8%	22.1%	15.7% / 6.4%	11.9% / 10.2%	10.2% / 5.5% / 1.7% / 4.7%	23.8%	17.0% / 6.8%	13.2% / 10.6%	11.1% / 6.0% / 2.1% / 4.7%
							55.3%	17.9%	28.1%	35.7%	24.3% / 11.5%	17.9% / 17.9%	14.5% / 9.8% / 3.4% / 8.1%	34.0%	23.0% / 11.1%	16.6% / 17.4%	13.6% / 9.4% / 3.0% / 8.1%
					FDR	42.1%	0%	36.1%	20.5%	38.1%	37.3%	30.0%	27.3%	33.3%	31.0%	24.4%	21.2%
					FOR	0%	95.6%	30.9%	48.5%	61.8%	72.8%	79.4%	82.4%	58.8%	70.6%	77.2%	80.9%

1015 **Table 3.** Assessing FDR of the proximal non-rotating *qpAdm* protocol combined with model
1016 competition (Narasimhan et al. 2019). For each simulation setup, the analysis relies on simulation
1017 replicate no. 1 only. For constructing pipelines resembling those common in the archaeogenetic
1018 literature, we used five methods: proximal non-rotating *qpAdm*, *qpAdm* model competition (two
1019 alternative protocols), "admixture" f_3-statistics, 3D PCA with all individuals co-analyzed, and
1020 unsupervised *ADMIXTURE* with all individuals co-analyzed. For declaring a positive result, support of
1021 an admixture model by all methods in the pipeline was required, hence the order of methods is not
1022 important except for the first method, which was the proximal non-rotating *qpAdm* in all cases. In
1023 each column, methods comprising a pipeline are color-coded and numbered by their order. All
1024 feasible two-way *qpAdm* models emerging as outcomes of the proximal non-rotating protocol were
1025 classified into FP and TP (highlighted in red and green in the leftmost column, respectively). The other
1026 columns are structured like bifurcating trees: FP *qpAdm* models supported by method no. 2; FP
1027 *qpAdm* models not supported by method no. 2; TP *qpAdm* models supported by method no. 2; TP
1028 *qpAdm* models not supported by method no. 2. The same principle is used for representing results of
1029 more complex pipelines. All model counts are normalized by the number of feasible *qpAdm* models
1030 (FP + TP) that are outcomes of the first method. Percentages of models supported/not supported by
1031 the last method in the pipeline are highlighted in green and red, respectively. FDR values are shown
1032 for these different pipelines. Fractions of TP *qpAdm* models that are pruned out by progressively more
1033 stringent support requirements are also shown (false omission rate or FOR).

1034

1035 In many publications (e.g., Haak et al. 2015, Mathieson et al. 2015, Antonio et al. 2019,
1036 Mathieson et al. 2018, Prendergast et al. 2019, Marcus et al. 2020, Papac et al. 2021, Wang
1037 et al. 2021, Yaka et al. 2021, Changmai et al. 2022a, 2022b, Patterson et al. 2022) this first
1038 non-rotating step remains the only *qpAdm* protocol used (in its distal or proximal forms). In
1039 a model competition protocol, subsequent analysis is focused on targets for whom two or
1040 more alternative *qpAdm* models emerge as feasible at the first step. For each target,
1041 alternative proxy sources are pooled and rotated between the "left" and "right" sets, testing
1042 only the models that emerged as feasible at the first step and applying a composite feasibility
1043 criterion (e.g., p-value > 0.01, estimated admixture proportions ± 2 SE are within 0 and 1).

1044 Rotation of alternative proxy sources can be performed in various ways: "whatever is not on
1045 the left is on the right" (Brielle et al. 2023), or placing alternative sources in the "right" set
1046 one by one (Calhoff et al. 2021, Maróti et al. 2022, Brielle et al. 2023). In the latter case several
1047 "right" sets are tested for each model, and the model is considered supported by the model
1048 competition protocol only if it is not rejected under any of these "right" sets (Maróti et al.
1049 2022). The reasoning behind this protocol is as follows: model rejection due to violations of
1050 the topological assumptions of *qpAdm* is not expected for a model composed of sources very
1051 close to the true ones since in this case branches private to the proxy sources are short, and

1052 it is unlikely that gene flows to or from the "right" population set happened on these short
1053 branches. Models composed of sources closely related to the true ones are also not expected
1054 to be rejected when more distant proxy sources are placed in the "right" set (Harney et al.
1055 2021; see also a case study in **Fig. S4h**).

1056 For reasons detailed in the Discussion section, we explored the *qpAdm* model competition
1057 protocol and multi-method admixture inference pipelines on one replicate per simulation
1058 setup, and three (**Table 3**) or four (**Table 1**) simulation setups were involved in this analysis.
1059 The two alternative model competition protocols described above were applied to targets
1060 for whom more than one model was feasible given a fixed "right" set composed of six groups.
1061 If only one model was feasible for a target, such a model was evaluated as passing model
1062 competition (such models accounted only for 7%–11% of models feasible at the first step).
1063 The model competition protocols failed to improve *qpAdm* performance: FDR ranged from
1064 29% to 46% (as compared to 36%–46% prior to the model competition step), and both model
1065 competition protocols demonstrated very similar results (**Table 3**). FOR of the model
1066 competition protocols varied from 59% to 72%. FDR also remained high for models supported
1067 by proximal non-rotating *qpAdm* & PCA or by proximal non-rotating *qpAdm* & model
1068 competition & PCA (**Table 3**).

1069 Considering all the simulation setups and replicates shown in **Table 2**, there were only 1,591
1070 instances when a two-way admixture model was supported by both the proximal rotating
1071 and proximal non-rotating protocols on the same simulated data (this corresponds to
1072 1,591×2 = 3,182, or 9.6%, of 33,072 unique combinations of *qpAdm* protocols, simulated
1073 datasets, and feasible two-way admixture models that we found; see **Suppl. dataset 1**). In
1074 contrast, there were 5,844 (17.7%) and 24,046 (72.7%) instances when a model was
1075 supported exclusively by the proximal non-rotating and rotating protocols, respectively.
1076 Notably, FP models supported by the proximal non-rotating *qpAdm* protocol largely lacked
1077 support by an unsupervised *ADMIXTURE* analysis (**Table 3**), in contrast to outcomes of the
1078 proximal rotating protocol (**Table S1**). FDR of a pipeline composed of these two methods
1079 ranged from 5% to 21% across three simulation setups tested (**Table 3**). Adding a model
1080 competition step to this pipeline increased both FDR and FOR in 4 of 6 cases; and, in general,

1081 the proximal non-rotating *qpAdm* protocol combined with *ADMIXTURE* is the best-
1082 performing protocol in this analysis (**Table 3**) according to FDR and FOR.

1083 The fact that our *ADMIXTURE* analysis supports a large fraction of FP two-way mixture models
1084 emerging as outcomes of the proximal rotating *qpAdm* protocol reflects known problems in
1085 modelling with *ADMIXTURE* very ancient individuals in the context of modern populations.
1086 These individuals are often modelled (Raghavan et al. 2014, Haak et al. 2015, Moreno-Mayar
1087 et al. 2018a) as complex mixtures of ancestry components typical for modern populations,
1088 which is obviously an artefact. Sampling dates for unique targets from FP models supported
1089 by both proximal rotating *qpAdm* and *ADMIXTURE* ranged from the present to 665
1090 generations in the past (median = 406 generations), while the median sampling date for
1091 unique targets from both FP and TP models supported by proximal rotating *qpAdm* was 215
1092 generations in the past (for comparison across simulation setups, all the dates were rescaled
1093 to a maximum simulation depth of 800 generations). The proximal non-rotating protocol by
1094 design did not consider the oldest groups as targets for *qpAdm* and *ADMIXTURE* analyses,
1095 thus largely avoiding this problem (sampling dates for unique targets from FP models
1096 supported by both proximal non-rotating *qpAdm* and *ADMIXTURE* ranged from the present
1097 to 366 generations in the past; median = 44 generations).

1098 Above we have discussed multi-method pipelines based on the proximal non-rotating *qpAdm*
1099 protocol. Combining distal *qpAdm* protocols with PCA allows to reduce FDR of both rotating
1100 and non-rotating protocols further, to 10%–24%, and distal *qpAdm* protocols combined with
1101 an unsupervised *ADMIXTURE* analysis demonstrated even better FDR values ca. 0%–8%
1102 (**Table 1**). If target and proxy source populations are sampled at approximately the same time
1103 (such as those from our simulation setup no. 4 and from the proximal analysis in Narasimhan
1104 et al. 2019, **Fig. 2c**) applying this approach is impossible. However, if our simulations with all
1105 branches extended to the present are treated in the same way as their topological
1106 counterparts with date-variable sampling, performance gains (decrease in FDR) of temporal
1107 stratification of admixture models are similar to those mentioned above (**Table 1**). In this case
1108 the temporal stratification procedure retains models with the latest admixture event in
1109 target's history that is more recent than (or as recent as) the latest admixture events in proxy
1110 sources' history. However, in the case of simulation setup no. 4 performance gains of the
1111 "*qpAdm* + PCA" and "*qpAdm* + *ADMIXTURE*" method combinations were moderate (**Table 1**).

1112

1113 ## Discussion

1114 In this study we explored performance of various *qpAdm* protocols on a collection of
1115 simulated genetic histories in the shape of complex admixture graphs of random topology
1116 and the same complexity, where admixture history of target groups may vary from the
1117 simplest (no admixture) to very complex. It is because of this research design and other
1118 limitations discussed below that our study is focused mostly on one performance metric: false
1119 discovery rate or FDR. In simple terms, we focused our analysis only on models of a chosen
1120 complexity class (two-way models) supported by a *qpAdm* protocol (feasible models),
1121 classified them manually into false and true positives according to a set of topological rules,
1122 and subjected them to further screening by PCA and/or *ADMIXTURE* methods. We did not
1123 attempt to classify rejections of two-way models by *qpAdm* or other methods into false
1124 rejections due to violations of the topological assumptions of *qpAdm* (**Fig. 1**) and true
1125 rejections when the true admixture history of the target does not fit a two-way model. This
1126 problem was deliberately left out since in the literature more attention is paid to
1127 interpretation of "fitting" ("feasible" or "positive") than of rejected *qpAdm* models.

1128 Another limitation of our study is that we had to use idealized versions of *qpAdm*, PCA, and
1129 *ADMIXTURE* protocols, while in the archaeogenetic literature manual adjustment of
1130 analytical protocols is common: protocols often vary from one target group to another (see,
1131 e.g., Lazaridis et al. 2016, Zhang et al. 2021, Brielle et al. 2023, Lee et al. 2023) and from study
1132 to study. These extensive details are very hard to formalize and reproduce. In the case of
1133 *qpAdm* protocols, certain groups of populations may be placed exclusively in "right" or in
1134 "left" sets, with the rest rotated between these sets, and relative sizes and compositions of
1135 these three groups vary from study to study: in the case of model competition protocols, this
1136 rotated subset is small, and rotation may be restricted to a particular model complexity class,
1137 but in other cases it may encompass all or nearly all populations analyzed (see, e.g.,
1138 Narasimhan et al. 2019, Librado et al. 2021, Bergström et al. 2022, Lazaridis et al. 2022,
1139 Oliveira et al. 2022, Taylor et al. 2023). Reproducing all aspects of PCA and *ADMIXTURE*
1140 protocols used in the literature is also hardly possible on simulated data. For instance, PCs in
1141 archaeogenetic studies are usually calculated based on present-day populations, and ancient

1142 individuals are projected on the resulting PCs (e.g., Haak et al. 2015, Mathieson et al. 2018,
1143 Narasimhan et al. 2019, Furtwängler et al. 2020, Marcus et al. 2020, Lazaridis et al. 2022). In
1144 contrast, in our study all simulated individuals were co-analyzed for calculating PCs since this
1145 analysis was based on high-quality data only. Unsupervised *ADMIXTURE* analyses in the
1146 literature are usually performed on worldwide or continent-wide panels of populations that
1147 often overlap just partially with population sets used for *qpAdm* analyses (see, for instance,
1148 Rasmussen et al. 2010, Haak et al. 2015, Harney et al. 2018, Moreno-Mayar et al. 2018a,
1149 Zhang et al. 2021, Changmai et al. 2022a, Brielle et al. 2023), while in our study identical
1150 population sets were used for *qpAdm*, PCA, and *ADMIXTURE* analyses.

1151 Another important caveat is that complexity of genetic history in the region and period of
1152 interest often remains unknown and it is difficult to judge if a particular admixture graph
1153 complexity is adequate for simulating the real history. However, we have not explored *qpAdm*
1154 performance over a range of simulated admixture graph complexities, over a range of model
1155 feasibility criteria (except for those in **Table S2**), for models more complex than two-way, and
1156 have estimated FDR and FOR instead of false positive and false negative rates due to an
1157 important technical limitation: the process of model classification into true and false ones
1158 cannot be fully automated since it requires careful interpretation of the simulated topology
1159 and simulated admixture proportions (this is illustrated by the case studies in **Fig. S4**). For
1160 similar reasons, some comparisons of method performance in this study, such as *qpAdm* vs.
1161 "*qpAdm* combined with *ADMIXTURE*", are qualitative rather than quantitative: we applied
1162 the PCA and *ADMIXTURE* methods to one simulation replicate only per simulation setup since
1163 automated classifiers of admixture models into positive and negative ones based on 3D PCA
1164 and *ADMIXTURE* results were not available. Despite these limitations, our simulations
1165 reproduce the most important aspects of typical *qpAdm* protocols.

1166 Finally, a notable limitation of our analysis is that isolation by distance (stepping-stone)
1167 models were not considered. Strictly speaking, it remains unknown if admixture history of
1168 humans, wolves (Bergström et al. 2022), and horses (Librado et al. 2021, Taylor et al. 2023),
1169 to cite some recent examples from the archaeogenetic literature, is approximated reasonably
1170 well by complex admixture graphs composed of isolated populations and episodes of
1171 admixture, or rather by two-dimensional stepping-stone models (with additional features
1172 such as non-symmetric gene flows, extinction and repopulation), or a combination thereof.

1173 Performance of *qpAdm* on a very simple one-dimensional stepping-stone history was
1174 explored just briefly by Harney et al. (2021) and was deemed poor.

1175 We demonstrated that application of the proximal rotating *qpAdm* protocol that can be
1176 summarized as "whatever is not on the right is on the left" without any temporal stratification
1177 of the "right" and "left" sets and of proxy sources and targets carries a risk of an FDR above
1178 50% or 60%. Adding further levels of support (considering only models supported by PCA
1179 and/or an *ADMIXTURE* analysis) does not help to decrease FDR drastically in this case (**Table**
1180 **S1, Suppl. text 1**).

1181 The proximal rotating protocol is an extreme example of *qpAdm* protocols that is rarely
1182 encountered in the archaeogenetic literature (Calhoff et al. 2021, Oliveira et al. 2022) but
1183 serves as a reference point in our analysis. Other protocols such as distal rotating (e.g.,
1184 Narasimhan et al. 2019, Librado et al. 2021, Allentoft et al. 2022, Bergström et al. 2022,
1185 Lazaridis et al. 2022, Taylor et al. 2023), distal non-rotating (e.g., Haak et al. 2015, Mathieson
1186 et al. 2015, Lazaridis et al. 2016, Antonio et al. 2019, Marcus et al. 2020, Yang et al. 2020,
1187 Papac et al. 2021, Yaka et al. 2021, Patterson et al. 2022), and proximal model competition
1188 (e.g., Narasimhan et al. 2019, Calhoff et al. 2021, Zhang et al. 2021, Maróti et al. 2022, Brielle
1189 et al. 2023, Lee et al. 2023) are often used in practice, and FDR of these three classes of
1190 protocols on our simulated data ranged from 12% to 46% across simulation parameter
1191 combinations and replicates (**Tables 1 and 3**). These FDR for best-performing standalone
1192 *qpAdm* protocols are high but should not be over-interpreted since they are expected to
1193 depend on the complexity of simulated histories and on the amount of data (**Table 2**), which
1194 also depends on the time depth of simulated histories (**Fig. S3**). Only one graph complexity
1195 level was tested, that is 13 groups and 10 admixture events; and only one time depth, 800
1196 generations, was tested in a high-throughput way (**Table 1**). Thus, it is hard to extrapolate
1197 this aspect of our results to real analyses and predict FDR levels on real data.

1198 Temporal stratification tested in this study and practiced in the literature is of two sorts: 1)
1199 most or all populations in the "right" set are sampled deeper in the past than those in the
1200 "left" set (non-rotating protocols); 2) a target group post-dates (or is as old as) all its proxy
1201 sources (distal protocols). We showed that both temporal stratification approaches helped

1202 to decrease FDR of *qpAdm* admixture screens significantly, and the latter approach
1203 demonstrated the best FDR among standalone *qpAdm* protocols (**Table 2**).

1204 Although restricting analyses to distal models is often *necessary* for reducing FDR below an
1205 arbitrary threshold at 10%, it is not *sufficient* for reaching this objective (**Table 1**) given the
1206 complexity of our simulated admixture graph-shaped histories and the amounts of data we
1207 generated. Respecting this threshold, only the following admixture screening protocols
1208 demonstrated acceptable performance (we did not consider protocols demonstrating FOR
1209 above 90% as useful in practice):

1210 1) proximal non-rotating *qpAdm* with a requirement that admixture models are supported by
1211 both *qpAdm* and an unsupervised *ADMIXTURE* analysis (**Tables 1 and 3**), under simulation
1212 setups no. 1 and 2 (groups sampled at different dates in the past, maximal simulated history
1213 depth = 800 generations, 300-Mbp-sized or 1000-Mbp-sized genomes simulated);

1214 2) distal non-rotating or rotating *qpAdm* with a requirement that admixture models are
1215 supported by both *qpAdm* and an unsupervised *ADMIXTURE* analysis (**Table 1**), under
1216 simulation setups no. 1, 2, and 3 (groups sampled at different dates in the past, maximal
1217 simulated history depth = 800 or 3000 generations, 300-Mbp-sized or 1000-Mbp-sized
1218 genomes simulated).

1219 FDR of these protocols was 0% – 8% (**Table 1**). In contrast, adding a model competition step
1220 to the proximal non-rotating *qpAdm* protocol did not help to reduce FDR below 10%. The
1221 performance of this type of protocols is explored in detail in **Table 3**.

1222

1223 ***Best practices***

1224 As outcomes of this study, we make the following suggestions for improving robustness of
1225 admixture inference in archaeogenetics:

1226 1. Our results suggest that temporal stratification of targets and proxy sources is a very
1227 efficient way of reducing FDR of *qpAdm* protocols (**Tables 1 and 2, Fig. 6**). The distal
1228 rotating and non-rotating protocols invariably demonstrated FDR significantly lower than

1229 those of the proximal non-rotating and rotating protocols (**Table 2**). Although the
1230 proximal model competition protocol (Narasimhan et al. 2019, Calhoff et al. 2021, Zhang
1231 et al. 2021, Maróti et al. 2022, Brielle et al. 2023, Lee et al. 2023) was not tested on
1232 multiple simulation or subsampling replicates (**Table 3**), we note that it demonstrated
1233 FDR values higher than those of the distal non-rotating protocol (**Table 1**). These results
1234 and our case studies (**Fig. 4, Fig. S4a-e**) imply that *qpAdm* protocols where all populations
1235 are sampled at present (similar to our setup no. 4; see also Jeong et al. 2019, Changmai
1236 et al. 2022a) or where present-day groups are used as proxy ancestry sources for ancient
1237 groups (e.g., Mathieson et al. 2015, van de Loosdrecht et al. 2018, Narasimhan et al. 2019,
1238 Prendergast et al. 2019, Shinde et al. 2019, Wang et al. 2020, Wang et al. 2021, Changmai
1239 et al. 2022b, Calhoff et al. 2021, Oliveira et al. 2022) are less reliable than those where
1240 target groups are not allowed to pre-date their proxy sources. While the proximal non-
1241 rotating *qpAdm* protocol demonstrated FDR significantly lower than that of the proximal
1242 rotating protocol (**Table 2**), the distal rotating protocol was in terms of FDR as good as the
1243 distal non-rotating protocol (on low-quality data) or significantly better (on high-quality
1244 data, **Table 2**).
1245 2. Another way of radically improving FDR of *qpAdm* protocols is combining *qpAdm* with an
1246 unsupervised *ADMIXTURE* analysis. These two approaches should possibly be combined
1247 for optimal performance (**Table 1**).
1248 3. Adding 3.3 times more data led to a small but significant decrease in FDR only in the case
1249 of high-quality diploid data, but not in the case of pseudo-haploid data with high missing
1250 rates (**Table 2**). This observation deserves further investigation.
1251 4. It is safest to use the *qpAdm* method in controlled conditions, when relationships among
1252 populations are understood well enough to exclude violations of the topological
1253 assumptions, when radiocarbon or context dates of ancient populations are reliable and
1254 allow accurate temporal stratification, or when sets of potential proxy sources are well-
1255 constrained based on archaeological or historical scholarship: see, for instance, the
1256 earliest publications where the *qpAdm* method was employed (Haak et al. 2015,
1257 Mathieson et al. 2015) and some recent studies (e.g., Marcus et al. 2020, Papac et al.
1258 2021, Yaka et al. 2021, Changmai et al. 2022a, 2022b, Patterson et al. 2022). Obviously,
1259 the amount of new information that the *qpAdm* method provides in these conditions is
1260 limited. However, considering that it was possible to reach FDR levels as low as 0% to 8%

1261 on our simulated data, we do not recommend avoiding *qpAdm*-based high-throughput
1262 admixture screens altogether.

1263 5. Summing up all the results above, for reducing FDR of *qpAdm* admixture screens to nearly
1264 0% we suggest using large SNP sets with low missing data rates, using the rotating *qpAdm*
1265 protocol with a strictly enforced rule that targets do not pre-date their proxy sources, and
1266 performing an unsupervised *ADMIXTURE* analysis to verify feasible *qpAdm* models.

1267 6. Our study has multiple limitations and caveats discussed above, mostly related to
1268 difficulties in simulating all the details of published *qpAdm*, PCA, and *ADMIXTURE*
1269 protocols, to uncertainties about the level of admixture graph complexity that is adequate
1270 for simulating real population histories (all our simulations were of the same complexity:
1271 13 groups and 10 pulse-like admixture events), to uncertainty about applicability of
1272 admixture graph models in general since stepping-stone models may constitute a more
1273 adequate simplification if human population history, to difficulties in interpreting
1274 topologies of random admixture graphs in an automated way for classifying even simple
1275 admixture models into true and false ones, and to difficulties in interpreting 3D PCA and
1276 *ADMIXTURE* results in an automated way. Nevertheless, our results surpass in scale
1277 previous simulation studies of *qpAdm* protocols (Lazaridis et al. 2017, Ning et al. 2020,
1278 Harney et al. 2021) by several orders of magnitude and may serve as a guide for users of
1279 high-throughput *qpAdm* protocols.

1280 7. Feasible *qpAdm* models are sometimes ranked by *p*-values, with a model having the
1281 highest *p*-value highlighted as the most plausible one (see, for instance, Lazaridis et al.
1282 2022, van de Loosdrecht et al. 2018, Oliveira et al. 2022, Taylor et al. 2023). *qpWave p*-
1283 values for pairs of individuals were also used in lieu of genetic distances in the former
1284 study (Lazaridis et al. 2022). Of 1,201 instances when both false and true feasible *qpAdm*
1285 models were found for the same target group on the same data (all simulation setups,
1286 simulation/subsampling replicates, and *qpAdm* protocols), a model having the highest *p*-
1287 value was an FP in 463 (38.6%) cases, and the difference in maximal *p*-values between the
1288 TP and FP model classes was significant according to the paired two-sided Wilcoxon test
1289 (TP > FP, *p*-value = 2.2×10^{-16}). Thus, our limited analysis suggests that the approach of
1290 ranking *qpAdm* models by *p*-values is justified (see also related results in **Fig. S6** and **Table**
1291 **S2**), but it generates noisy results.

1292 8. f_3-statistic is a simple method for proving that a population is admixed, and it
1293 demonstrated FDR values much lower (6%, see **Suppl. text 1**) than those of standalone
1294 *qpAdm* protocols, but f_3-statistics are applicable only to recent admixture events and/or
1295 populations of large effective size since post-admixture drift on the target lineage
1296 obscures the signal (Patterson et al. 2012, Peter 2022). Moreover, calculating f_3-statistics
1297 for a target composed of a single pseudo-haploid individual is impossible since a
1298 heterozygosity estimate is required (Maier et al. 2023), and such singleton groups are
1299 common in archaeogenetic studies. Researchers should also be aware that f_3-statistics
1300 are defined on unrooted trees, and that may lead to rare but strong false signals of
1301 admixture (**Fig. S4f**).

1302 The risks of interpreting *qpAdm* results for proximal models discussed above are apparently
1303 due to topological properties of the histories in the form of admixture graphs that we
1304 simulated: complexity of genetic history of a sampled population randomly picked from these
1305 graphs tends to grow with time from the root, as well as the chance to encounter violations
1306 of the topological assumptions of *qpAdm* (**Fig. 1**) grows with the length of the branch between
1307 a true ancestry source and a sampled proxy source or between a target group when it is
1308 formed by admixture and when it is sampled. Although it is very often assumed that real
1309 genetic history of humans and many other species is reasonably approximated as complex
1310 admixture graphs with these properties, this assumption remains unproven.

1311

1312 Methods

1313 *Simulating random admixture graphs with msprime v.1.1.1*

1314 For simulating genetic data, we used *msprime v.1.1.1* which allows accurate simulation of
1315 recombination and of multi-chromosome diploid genomes relying on the Wright-Fisher
1316 model (Nelson et al. 2020, Baumdicker et al. 2022). We simulated three or ten diploid
1317 chromosomes (each 100 Mbp long) by specifying a flat recombination rate (2×10^{-8} per nt per
1318 generation) along the chromosome and a much higher rate at the chromosome boundaries
1319 ($\log_e 2$ or ~0.693 per nt per generation, see
1320 https://tskit.dev/msprime/docs/stable/ancestry.html#multiple-chromosomes). A flat

1321 mutation rate, 1.25×10^{-8} per nt per generation (Scally & Durbin 2012), and the binary
1322 mutation model were used. To maintain the correct correlation between chromosomes, the
1323 discrete time Wright-Fischer model was used for 25 generations into the past, and deeper in
1324 the past the standard coalescent simulation algorithm was used (as recommended by Nelson
1325 et al. 2020).

1326 Genetic histories in the form of admixture graphs of random topology including 13
1327 populations and 10 pulse-like admixture events were generated using the
1328 *random_admixturegraph* and *random_sim* functions from the *ADMIXTOOLS 2* package
1329 (https://uqrmaie1.github.io/admixtools/reference/random_sim.html), which produced
1330 scripts for running the *msprime v.1.1.1* simulator. Demographic events were separated by
1331 date intervals ranging randomly between 20 and 120 generations, with an upper bound on
1332 the graph depth at 800 generations (or ca. 23,000 years in the case of humans). In another
1333 set of simulations, all the dates were scaled up 3.75 times, with an upper bound on the graph
1334 depth at 3,000 generations (or 87,000 years in the case of humans). To be more precise,
1335 demographic events were not placed in time entirely randomly, but were tied to one or few
1336 other events of the same "topological depth" within the graph, as illustrated by ten examples
1337 of simulated topologies in **Fig. S2**. The same principle was applied to sampling dates, which
1338 were tied to other demographic events such as divergence and admixture of other
1339 populations. These restrictions were used to ensure topological consistency of random
1340 graphs.

1341 Ten diploid individuals with no missing data were sampled from each population at "leaves"
1342 of the graph. Effective population sizes were constant along each edge and were picked
1343 randomly from the range of 1,000 – 10,000 diploid individuals. Admixture proportions for all
1344 admixture events varied randomly between 10% and 50%. This setup generates groups
1345 sampled at widely different dates in the past or, in other words, located at various genetic
1346 distances from the root. Alternatively, all terminal branches were extended to the "present"
1347 of the simulation and sampled at "present", keeping their respective effective population
1348 sizes and topological relationships unchanged. Thus, another set of simulations was
1349 generated for the same topologies, where groups were more drifted with respect to each
1350 other (see F_{ST} distributions in **Fig. S1**).

1351 In summary, four sets of independent simulations ("simulation setups") differing by the
1352 amount of data generated and by population divergence metrics were performed for a set of
1353 40 random admixture graph topologies (see an overview in **Fig. 3**):

1354 1) three 100-Mbp-sized chromosomes; groups sampled at different points in time; maximal
1355 simulated history depth at 800 generations (10 simulation replicates, median number of
1356 polymorphic sites = 669,655, see **Fig. S3**);

1357 2) ten 100-Mbp-sized chromosomes; groups sampled at different points in time; maximal
1358 simulated history depth at 800 generations (10 simulation replicates, median number of
1359 polymorphic sites = 2,229,459);

1360 3) three 100-Mbp-sized chromosomes; groups sampled at different points in time; maximal
1361 simulated history depth at 3,000 generations (one simulation replicate, median number of
1362 polymorphic sites = 1,074,336);

1363 4) three 100-Mbp-sized chromosomes; all terminal branches extended to the "present" of the
1364 simulation and sampled at that point; maximal simulated history depth at 800 generations
1365 (one simulation replicate, median number of polymorphic sites = 838,297).

1366 To create more realistic datasets, we performed randomised subsampling of polymorphic
1367 sites and individuals (replicates no. 1 of the first and second simulation setups were used for
1368 this, see the list above). First, we randomly sampled alleles at heterozygous sites, creating
1369 pseudo-haploid data. Then we introduced missing data by randomly selecting a missing rate
1370 between 5% and 95%, followed by randomly selecting sites according to the missing rate. This
1371 site subsampling was repeated for each individual independently. Lastly, we randomly
1372 sampled *n* (from 1 to 10) individuals from each population independently. The subsampling
1373 procedure described above was conditioned on the number of sites polymorphic in the set
1374 of 13 simulated populations and was repeated until a subsampling replicate with more than
1375 20,000 (for 300-Mbp-sized genomes) or 66,000 such sites (for 1000-Mbp-sized genomes) was
1376 obtained. We generated 10 independent subsampled replicates for each topology and
1377 simulation setup (800 replicates in total).

1378 Polymorphism data in the *EIGENSTRAT* format were generated from the tree sequences using
1379 the *TreeSequence.genotype_matrix* function (https://tskit.dev/tskit/docs/stable/python-
1380 api.html#tskit.TreeSequence.genotype_matrix) and used for all subsequent analyses (*f*-
1381 statistics and *qpAdm*, PCA, *ADMIXTURE*).

1382 For all the work on *f*-statistics and *qpAdm*, the *ADMIXTOOLS 2* software package (Maier et al.
1383 2023) was used. For diploid SNP sets without missing data, we first calculated all possible f_2-
1384 statistics for 4-Mbp-sized genome blocks (with the "*maxmiss=0*",
1385 "*adjust_pseudohaploid=TRUE*", and "*minac2=FALSE*" settings) and then used them for
1386 calculating f_3- and f_4-statistics as linear combinations of f_2-statistics and for testing *qpAdm*
1387 models using the *qpadm* function in *ADMIXTOOLS 2*
1388 (https://uqrmaie1.github.io/admixtools/) under default settings. Inferred admixture
1389 proportions were not constrained between 0 and 1. For pseudo-haploid SNP sets with missing
1390 data and uneven group sizes, the *qpadm* function was applied directly to genotype files, with
1391 the "*allsnps=TRUE*" setting. In other words, f_4-statistics utilized by *qpAdm* and f_3-statistics
1392 were calculated off the genotype files without intermediate f_2-statistics, and removal of
1393 missing data was done for each population quadruplet or triplet separately. This setup is
1394 often used in the literature in the presence of missing data (e.g., Harney et al. 2018, Harney
1395 et al. 2019, Narasimhan et al. 2019, Lazaridis et al. 2022).

1396

1397 ***Rotating qpAdm protocols***

1398 *QpWave* tests were performed on sets of 13 groups divided randomly into 2 "left" and 11
1399 "right" groups, testing all possible bisections of this form. *QpAdm* was applied to the same
1400 sets of 13 groups divided randomly into 3 "left" and 10 "right" groups, testing all possible
1401 bisections of this form for all possible target groups in "left" sets (858 such models per
1402 simulated history). This proximal rotating protocol was applied to all simulation setups.
1403 Subsequent work was focused only on feasible *qpAdm* models defined as follows: 1) *p*-values
1404 calculated by *qpWave* for one-way models "target = proxy source$_1$", "target = proxy source$_2$",
1405 and "proxy source$_1$ = proxy source$_2$" are all below 0.01; 2) in the case of the two-way model
1406 "target = proxy source$_1$ + proxy source$_2$", estimated admixture proportions ± 2 standard

1407 errors are between 0 and 1; 3) the *p*-value calculated by *qpAdm* for the two-way model ≥
1408 0.01.

1409 For exploring performance of the distal rotating protocol, feasible two-way *qpAdm* models
1410 were simply filtered according to sampling dates of target groups and proxy sources. If target
1411 group's sampling date was equal to or smaller than sampling dates of both proxy sources,
1412 such a model was considered distal.

1413

1414 ***Non-rotating and model competition qpAdm protocols***

1415 In the non-rotating protocol, for each simulated admixture graph six oldest sampled groups
1416 were selected as a fixed "right" set (ties in sampling dates were resolved in alphabetical order;
1417 these "right" sets remained unchanged for a given topology across all independent
1418 simulations), and for the remaining seven groups all possible one-way and two-way
1419 admixture models were tested (105 models), applying the same composite feasibility
1420 criterion that was used above for the rotating protocol. This is the proximal non-rotating
1421 protocol, and alternatively we focused on distal admixture models only (distal non-rotating
1422 protocol).

1423 In the proximal model competition protocol, subsequent analysis was focused on targets for
1424 whom two or more alternative *qpAdm* models emerged as feasible at the first step. For each
1425 target, alternative proxy sources were pooled and rotated between the "left" and "right" sets,
1426 testing only the models that emerged as feasible at the first step and applying the composite
1427 feasibility criterion (*p*-value ≥ 0.01, estimated admixture proportions ± 2 SE are between 0
1428 and 1). Rotation of alternative proxy sources was performed in two alternative ways:
1429 "whatever is not on the left is on the right", or placement of alternative sources in the "right"
1430 set one by one. In the latter case several "right" sets were tested for each model, and the
1431 model was considered supported by the model competition protocol only if it was not
1432 rejected under any of these "right" sets (the latter protocol follows Maróti et al. 2022). If only
1433 one model was feasible for a target, such a model was evaluated as passing the model
1434 competition procedure. A distal model competition protocol was not tested in this study.

1435 For testing statistical significance of differences in FDR between *qpAdm* protocols, the
1436 following approach was used. FDR was calculated either on low-quality data for 10 random
1437 site/individual subsampling replicates derived from simulation replicate no. 1 (simulation
1438 setups no. 1 and 2) or on high-quality data for 10 independent simulation replicates
1439 (simulation setups no. 1 and 2). Comparisons of four *qpAdm* protocols (rotating and non-
1440 rotating, proximal and distal) were performed independently on these four sets of replicates,
1441 using the two-sided paired Wilcoxon test (**Table 2**). Comparisons of the same *qpAdm*
1442 protocols on lower and higher amounts of data (300-Mbp-sized vs. 1,000-Mbp-sized
1443 simulated genomes) were performed using the two-sided non-paired Wilcoxon test since
1444 simulation replicates were independent unlike alternative *qpAdm* protocols applied to the
1445 same data (**Table 2**).

1446

1447 ***Classifying two-way admixture models into false and true positives***

1448 Since the simulated admixture graph topologies were complex and random, target groups
1449 modelled with *qpAdm* had very complex admixture history in some cases, being a part of
1450 gene flow networks. In this context it is hard to draw a strict boundary between true and false
1451 admixture models composed of a target and only two proxy sources. Two-way admixture
1452 models were considered false only if at least one of the following criteria was satisfied
1453 (considering only graph topologies and admixture proportions):

1454 1. The target and at least one of the proxy sources are simulated as strictly cladal (**Fig. 4, Fig.**
1455 **S4a**). In this case the target may either be unadmixed, or it may have experienced gene
1456 flows earlier in its history that do not break its cladality with one of the proxy sources.
1457 2. A proxy source lineage is a recipient of gene flow from the target lineage (after the last
1458 admixture event in target's history), possibly mediated by other lineages (**Fig. 4, Fig. S4a,**
1459 **b**). In other words, the incorrect proxy source is a descendant of the target lineage, i.e.,
1460 the expected gene flow direction is reversed.
1461 3. A proxy source does not represent any true source. In other words, it is symmetrically
1462 related to all true sources of ancestry in the target (**Fig. S4c, d**). Alternatively, both proxy

1463 sources represent the same true source, and are symmetrically related to all the other
1464 true sources (**Fig. S4e**).
1465 4. A proxy source shares genetic drift with the corresponding true source that is not shared
1466 by the second proxy source (and the same is true for the other proxy source and another
1467 true source, i.e., condition no. 3 above is not satisfied), however less than 40% of its
1468 ancestry is derived from the true source (**Fig. S4a, d**).
1469 5. The target gets a gene flow from a deep-branching source not represented by any
1470 sampled population, and an inappropriate proxy source is included in the fitting model
1471 (**Fig. S4f, g**).

1472 We illustrate these topological rules with eight case studies of FP and feasible *qpAdm* models
1473 in **Fig. 4** and **Fig. S4a-g**. Two-way models for targets whose population history is best
1474 approximated with three-way and more complex models were considered as TP if they
1475 included source proxies (that do *not* satisfy the criteria above) for at least two of three or
1476 more true ancestry sources (see three case studies in **Fig. S4k-m**).

1477

1478 ***Principal component analysis***

1479 PCA was performed for one simulation replicate per simulation setup. Only high-quality data
1480 was used, and all individuals sampled from a simulation were co-analyzed. Prior to the
1481 analysis, linked sites were pruned with *PLINK v.2.00a3LM* (Chang et al. 2015) using the
1482 following settings: window size, 2000 SNPs; window step, 100 SNPs; r^2 threshold = 0.5
1483 (argument "--indep-pairwise 2000 100 0.5"). PCA was also performed using *PLINK v.2.00a3LM*
1484 under default settings, calculating 10 PCs. Interactive three-dimensional plots visualizing PC1,
1485 PC2, and PC3 were made using the *plotly* R package. A two-way admixture model was
1486 considered supported by PCA if:

1487 1. the target group (the center of the cluster of target individuals, to be precise) lay between
1488 the clusters of proxy source individuals on a straight line in the three-dimensional PC
1489 space;
1490 2. or if it was located at a distance of no more than three target cluster diameters from that
1491 straight line connecting the proxy source clusters.

1492 The second pattern was more common among both TP and FP two-way admixture models:
1493 1.5 and 1.3 times, respectively (across all non-subsampled simulated datasets). This situation
1494 is expected since many targets represent three-way and more complex mixtures, and due to
1495 post-admixture genetic drift (Peter 2022).

1496

1497 ***ADMIXTURE analysis***

1498 *ADMIXTURE* analysis was performed for one simulation replicate per simulation setup. Only
1499 high-quality data was used, and all individuals sampled from a simulation were co-analyzed.
1500 Prior to the analysis, linked sites were pruned with *PLINK v.2.00a3LM* (Chang et al. 2015)
1501 using the following settings: window size, 2000 SNPs; window step, 100 SNPs; r^2 threshold =
1502 0.5 (argument "--indep-pairwise 2000 100 0.5"). *ADMIXTURE v.1.3* (Alexander et al. 2009)
1503 was used in the unsupervised mode under the default settings. The algorithm was run on
1504 each SNP dataset only once, with the number of hypothetical ancestral populations (*K*)
1505 ranging from 3 to 10. This range was selected since the total number of populations in each
1506 simulated history was 13. A two-way admixture model was considered supported by
1507 *ADMIXTURE* analysis if:

1508 1. for at least one *K*, at least 5 of 10 target individuals were modelled as a mixture of at least
1509 two ancestry components, with a minor ancestry component exceeding 2%;
1510 2. typically, ancestry component *A* in the target group was shared with at least 5 individuals
1511 in proxy source 1, but not in proxy source 2, and ancestry component *B* was shared with
1512 at least 5 individuals in proxy source 2, but not in proxy source 1 (see examples in **Fig. 4**
1513 and **Fig. S4**); in some cases, both components *A* and *B* were found in the proxy sources,
1514 but in different proportions;
1515 3. if only one ancestry component in the target was shared with the two proxy sources, the
1516 model was considered unsupported;
1517 4. ancestry components in the target that are absent in any of the sources were ignored
1518 since three-way and more complex admixture histories are common in the set of random
1519 admixture graphs explored here (see examples of such topologies in **Fig. S4k-m**);

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 (see examples of such topologies in **Fig. S4i, j, m**).

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 the 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 the 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.

Data Availability Statement

The study is based on simulated data exclusively. **Suppl. dataset 1** contains summaries of 33,072 feasible two-way *qpAdm* models on which **Table 2** and **Fig. 6** are based; it is deposited at the GSA Figshare portal. Principal software packages used in the study are *ADMIXTOOLS 2* (Maier et al. 2023, https://github.com/uqrmaie1/admixtools) and *msprime v. 1.1.1*

(Baumdicker et al. 2022, https://github.com/tskit-dev/msprime). For details on data simulations and computational analyses see lines 1313-1541 in this manuscript.

Acknowledgements

O.F, P.C., and P.F. were supported by the Czech Science Foundation (project no. 21-27624S led by P.F.). L.A.V. and P.F. were also supported by the John Templeton Foundation (grant no. 61220 to David Reich) and by a gift from Jean-Francois Clin. L.A.V., O.F., P.C., P.F., and U.I. were supported by the Czech Ministry of Education, Youth and Sports (program ERC CZ, project no. LL2103 led by P.C.). C.D.H. and M.P.W were funded by the National Institutes of Health under award number R35GM146886. 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. Diaconis P, Goel S, Holmes S. Horseshoes in multidimensional scaling and local kernel methods. *Ann. Appl. Stat*. **2** (2008) 777–807. doi: 10.1214/08-AOAS165.
14. Duforet-Frebourg N, Slatkin M. Isolation-by-distance-and-time in a stepping-stone model. *Theor. Popul. Biol*. **108** (2016) 24–35. doi: 10.1016/j.tpb.2015.11.003.
15. 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.
16. Estavoyer M, François O. Theoretical analysis of principal components in an umbrella model of intraspecific evolution. *Theor. Popul. Biol*. **148** (2022) 11–21. doi: 10.1016/j.tpb.2022.08.002.
17. 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.
18. 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.

19. Flegontov P, Işıldak U, Maier R, Yüncü E, Changmai P, Reich D. Modeling of African population history using *f*-statistics is biased when applying all previously proposed SNP ascertainment schemes. *PLoS Genet*., in press.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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.

25. 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.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.

31. 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.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.
37. 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.
38. 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.
39. 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.
40. 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.
41. 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.
42. 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.
43. 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.

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

45. 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.

46. 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.

47. 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.

48. 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.

49. 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.

50. 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.

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

52. 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.

53. 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.

54. 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.

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

56. 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.

57. 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.

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

59. Peter BM. A geometric relationship of *F2*, *F3* and *F4*-statistics with principal component analysis, *Philos Trans R Soc Lond B Biol Sci*. **377** (2022) 20200413. doi: 10.1098/rstb.2020.0413.

60. 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.

61. 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.

62. 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.
63. 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.
64. Reich D, Price AL, Patterson N. Principal component analysis of genetic data, *Nat. Genet*. **40** (2008) 491–492.
65. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history, *Nature*. **461** (2009) 489–494. doi: 10.1038/nature08365.
66. 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.
67. 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.

68. 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.

69. 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.

70. 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.

71. 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.

72. 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.

73. 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.

74. 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.

75. 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.

76. 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.

77. 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.

78. 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. doi: 10.1126/science.aba0909.

79. Zhang F, Ning C, Scott A, Fu Q, Bjørn R, Li W, Wei D, Wang W, Fan L, Abuduresule I, Hu X, Ruan

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, Wen S, Zhu H, Zhou H, Robbeets M, Kumar V, Krause J, Warinner C, Jeong C, Cui Y. The genomic origins of the Bronze Age Tarim Basin mummies. *Nature*. **599** (2021) 256–261. doi: 10.1038/s41586-021-04052-7.