An explanation for the neighbour repulsion phenomenon in Patterson's f-statistics

Gözde Atağ and Mehmet Somel

Department of Biological Sciences, Middle East Technical University, 06800 Ankara, Turkey

Abstract

Patterson's f-statistics are among the most heavily utilized tools for analysing genome-wide allele frequency data for demographic inference. Beyond studying admixture, f_3 and f_4 statistics are also used for clustering populations to identify groups with similar histories. However, previous studies have noted an unexpected behaviour of f-statistics: multiple populations from a certain region systematically show higher genetic affinity to a more distant population than to their neighbours, a pattern that is mismatched with alternative measures of genetic similarity as well as with spatial relationships. We call this counter-intuitive pattern "neighbour repulsion". Here, we first present a novel instance of neighbour repulsion, where genomes from Bronze Age East Anatolian sites show higher affinity towards Bronze Age Greece rather than their neighbours. This is observed both using f_3- and f_4-statistics, contrasts with archaeological/historical expectation, and also contradicts genetic affinity patterns captured using PCA or MDS on genetic distances. We then propose a simple demographic model to explain this pattern, where the neighbouring populations (such as East Anatolia) receive independent gene flow from an external source after their split from each other. Calculating f_3- and f_4-statistics using simulated genetic data with varying population genetic parameters, we confirm that low-level gene flow from an external source into populations from one region, as well as unidirectional gene flow between the studied regions, can create neighbour repulsion in f-statistics. Meanwhile, similar to our empirical observations, an MDS analysis of genetic distances can still capture common demographic history between neighbours. Overall, our results highlight the intricate effects of unaccounted population dynamics when inferring shared demographic history using f-statistics.

Introduction

Over the last decade, Patterson's f-statistics (Patterson *et al.* 2012) have become one of the most commonly applied tools in population genetics studies involving the analysis of genome-wide data from recombining chromosomes. Various f-statistics (f_2, outgroup-f_3, admixture-f_3, and f_4) quantify divergence, shared genetic drift and treeness among populations by comparing allele frequencies. These measurements enable identification of population structure, differentiation and admixture events and help decipher the demographic history of the studied populations (Peter 2016; Harris and DeGiorgio 2017). The apparent robustness of f-statistics to heterogeneity of sample size/coverage and ascertainment has gained them high popularity.

The outgroup-f_3 statistic in the form of f_3*(Outgroup; Pop2, Pop3)* measures shared drift between *Pop2* and *Pop3* relative to an outgroup, and its magnitude has been used as a measure of pairwise population affinity. The f_4-statistic measures treeness, such that f_4*(Outgroup, Pop1; Pop2, Pop3)* is expected to be 0 if *Pop1* shares the same drift with *Pop2* and *Pop3*, i.e. *Pop1* is a real outgroup to the latter two populations. Significant deviation from 0 can be interpreted as evidence for admixture (Patterson *et al.* 2012), but the f_4-statistic can also be used for studying population structure and shared demographic histories. If *Pop2* and *Pop3* share more of their demographic history, i.e. shared drift and/or similar admixture events that *Pop1* did not experience, and if there is no post-split admixture, we expect f_4*(Outgroup, Pop1; Pop2, Pop3)* to be non-significant, while f_4*(Outgroup, Pop2; Pop1, Pop3)* and f_4*(Outgroup, Pop3; Pop1, Pop2)* to both be positive. Gene flow from *Pop1* into *Pop3* in the same model will also cause f_4*(Outgroup, Pop1; Pop2, Pop3)* to be positive. Thus, in addition to testing admixture, reciprocal affinity in f_4-tests has frequently been used as a tool for clustering populations based on allele frequency similarities, where clustering implies similar histories. The outgroup-f_3 statistic has similarly been used to evaluate shared history between populations.

Nevertheless, f-statistics are not always straightforward to interpret. Complications of diverse type and origin have been noted, such as possible misinterpretation of an outgroup as an admixture source when using admixture-f_3 (Patterson *et al.* 2012), or the sensitivity of f_4-statistics to SNP ascertainment schemes as showcased for demographic modelling of African human populations using SNPs identified in non-African groups (Bergström *et al.* 2020; Flegontov *et al.* 2023). The most widespread type of issue arguably involves the interpretation of f_4 -statistics under the assumption of a simple demographic model where the real history is substantially more complicated. For instance, unrecognized spatial structure can be misinterpreted as admixture events (Durand *et al.* 2011). "Ghost admixture", i.e. gene flow from an unknown source, can

lead to erroneous identification of donor/recipient populations of introgression events (Rogers and Bohlender 2015; Tricou *et al.* 2022). A similar confounding effect can be created by backward gene flow, as shown for Neanderthal admixture estimates varying spuriously due to admixture between Eurasia and Africa within the last 45,000 years (Petr *et al.* 2019a).

Previous work using f-statistics has noted another unexpected pattern that has yet remained largely unexplored, to our knowledge. Using f_4-statistics, Kılınç et al. observed higher affinity of all Anatolian Neolithic groups to a Neolithic genome from Greece than to each other, a pattern that appeared difficult to interpret (Kılınç *et al.* 2017). Even more surprising was the observation that modern-day North African populations showed higher affinity to modern-day Sardinians than to each other using f_3- and f_4-statistics (Rodríguez-Varela *et al.* 2017). This pattern was hypothesized to stem from different degrees of gene flow from divergent sources, here sub-Saharan Africa, but the case was not studied any further.

Here, we provide another case of this pattern, which we call “neighbour repulsion”, by comparing genomes from the Bronze Age (BA) East Mediterranean. We then propose scenarios involving external and regional gene flow to explain neighbour repulsion and recapitulate the observed f-statistics using simulations under these models, with various conditions. Our results showcase how complex demographic histories can lead to counter-intuitive patterns of f_3- and f_4-statistics when used in the context of studying shared population history.

Results and Discussion

Our initial motivation was to investigate population structure and admixture patterns in the East Mediterranean during the Bronze Ages, a period which saw population growth and extensive interregional trade networks being established, as inferred from material culture data (Şahoğlu 2005; Kouka 2016; Massa 2016). We studied genetic affinities between published genomic data from n=6 populations from BA Greece (2400-1070 BCE) and n=5 populations from BA East Anatolia (3700-1303 BCE) (Lazaridis *et al.* 2017; Skourtanioti *et al.* 2020; Lazaridis *et al.* 2022b) (Table S1). We restricted the analysis to 1240K capture data to avoid confounding by technical factors. We note that the results described henceforth are also observed using Central and West Anatolian BA populations, which we do not show here; we instead focus on East Anatolia as analyses involving populations from this region are the most counter-intuitive.

First, we explored genetic clustering among populations with Principal Components Analysis (PCA) and Multidimensional Scaling (MDS) analysis. In the PCA, where the ancient genomes are projected on a PC space calculated from modern-day West Eurasians, genomes from Greece group separately from those from East Anatolia (Figure S1). Then, we calculated genetic distance between all pairs of populations using 1 - f_3*(Yoruba; Pop1, Pop2)* to represent distance between populations *Pop1* and *Pop2*. Applying MDS analysis on the resulting pairwise distance matrix, we again observed clear clustering in MDS space that mirrored geography (Figure 1A,B).

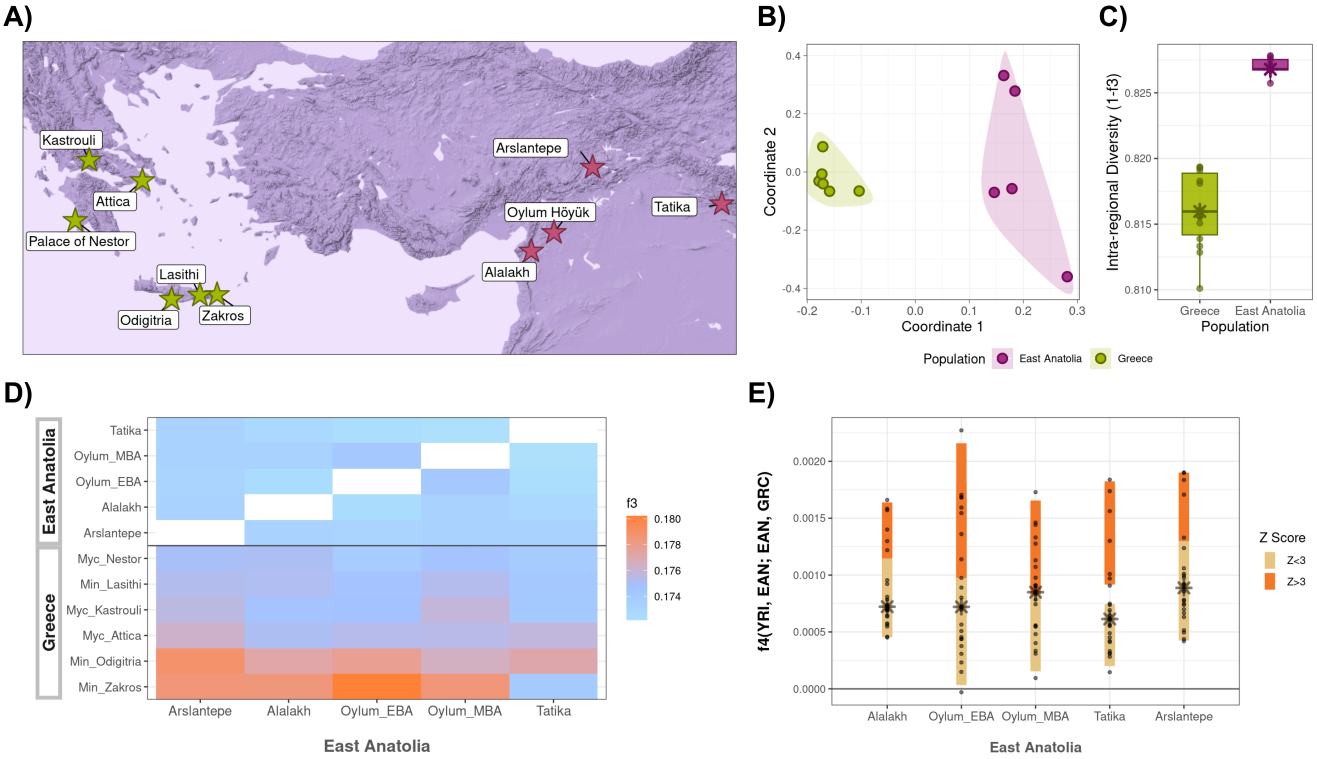


Figure 1: (A) Geographical distribution of the studied Bronze Age genomes from Greece and East Anatolia. **(B)** Genetic clustering of the populations in line with geography revealed by multidimensional scaling analysis (MDS), using 1 - outgroup-f_3 values as genetic distances. **(C)** Intra-regional diversity of the populations measured as pairwise 1 - outgroup-f_3 values within regions. **(D)** f_3 statistics of the form f_3*(Yoruba; East Anatolia, X)* where X corresponds to populations from Greece and East Anatolia. **(E)** f_4 statistics of the form f_4*(Yoruba, East Anatolia; East Anatolia, Greece)*, where positive values depict higher affinity of East Anatolian populations towards Greek populations.

Intriguingly, however, and in contrast to the patterns observed in PCA and MDS, we noticed that the values f_3*(Yoruba; EAnatoliaY, GreeceX)* were consistently higher than the values f_3*(Yoruba; EAnatoliaY, EAnatoliaZ)*, where *GreeceX* is one of the six populations from Greece, while *EAnatoliaY* and *EAnatoliaZ* are any of the five populations from East Anatolia (Figure 1C, Table S2). On face value, this would imply more similar histories between any East Anatolian population with those from Greece than other neighbouring East Anatolian populations.

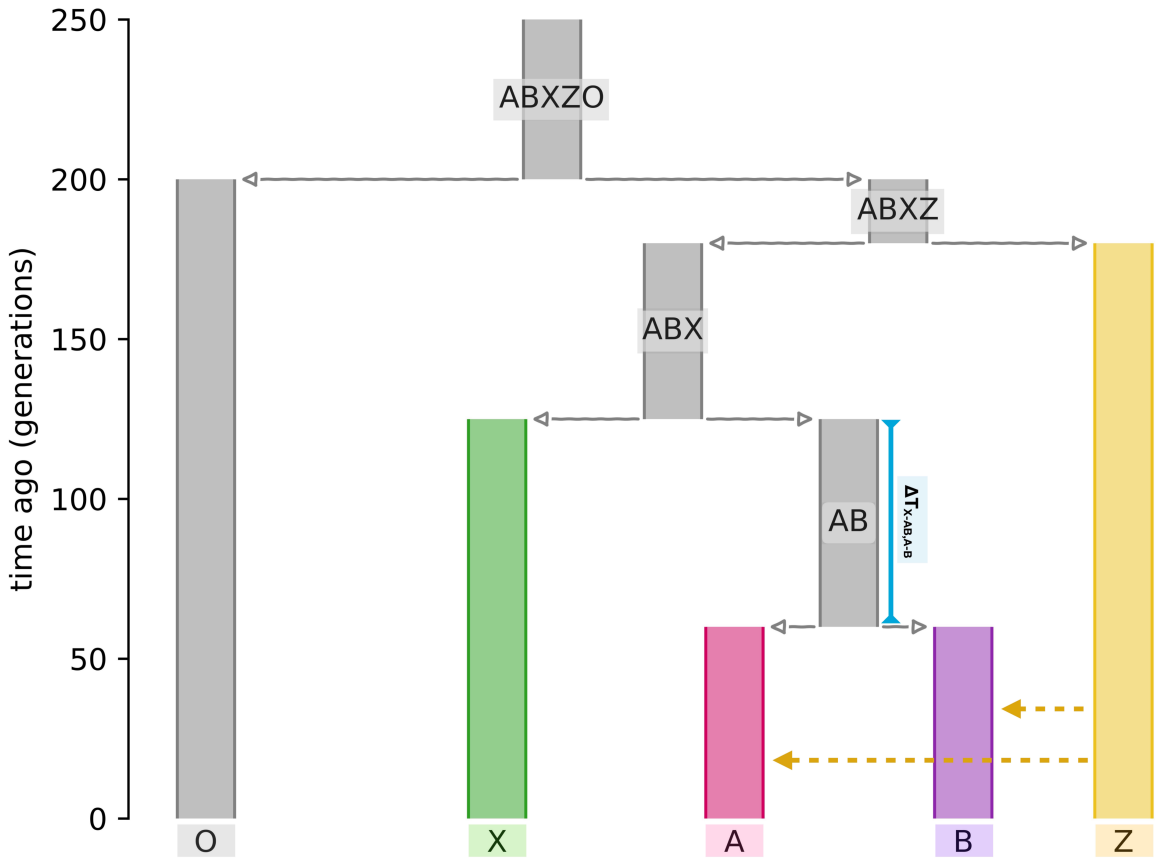
We next calculated f_4-statistics of the form f_4*(Yoruba, AnatoliaY; AnatoliaZ, GreeceX)*. In 99% of 120 comparisons we observed positive values (42% significant at Z>3, without correcting for multiple testing) (Figure 1D, Table S3). No f_4-statistic was significantly negative, i.e. no East Anatolian group significantly shared more drift with one of its neighbours over a population from Greece. This starkly contrasts with the expected geographical clustering, while it does mirror the outgroup-f_3 results.

We hypothesized two explanations for this neighbour repulsion pattern. One may be gene flow from Greece into each East Anatolian population, independently. Still, given the thousands of kilometers of distance between Greece and East Anatolia, and that this pattern was observed for all populations in the regions studied, this may not be a parsimonious explanation. An alternative explanation and possibly more likely scenario could involve admixture from an external and diverse source impacting East Anatolia, similar to that hypothesized by Rodríguez-Varela and colleagues (2017).

The ancestral components of both Anatolia and Greek Early/Middle Bronze Age genomes comprise Neolithic Anatolia and eastern-related ancestry from Caucasus/Mesopotamia (Clemente *et al.* 2021; Lazaridis *et al.* 2022b; Skourtanioti *et al.* 2023). We hypothesized that additional external gene flow into East Anatolia (e.g. from the Levant, Zagros, or Mesopotamia) that BA Greece did not receive could lead to differentiation within the former region and might explain the observed neighbour repulsion pattern. As a first test of this idea, we calculated intra-regional diversity within each site, using pairwise 1- outgroup-f_3 among genomes within a site as the diversity measure. This revealed significantly higher median diversity in BA East Anatolia than in BA Greece (Wilcoxon rank sum test p=3e-09).

We then sought to replicate neighbour repulsion using population genetic simulations under a simple demographic model with external admixture (Figure 2). We define A and B as two sister populations, X as a more distant lineage, Z as the source of external gene flow into A and B, and O as outgroup to all. Here, A, B and X as our focal populations; A and B could represent two BA East Anatolian groups, X could be a population from BA

Greece, and Z would be an unknown source that admixes only with East Anatolia. The ancestors of multiple populations are denoted by their merged names (i.e. the ancestor population of A and B is called AB). Note that the demographic histories of A and B are the same except for receiving independent gene flow from Z.



★ **f4(O,A; B,X) < 0 : <u>neighbour attraction</u>** ★ **f4(O,A; B,X) > 0 : <u>neighbour repulsion</u>**

Figure 2: Demographic model used for the population genetic simulations. The ancestral populations are named as combinations of the descending population names. Population O is the outgroup used in the calculation of f-statistics. Dashed lines represent equal rates of continuous migration from Z to A and B per generation. The blue line corresponds to the parameter $\Delta T_{X\text{-}AB,A\text{-}B}$.

Using this general model we ran multiple simulations where we varied the parameters $m_{Z\text{->}A.B}$, Ne_Z, Ne_{AB}, Ne_{ABX}, $T_{A\text{-}B}$, and $T_{X\text{-}AB}$ (Table 1). For simplicity, we refer to $T_{A\text{-}B}$ and $T_{X\text{-}AB}$ as a single variable that shows $T_{X\text{-}AB}$ - $T_{A\text{-}B}$, which we call $\Delta T_{X\text{-}AB,A\text{-}B}$. In each case we

sampled 25 genomes each from O, X, A, and B, and calculated various outgroup-f_3 and f_4-statistics.

Table 1: Parameter values used for the population genetic simulations. Migration rates are given per generation, divergence times are given in generations.

Parameter	Values	Definition
$m_{Z->A.B}$	0, 0.001, 0.005, 0.01, 0.1	Migration rate from Z to A and B
Ne_Z	5k, 15k, 35k	Effective population size of source Z
Ne_{AB}	5k, 15k, 35k	Effective population size of ancestral AB
Ne_{ABX}	1k, 3k, 5k	Effective population size of ancestral ABX
T_{A-B}	30, 60, 90	Divergence time of A and B
T_{X-AB}	100, 125, 150	Divergence time of X and AB
$\Delta T_{X-AB,A-B}$	10, 65, 120	Difference between T_{A-B} and T_{X-AB}
Ne_X*	1k, 3k, 5k	Effective population size of X
$m_{X->A,B}$*	0, 0.001, 0.005, 0.01, 0.1	Migration rate from X to A and B

* Given parameter values for Ne_X and $m_{X->A,B}$ were tested only for scenarios where Ne_{ABX} = 1k and $\Delta T_{X-AB,A-B}$ = 65. At the remaining scenarios, Ne_X was fixed at 5k and $m_{X->A,B}$ at 0.

In this model (Figure 2) one would anticipate $f_4(O, A; B, X)$ values to be negative since the populations A and B appear to share more similar histories with each other than with X. Indeed, in the absence of gene flow from Z into A/B (m=0), f_4 statistics are mostly negative, reflecting higher shared drift between A and B than with X (Figure 3, Figure S2). However, f_4 values change as we allow migration from Z into A/B (from m=0 to m=0.1), although in non-linear fashion. We observe a monotonic increase in $f_4(O, A; B, X)$ values between 0 and 0.01, i.e. a stronger trend towards neighbour repulsion (Figure 3). Under specific conditions, e.g. Ne_{ABX} = 1k, $0.001 \leq m \leq 0.1$ and $\Delta T_{X-AB,A-B} \leq 65$, all $f_4(O, A; B, X)$ values are positive (Figure 3, S2). The same applies to $f_4(O, B; A, X)$. This neighbour repulsion pattern is analogous to that observed in the empirical data where two East Anatolian populations showed higher affinity to populations from Greece than to each other (Figure 1D).

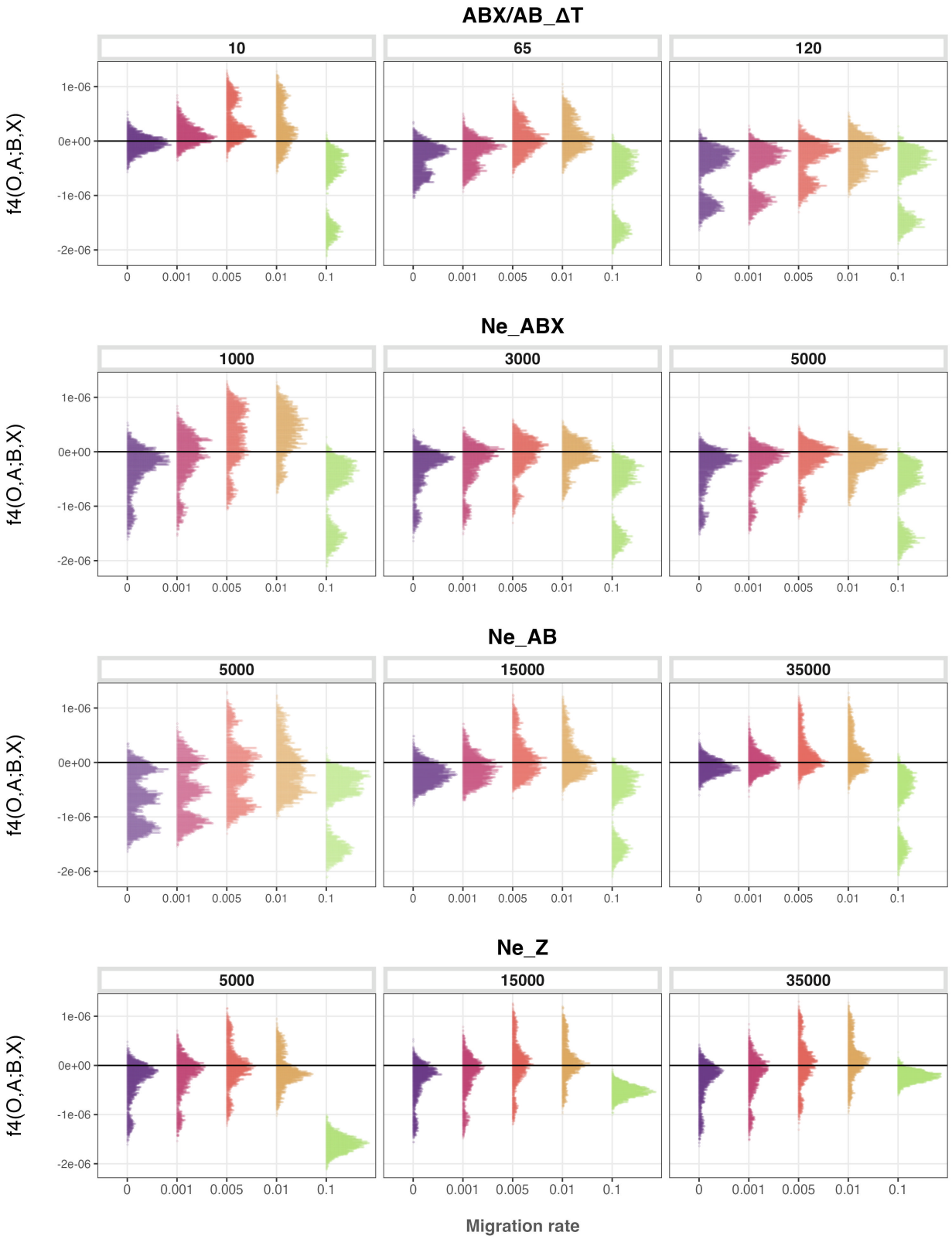


Figure 3: Change in f_4 statistics of the form $f_4(O, A; B, X)$ depending on the varying parameters in population genetic simulations, where negative values represent higher

affinity between A and B, and positive values higher affinity between A and X. The panels for each parameter contain all values of the remaining studied parameters and are explained in Table 1.

We note a number of conditions driving more positive $f_4(O, A; B, X)$ values under low-level migration in our model (Figure 2; Table 1): (a) Low $\Delta T_{X\text{-}AB,A\text{-}B}$, which means A and B share less drift relative to X (the phylogeny is more star-like), and also that A and B are impacted by Z-sourced admixture for fewer generations. (b) Low Ne_{ABX}, which causes stronger shared drift between A, B and X. (c) High Ne_{AB}, which leads to lower shared drift between A and B. (d) High Ne_Z, whereby Z is more diverse and thus Z-sourced gene flow differentiates A and B more effectively.

As opposed to results involving low-level migration (m=0.001 to m=0.01), at m=0.1, $f_4(O, A; B, X)$ values shift back to negative values, i.e. neighbour attraction (Figure 3, S2). This happens because high migration rates from Z leads to homogenization of A and B. The drop in f_4 values is particularly strong if Ne_Z is low: e.g. at m=0.1 and Ne_Z = 5k all results are negative (Figure 3, S2).

We then asked whether in simulations where we detect neighbour repulsion between A and B in f_4-statistics, we would still observe A and B clustering separately from X in the MDS analysis, a pattern we had observed in the empirical data (Figure 1). For this, we chose one random case from each scenario and calculated outgroup-f_3 statistics. We then compared the mean distance between X to A and B [$\bar{d}_{mds}$(X,A/B)] versus the distance between A and B [d_{mds}(A,B)] in MDS space. In 99% of all cases, populations A and B clustered separately from X, i.e. d_{mds}(A,B) was lower. This included scenarios where 100% of the f_4 values signified neighbour repulsion (Figure 4A,B). Meanwhile, under low-level migration we also observed $f_3(O; A, B) < f_3(O; A/B, X)$, reflecting neighbour repulsion (Figure 4C). Hence, similar to our observation with the East Mediterranean Bronze Age genomes (Figure 1B,D), MDS-based clusterings in our simulations contradict pairwise f_3 statistics.

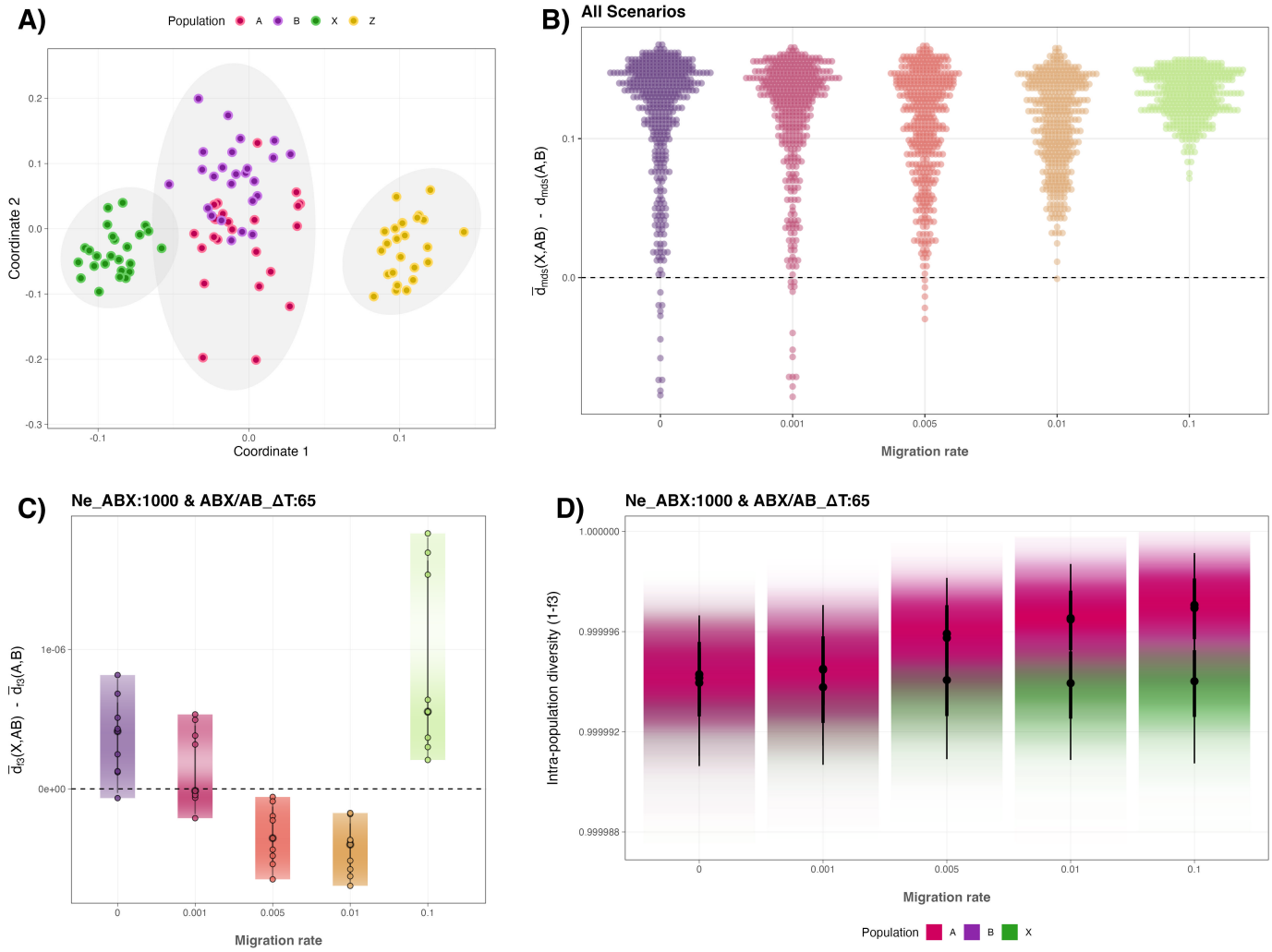


Figure 4: **(A)** MDS of simulated populations A, B, X and Z, using 1 - outgroup-f_3 values as genetic distances, for one sample scenario where $m_{Z\text{->}A.B}$: 0.01 ,Ne_Z : 15000, Ne_{AB} : 15000 , Ne_{ABX} : 1000, and $\Delta T_{X\text{-}AB,A\text{-}B}$: 65. **(B)** MDS clusterings between A and B versus X and B, calculated by the Euclidean distances between the mean MDS coordinates of the corresponding populations, for all scenarios. $\bar{d}_{mds}$(X,AB) is the mean distance of X to A and B. d_{mds}(A,B) is the distance between A and B. Positive values for $\bar{d}_{mds}$(X,AB) - d_{mds}(A,B) represent higher distance of X to A and B, thus distinct clustering of X from A and B. **(C)** Change in f3 values with migration rate from Z for scenarios where Ne_{ABX} : 1000 and $\Delta T_{X\text{-}AB,A\text{-}B}$: 65. $\bar{d}_{f3}$(X,A/B) is the mean 1-f3 distance of X to A or to B. $\bar{d}_{f3}$(A,B) is the mean 1-f3 distance between A and B. Negative values for $\bar{d}_{f3}$(X,A/B) - $\bar{d}_{f3}$(A,B) show higher affinity between A or B and X, than between A and B. **(D)** Intra-population diversity differences between populations A,B and X in relation to migration rate from Z, for scenarios where Ne_{ABX} : 1000 and $\Delta T_{X\text{-}AB,A\text{-}B}$: 65. The diversities are measured by pairwise 1- outgroup f3 values per population. At migration rates 0 and 0.001, the values for X have higher deviation and overlap with A and B, thus have lower visibility.

In these models, we might expect regional differences in within-population diversity, which could be a source of information with respect to admixture dynamics. Indeed, in the Bronze Age dataset we had observed significantly higher diversity in genomes from Anatolia versus Greece (Figure 1C). In our simulations that involved gene flow from Z, we likewise observed elevated intra-population diversity in A and B versus X (Figure 4D).

We finally studied neighbour repulsion under two additional models of gene flow. In one case, we asked how asymmetric Z-sourced gene flow, where only population B receives migrants from Z, would impact f_4 statistics. In this case, as might be expected, $f_4(O, A; B, X)$ tends to be positive, i.e. A shows neighbour repulsion, while $f_4(O, B; A, X)$ tends to be negative, i.e. B is still attracted to its neighbour A (Figure S3).

We then asked whether gene flow from X into A and into B independently can also lead to neighbour repulsion, without invoking the external source Z. We confirmed this was the case using both f_3- and f_4-statistics (Figure S4A). We further found that patterns observed both in the above simulations with external gene flow (Figure 3,4) and in the empirical data (Figure 1), namely lower diversity in X versus A/B and separate MDS clustering of X and A/B, can be replicated under gene flow from X, but only when the effective population size of X (Ne_X) is low relative to Ne_A and Ne_B (Figure S4B,C).

The results above reveal how low-level gene flow from an external source (Z) can lead to neighbour repulsion i.e. closely related pairs of populations (A and B) showing higher affinity to a lineage that has split earlier from them (X). We also show that independent regional gene flow from X into A and B can likewise create neighbour repulsion patterns, especially when Ne_X is relatively small.

We believe neighbour repulsion is an interesting case because in the scenarios we studied, A and B share more similar demographic histories with each other than with X, although this is not directly reflected in the f_3- and f_4-statistics. Whether post-split admixture from X or an external population Z is more likely would need to be evaluated using other sources of evidence, such as local ancestry inference or archaeological/historical evidence for regional gene flow. In the empirical cases we covered, such as neighbour repulsion in modern-day North Africa or Bronze Age East Anatolia, external gene flow may be a more parsimonious explanation than the latter.

Methods

Ancient genomes

We used 79 published Bronze Age genomes from the East Mediterranean (Supplementary Table 1), from a publicly available dataset comprising the ancient individuals analyzed in Lazaridis et al., 2022, genotyped at the 1240K array positions (Lazaridis *et al.* 2022a). We only used Bronze Age genomes from East Anatolia and Greece produced with the 1240K enrichment method in order to avoid any technical bias that can arise when co-analyzing different data types, e.g. shotgun versus capture.

Principal Components Analysis (PCA)

For the principal components analysis, we used the Human Origins Dataset (Lazaridis *et al.* 2014, 2016), containing 2,583 present-day genomes genotyped at the Affymetrix Human Origins Array positions, merged with the Bronze Age genomes. PCA was performed using *smartpca v.18140*, with option *lsqproject:YES*, implemented in *EIGENSOFT v.7.2.0* (Patterson *et al.* 2006). We projected the ancient individuals onto the first two principal components calculated from present-day Western Eurasian diversity based on populations from the Human Origins panel.

f_3 , f_4 - statistics and Multidimensional scaling (MDS)

We calculated outgroup f_3- and f_4-statistics using *admixr v0.9.1,* which runs *AdmixTools v.7.0.2,* with default parameters (Patterson *et al.* 2012; Petr *et al.* 2019b). Within-group genetic diversities were estimated by computing pairwise 1 - outgroup-f_3. All statistics were computed based on the 1240K dataset, using 3 modern-day Yoruba individuals as the outgroup. We estimated pairwise genetic distances with 1- outgroup f_3, computed a distance matrix of 1-f_3 values, and performed a classical multidimensional scaling analysis using the *cmdscale* algorithm implemented in the R package *stats* (R Core Team, 2023).

Simulations

We used *msprime v1.0* (Baumdicker *et al.* 2022) to simulate diploid sequences of 100Mbs, where we sampled 25 individuals from each population. The divergence time of the population used as outgroup (ABXZO) was given as 200 generations ago, followed by the divergence of ABXZ after 20 generations. The effective population size of each population other than the ones given in Table 1 were fixed as 5k, i.e. populations ABXZO,

O, ABXZ, X (except for scenarios where Ne_{ABX} = 1k and $\Delta T_{X\text{-}AB,A\text{-}B}$ = 65), A, and B. The mutation rate and the recombination rate were chosen as 1e-8 and 1.1e-8 per generation, respectively (Baumdicker *et al.* 2022). Each scenario had 100 replicates. One-way migration rates from Z were set symmetric into both A and B, except for one case (Ne_Z, Ne_{AB}, Ne_{ABX}, $\Delta T_{X\text{-}AB,A\text{-}B}$) where we tested the effect of asymmetric migration by allowing gene flow (m=0.01) only into B. For scenarios where Ne_{ABX} = 1k and $\Delta T_{X\text{-}AB,A\text{-}B}$ = 65, one-way migration from X into A and B was also tested (Table 1). All combinations of values for parameters $m_{Z\text{->}A.B}$, Ne_Z, Ne_{AB}, Ne_{ABX}, and $\Delta T_{X\text{-}AB,A\text{-}B}$ given in Table 1 were tested. The outgroup-f_3 and f_4 statistics were calculated using *tskit v0.5.6*, which follows the calculation in Patterson et al., 2012. The MDS analysis was performed the same way as above. In order to compare the MDS distances between populations, we used the mean MDS x and y coordinates of each individual, and calculated the mean coordinate point per population. We then calculated Euclidean distances between the mean coordinates of the corresponding populations.

Data Availability

The code and parameters for generating the simulated data are deposited under https://github.com/goztag/msprime_fstats/tree/main.

Acknowledgements

We thank Emilia Huerta-Sanchez, Torsten Günther, Benjamin Vernot, Pavlos Pavlidis, Dilek Koptekin, Yetkin Alıcı and members of the METU CompEvo Team for helpful discussions. The work was supported by the H2020 ERC Consolidator grant (no. 772390 NEOGENE to M.S.), H2020-WIDESPREAD-05-2020 TWINNING grant (no. 952317 NEOMATRIX to M.S.) and Scientific and Technological Research Council of Turkey (TÜBİTAK) through the 2210/A National Scholarship Programme for MSc Students (to G.A.).

References

Baumdicker F., G. Bisschop, D. Goldstein, G. Gower, A. P. Ragsdale, *et al.*, 2022 Efficient ancestry and mutation simulation with msprime 1.0. Genetics 220: iyab229. https://doi.org/10.1093/genetics/iyab229

Bergström A., S. A. McCarthy, R. Hui, M. A. Almarri, Q. Ayub, *et al.*, 2020 Insights into human genetic variation and population history from 929 diverse genomes. Science 367: eaay5012. https://doi.org/10.1126/science.aay5012

Clemente F., M. Unterländer, O. Dolgova, C. E. G. Amorim, F. Coroado-Santos, *et al.*, 2021 The genomic history of the Aegean palatial civilizations. Cell 184: 2565-2586.e21. https://doi.org/10.1016/j.cell.2021.03.039

Durand E. Y., N. Patterson, D. Reich, and M. Slatkin, 2011 Testing for ancient admixture between closely related populations. Mol Biol Evol 28: 2239–2252. https://doi.org/10.1093/molbev/msr048

Flegontov P., U. Işıldak, R. Maier, E. Yüncü, P. Changmai, *et al.*, 2023 Modeling of African population history using f-statistics is biased when applying all previously proposed SNP ascertainment schemes. PLOS Genetics 19: e1010931. https://doi.org/10.1371/journal.pgen.1010931

Harris A. M., and M. DeGiorgio, 2017 Admixture and Ancestry Inference from Ancient and Modern Samples through Measures of Population Genetic Drift. Hum Biol 89: 21–46. https://doi.org/10.13110/humanbiology.89.1.02

Kılınç G. M., D. Koptekin, Ç. Atakuman, A. P. Sümer, H. M. Dönertaş, *et al.*, 2017 Archaeogenomic analysis of the first steps of Neolithization in Anatolia and the Aegean. Proceedings of the Royal Society B: Biological Sciences 284: 20172064. https://doi.org/10.1098/rspb.2017.2064

Kouka O., 2016 The Built Environment and Cultural Connectivity in the Aegean Early Bronze

Age, in Molloy, B.P.C (ed.), Of Odysseys and Oddities, Sheffield Studies in Aegean Archaeology, Oxbow Books, 203-222.

Lazaridis I., N. Patterson, A. Mittnik, G. Renaud, S. Mallick, *et al.*, 2014 Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513: 409–413. https://doi.org/10.1038/nature13673

Lazaridis I., D. Nadel, G. Rollefson, D. C. Merrett, N. Rohland, *et al.*, 2016 Genomic insights into the origin of farming in the ancient Near East. Nature 536: 419–424. https://doi.org/10.1038/nature19310

Lazaridis I., A. Mittnik, N. Patterson, S. Mallick, N. Rohland, *et al.*, 2017 Genetic origins of the Minoans and Mycenaeans. Nature 548: 214–218. https://doi.org/10.1038/nature23310

Lazaridis I., S. Alpaslan-Roodenberg, R. Pinhasi, and D. Reich, 2022a The genetic history of the Southern Arc: A bridge between West Asia and Europe

Lazaridis I., S. Alpaslan-Roodenberg, A. Acar, A. Açıkkol, A. Agelarakis, *et al.*, 2022b The genetic history of the Southern Arc: A bridge between West Asia and Europe. Science 377: eabm4247. https://doi.org/10.1126/science.abm4247

Massa M., 2016 Networks before Empires: cultural transfers in western and central Anatolia during the Early Bronze Age

Patterson N., A. L. Price, and D. Reich, 2006 Population Structure and Eigenanalysis. PLOS Genetics 2: e190. https://doi.org/10.1371/journal.pgen.0020190

Patterson N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, *et al.*, 2012 Ancient Admixture in Human History. Genetics 192: 1065–1093. https://doi.org/10.1534/genetics.112.145037

Peter B. M., 2016 Admixture, Population Structure, and F-Statistics. Genetics 202: 1485–1501. https://doi.org/10.1534/genetics.115.183913

Petr M., S. Pääbo, J. Kelso, and B. Vernot, 2019a Limits of long-term selection against Neandertal introgression. Proc. Natl. Acad. Sci. U.S.A. 116: 1639–1644. https://doi.org/10.1073/pnas.1814338116

436 Petr M., B. Vernot, and J. Kelso, 2019b admixr—R package for reproducible analyses using
437 ADMIXTOOLS. Bioinformatics 35: 3194–3195.
438 https://doi.org/10.1093/bioinformatics/btz030
439 R Core Team, 2023 _R: A Language and Environment for Statistical Computing_. R
440 Foundation for Statistical Computing, Vienna, Austria. <https://www.R-
441 project.org/>.
442 Rodríguez-Varela R., T. Günther, M. Krzewińska, J. Storå, T. H. Gillingwater, *et al.*, 2017
443 Genomic Analyses of Pre-European Conquest Human Remains from the Canary Islands
444 Reveal Close Affinity to Modern North Africans. Current Biology 27: 3396-3402.e5.
445 https://doi.org/10.1016/j.cub.2017.09.059
446 Rogers A. R., and R. J. Bohlender, 2015 Bias in estimators of archaic admixture. Theoretical
447 Population Biology 100: 63–78. https://doi.org/10.1016/j.tpb.2014.12.006
448 Şahoğlu V., 2005 The Anatolian Trade Network and the Izmir region during the Early Bronze
449 Age. Oxford Journal of Archaeology 24: 339–361. https://doi.org/10.1111/j.1468-
450 0092.2005.00240.x
451 Skourtanioti E., Y. S. Erdal, M. Frangipane, F. Balossi Restelli, K. A. Yener, *et al.*, 2020
452 Genomic History of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern
453 Caucasus. Cell 181: 1158-1175.e28. https://doi.org/10.1016/j.cell.2020.04.044
454 Skourtanioti E., H. Ringbauer, G. A. Gnecchi Ruscone, R. A. Bianco, M. Burri, *et al.*, 2023
455 Ancient DNA reveals admixture history and endogamy in the prehistoric Aegean. Nat
456 Ecol Evol 1–14. https://doi.org/10.1038/s41559-022-01952-3
457 Tricou T., E. Tannier, and D. M. de Vienne, 2022 Ghost Lineages Highly Influence the
458 Interpretation of Introgression Tests. Systematic Biology 71: 1147–1158.
459 https://doi.org/10.1093/sysbio/syac011