



Impact of cultural and genetic structure on food choices along the Silk Road

Serena Aneli^{a,b,1}, Massimo Mezzavilla^{a,c}, Eugenio Bortolini^{d,e}, Nicola Pirastu^f, Giorgia Giroto^{c,g}, Beatrice Spedicati^g, Paola Berchiaglia^h, Paolo Gasparini^{c,g}, and Luca Pagani^{a,i}

Edited by Marcus Feldman, Stanford University, Stanford, CA; received June 8, 2022; accepted October 3, 2022

The complex interplay between genetics, culture, and environment forms an individual's biology, influencing their behavior, choices, and health. However, to what extent information derived from this intertwined network could be quantitatively summarized to provide a glance at an individual's lifestyle is difficult to say. Here, we focused on dietary preferences as cultural proxies and genome-wide data of 543 individuals from six historical Silk Road countries: Georgia, Armenia, Azerbaijan, Uzbekistan, Kazakhstan, and Tajikistan. These lands favored the dispersal of innovations, foods, and DNA half-way across Eurasia, thus representing an ideal subject to explore interactions of cultural factors and genetic ancestry. We used discriminant analysis of principal components to infer cultural clusters, where mixed memberships are allowed. Five different clusters emerged. Of these, clusters 1 and 3, driven by aversion to pork and alcoholic beverages, mirrored genetic admixture patterns with the exception of Azerbaijan, which shares preferences supported by Islamic culture with Eastern countries. Cluster 3 was driven by protein-rich foods, whose preference was significantly related to steppe pastoralist ancestry. Sex and age were secondary clustering factors, with clusters formed by male and young individuals being related to alcohol preference and a reduced liking for vegetables. The soft clustering approach enabled us to model and summarize the individual's dietary information in short and informative vectors, which show meaningful interaction with other nondietary attributes of the studied individuals. Encoding other cultural variables would help summarize an individual's culture quantitatively, thus ultimately supporting its inclusion as a covariate in future association studies.

cultural clustering | food culture | genetic structure | Silk Road

Over the past forty years, a vast body of empirical and theoretical studies investigated the mechanisms of cultural transmission and culture change in various chronological and geographic contexts, drawing especially on parallels with genetic evolutionary processes (1–4). Recent examples include works aimed at disentangling the many forces underlying the differential persistence of cultural variants (5–17); the relationship between population and social structure, interaction networks, and the transmission of culture (18–22); patterns of cultural transmission over time and space (23, 24); the effects of intergroup interaction and branching–blending dynamics (25–32); the relationship between cultural diversity and demography (33–39); and possible relationships between biological and cultural variability (40–50).

Among the ubiquitous cultural variables shared by all human groups (e.g., language, traditional tales, beliefs), dietary and culinary habits have also been investigated through the same approach. These works aimed at theory building and model fitting concerning the link between mobility, population density, and resource quality (51), as well as at exploring the adaptive consequences of food choices (52) and change over time in human diet (53). Dietary choices are particularly relevant as they emerge at the interface between individual preferences, group-level systematic choices, cultural and environmental constraints, energy requirements, ecology, and many evolutionary and developmental processes. At the same time, food literally becomes the person who consumes it (“Man is what he eats” is a statement coined by philosopher Ludwig Feuerbach), and what we eat directly affects our health and viability. A number of food sharing practices emerged throughout human history and became a relevant part of group identity and memory. Humans attach nonfunctional meaning to food, food processing, and food consumption, which becomes a crucial moment for interaction, exchange, and the emergence of complex social structures. Such a widespread and pivotal form of cultural expression offers a privileged perspective on cultural adaptation, as well as culture change and transmission across geography and time. However, more importantly, it poses a direct link to genetic and biological variability (54). Nevertheless, to date there is little evidence on the impact that geographic, cultural,

Significance

Among the universal cultural variables shared by human groups (e.g., language, traditional tales, beliefs), dietary habits play a key role at the individual level, such that “you are what you eat,” directly affecting one's health and viability. In this work, we explored food preferences for 79 different foods within six populations along the historical Silk Road route. By using a probabilistic clustering technique, we summarized the cultural variables to cluster memberships and analyzed them in the light of sex, age, and genetic information from the same individuals. In doing so, we highlighted different facets of cultural admixture and stratification, reflecting religious dietary restrictions, age-related personal preferences, and pastoralist traditions, which in some cases genetics alone failed to identify.

Author contributions: S.A. and L.P. designed research; S.A. and M.M. performed research; N.P., G.G., B.S., P.B., and P.G. contributed new analytic tools; S.A. and M.M. analyzed data; and S.A., M.M., E.B., and L.P. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: serena.aneli@unito.it.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2209311119/-/DCSupplemental>.

Published November 14, 2022.

and genetic structure have on the cross-cultural distribution of food preferences and culinary traditions.

Here we investigated food choices as proxies for cultural group membership, aiming at exploring whether some cultural factors, such as religion, or other personal characteristics (e.g., sex, age, and genetic ancestry) may shape the interindividual differences in food choices and, consequently, the cultural grouping. We focused on 538 individuals from six populations of the Silk Road and investigated their food preferences for 79 different foods along with genotypes, sex, and age. The Silk Road has been a trading route between Europe and Asia from the second century BCE to the 15th century CE, thus representing an ideal case to study patterns of cultural and genetic variation and admixture. These lands, dotted by myriad trading posts, were crossed for millennia by various cultural, political, religious, and culinary innovations, thus favoring the dispersal of technological innovations, artifacts, seeds, and human DNA halfway across Eurasia (55). Previous works (56) highlighted in the populations living along this ancient route a complex pattern of admixture events and genetic diversity, as well as phenotypic differentiation in terms of bitter taste perception (57) and eye color (58). In particular, some of those populations experienced different admixture events during the Bronze Age (59) as well as more recent migrations and invasions such as the Mongol expansion in the 13th and 14th centuries CE. For these reasons, these populations are a well-suited natural laboratory for exploring the connection between genetic and cultural factors in modern multiethnic societies characterized by a complex demographic history. This work aimed to investigate whether food preferences can be effectively used to summarize an individual's membership to a group inferred through clustering techniques. A positive result would open up possibilities for modeling complex patterns of cultural admixture in an ever-connected world, characterized by the increased presence

of genetically and possibly culturally admixed individuals, similarly to what is done in genetic ancestry analysis.

Results

Genetic, Culinary, and Geographic Patterns Along the Silk Road. The populations living along the Silk Road, which spans a wide geographic region connecting Europe to the Far East (Fig. 1), experienced a complex demographic history, as evidenced by analyses of modern genomes (56). As evidenced by admixture analyses on the populations from six countries along this route (Georgia, Armenia, Azerbaijan, Uzbekistan, Kazakhstan, and Tajikistan; supervised and unsupervised analyses in Fig. 1 [Inset] and *SI Appendix*, Fig. S1), individuals living on opposite sides of the Caspian Sea exhibit different genetic structures, with Eastern Asian influxes (proxies for the ancient Kazakh/Mongolian population) being predominant in Kazakhstan in the east and reaching only Azerbaijan among the western countries. Whether cultural phenomena such as dietary preferences would have been also hampered by geographic distances and geomorphological barriers, namely the Caspian Sea, is yet to be established.

In order to investigate whether cultural and genetic variation followed similar patterns, we analyzed food preference questionnaires for 79 patterns, compiled by 543 individuals living in the above-mentioned countries. The participants were asked to rate the foods on a 5-point scale from 1 (“dislike extremely”) to 5 (“like extremely”) or as “never tasted.” Using the unweighted pair group method with arithmetic mean clustering algorithm (UPGMA), we found six food subsets that, according to their compositions, were named “alcohol” (e.g., red wine), “pork” (e.g., ham), “sweets” (e.g., milk chocolate), “vegetables” (e.g., peppers), “animal proteins” (e.g., whole milk), and “fruits” (e.g., apples; *SI Appendix*, Fig. S2). The “never tasted” answers

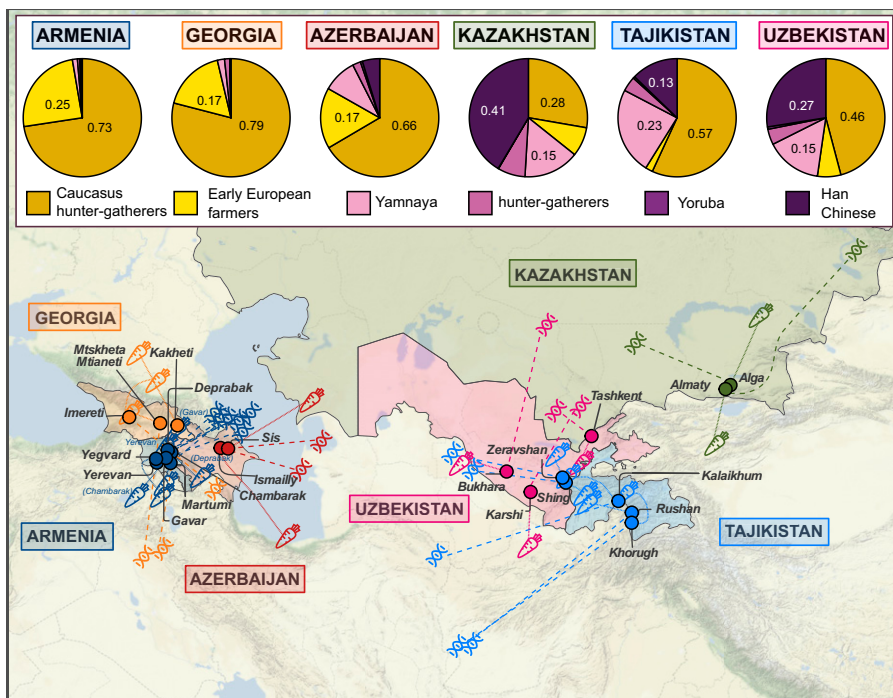


Fig. 1. Genetic and culinary affinities along the Silk Road. Map showing the geographic location of the cities (dots) from the six Silk Road countries sampled in this study (dots are colored according to the country); pseudocoordinates of each city from genetic (DNA icon) and food preference distances (carrot icon) are also shown. Where the overlapping of shapes and lines makes it difficult to distinguish precise locations and labels, we wrote the name of the city between brackets close to its respective icon (see also *SI Appendix*, Fig. S4 for a clearer representation). Within the inset, pie charts representing the ancestral genetic components for each country as inferred with a supervised admixture analysis. The genetic components at the individual level are reported in *SI Appendix*, Fig. S1A and have been averaged by country in order to compute the pie percentages.

were not randomly distributed between countries (*Material and Methods* and *SI Appendix*, Fig. S3), thus potentially pointing toward cultural information by way of shared and systematic dietary habits among the countries. For this reason, we replaced them with the minimum value of the questionnaire, which is 1, indicating that the subjects never ate or experienced a particular food.

Then, we delved into the relationship between the genetic, food preference, and geographic patterns by projecting genetic and dietary data onto the geographic coordinates with SpaceMix (60) in Fig. 1 (see *SI Appendix*, Fig. S4 for a more detailed representation), thus obtaining for each sampled city their genetic and dietary pseudocoordinates. The dietary pseudocoordinates are significantly less scattered than the genetic ones and gravitate generally closer to their geographic sampling point (Wilcoxon P value for dietary vs. genetic distances = $8.47e-8$; Wilcoxon effect size = 0.743). Azerbaijan is an exception to this trend, with its dietary preferences and, to a lesser extent also its genes, moving it toward the east. We noticed that the genetic pseudocoordinates of eastern countries tend to strikingly depart from the actual geographic positions: While the East Asian component of Kazakhstan and Uzbekistan pushes them roughly in the same direction, the reduced amount of this component in Tajikistan, which is replaced by higher proportions of Caucasus hunter-gatherer and Yamnaya signatures, draws it toward the Caucasus. Indeed, both geographic and genetic longitudes are strongly correlated with East Asian ancestry proportions (Pearson correlations of 0.83 and 0.95, respectively). Conversely, the genetic pseudocoordinates of western countries tend to match the geographic ones, with, once again, the exception of Azerbaijan.

Cultural Clustering of Food Preferences. We started by visually exploring the individuals' food preference profiles in a two-dimensional subspace by using different dimensionality reduction techniques (multidimensional scaling, principal component analysis (PCA), t -distributed stochastic neighbor embedding, and Uniform Manifold Approximation and Projection, *SI Appendix*, Fig. S5), whose results were found to be highly correlated (mean Procrustes correlation of 0.86; see *Materials and Methods*). In all analyses, the country of origin seems to be a great driver of sample scattering, with eastern and western countries stretching along the first dimension. Indeed, focusing specifically on the PCA, we found that the first component is strongly correlated with longitude ($R = 0.83$, $P = 3e-06$), while the second and the third are weakly related to age and sex ($R = 0.41$, $P < 2.2e-16$ and Wilcoxon test $P = 1.2e-15$; *SI Appendix*, Fig. S6).

As expected given the continuous nature of interindividual and cross-cultural food preferences, neat groups are not detectable. Hence, we delved into the peculiar food preferences contributing to the east-west stretching along the first PCA component, by relying on soft clustering techniques. Instead of grouping the items so that each item is only assigned to one cluster, soft clustering algorithms perform probabilistic clustering by quantifying the likelihood of a given individual belonging to any of the inferred clusters, with such probabilities being then interpreted as admixture proportions. To do so, we used the discriminant analysis of principal components (DAPC) technique (see *Materials and Methods* (61)). This method, originally designed to identify groups of genetically related individuals, can be applied to any quantitative data to infer the presence of different clusters and describe each individual as a combination of such clusters, thus allowing for mixed memberships. According to the Bayesian Information Criterion (BIC) evaluation (*SI Appendix*, Fig. S7),

we described the participants based on their probabilities of assignment to five distinct clusters (see *Materials and Methods*; we also reported the results for four, six, and seven clusters in *SI Appendix*, Figs. S8–S10).

We ran DAPC on individual food preferences and explored the inferred clusters' probabilities of assignment at the individual level (*SI Appendix*, Fig. S1C). We observed that 13% of them showed mixed memberships, with the probability of assignment for the second most likely cluster being at least 10% (*SI Appendix*, Fig. S1C).

In order to quantify food preference variability at population and individual levels and to compare it with genome-wide variability, we decomposed the overall dietary variance into between countries, between and within cities, and among individuals for each cluster distribution of assignments. We used the analysis of molecular variance (AMOVA) (62) to partition the genetic variance into the same layers (see *Materials and Methods*). As it happens for genetic variation, the highest fraction of the cultural cluster variance is explained by interindividual preferences ("Among individuals" layer, Fig. 2). Nevertheless, while the other variance partitions at the genetic level are negligible, the "Between countries" layer accounts for 20% of the overall variance on average, with cluster 1 reaching up to 38% (Fig. 2). Conversely, the low variance within and between cities suggests that food preferences are somewhat shared among people from the same location.

Notably, at the city level, the cultural admixture analysis shows a neat separation between Georgia and Armenia on one side and Azerbaijan, Kazakhstan, Tajikistan, and Uzbekistan on the other (Fig. 3A and *SI Appendix*, Table S1), thus partially mirroring the east-west cline highlighted by the PCA (*SI Appendix*, Fig. S5A). In particular, while clusters 1 (light blue) and 3 (red) seem predominant in eastern countries, as well as in Azerbaijan, clusters 2 (yellow), 4 (green), and 5 (blue) reach higher proportions in Armenia and Georgia. At the same time, cities within the same country are generally homogeneous for what concerns their mean individual cluster proportions. Notably, the average proportions of clusters 4 and 5, which are widely present in western countries, reach 30% in Bukhara (Uzbekistan), as well as in the two Kazakh cities (Alga and Almaty, Fig. 3A).

In order to pinpoint the main determinants of the inferred clustering, we explored individual assignments to the clusters by comparing intercluster food preferences (Fig. 3B and C), as well as considering other cultural aspects (religion), and

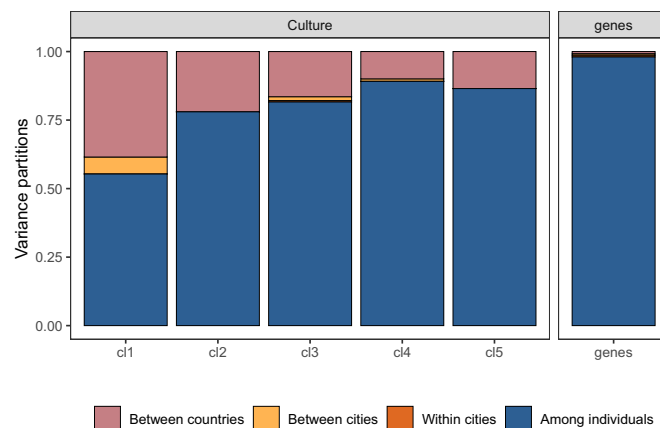


Fig. 2. Cultural and genetic variance partitions. Culture: Food-based variance layers for each distribution of cluster (cl) assignments were obtained through a linear mixed effect analysis. Genes: Genetic variance layers were obtained via AMOVA.

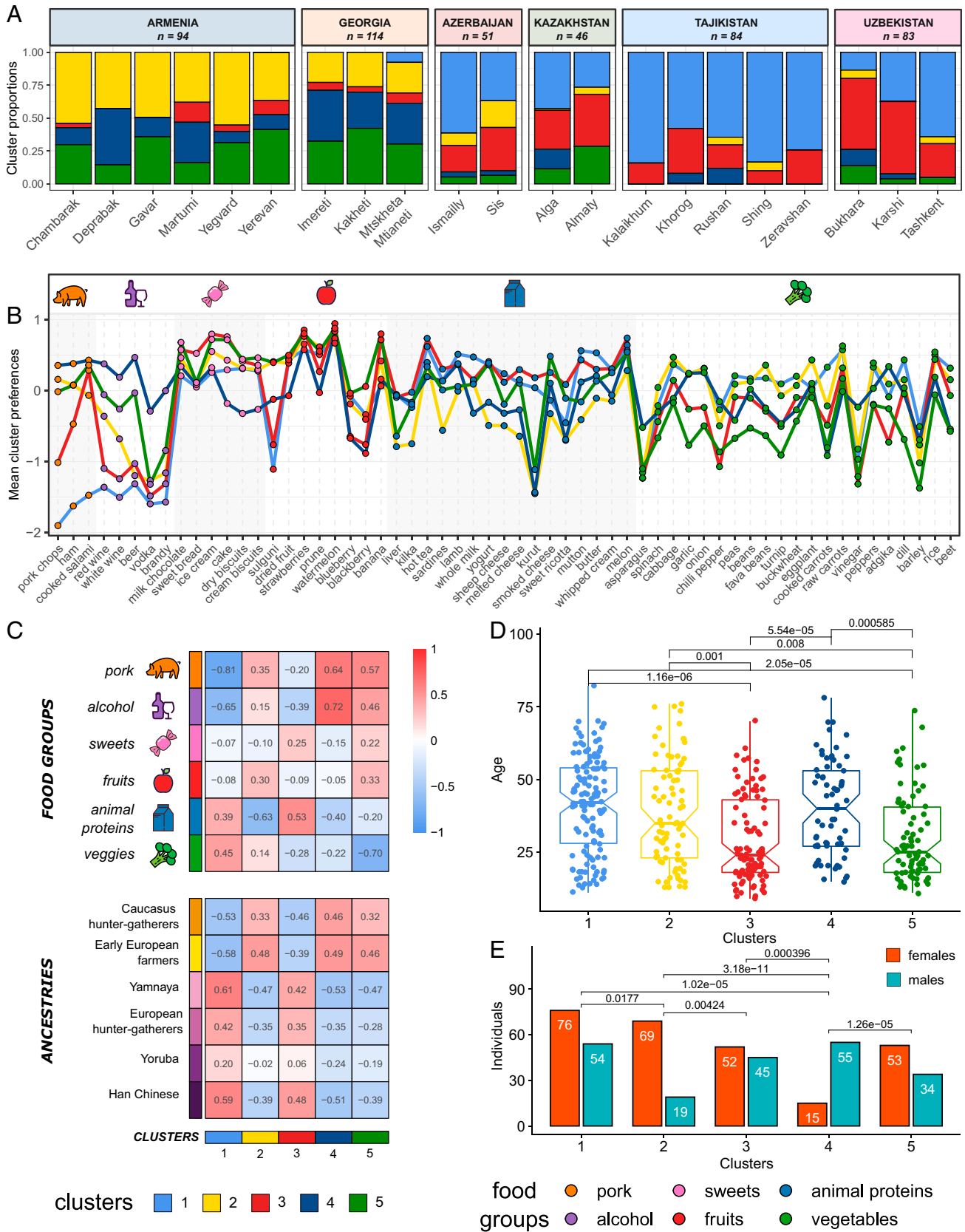


Fig. 3. DAPC clustering analyses with five clusters. (A) DAPC clustering analyses with $K = 5$ reporting the inferred clusters' probabilities of assignment averaged on the city of origin. (B) Mean food preference values of individuals assigned to each cluster (line and dots colors refer to clusters of individuals and food groups, respectively; lines connecting each food and each food group are for illustrative purposes only). Only the foods whose mean food preferences were significantly different (after Bonferroni correction) between clusters are shown. (C–E) Statistical analyses exploring the putative associations of the inferred clustering patterns with known individual information: food groups and genetic ancestral components (C, Spearman correlation), age (D, Bonferroni-adjusted P values refer to pairwise Wilcoxon test), and sex (E, Bonferroni-adjusted P values refer to pairwise χ^2 test).

factors potentially modulating the interindividual differences in food choices such as age and sex, as well as genetic ancestry (Fig. 3 C–E).

Both clusters 1 and 3, which are enriched in eastern countries and Azerbaijan, are negatively correlated with pork and alcohol preference ($R = -0.81$ and -0.65 for cluster 1 and $R = -0.20$ and $R = -0.39$ for cluster 3, Fig. 3 B and C), a pattern that, given the Muslim majority within such countries (63), may be at least partially explained by Islamic dietary restrictions (Pearson correlation between per-country Muslim proportion and clusters 1 and 3 assignments of 0.98, $P = 0.00047$). The compelling religious guidelines on dietary choices are mirrored by the higher fraction of variance explained by between-country variability for cluster 1 than for the others; in this case, religious constraints prevail on personal preferences, thus explaining almost 40% of the total variance (Fig. 2). Clusters 1 and 3 are also associated with greater preferences for animal protein foods, with cluster 3 showing a more robust correlation ($R = 0.39$ and 0.53 for clusters 1 and 3, respectively). This observation may partly be explained by the pastoral economies of local populations, here proxied by genetics, with Yamnaya and Han Chinese ancestral components being correlated with both clusters (Fig. 3C). Typical products from regional cuisines, such as the Georgian cheese *sulguni* and the steppe dairy-based product *kurut*, concur in strengthening the otherwise religion-based separation (Fig. 3B).

Moreover, clusters 1 and 3 differ in the age of the participants, with cluster 3 individuals being significantly younger than in cluster 1 (Fig. 3D, adjusted $P = 1.2e-06$). This trait may also explain both the greater preference for sweets and the aversion for vegetables of cluster 3 subjects (Fig. 3C).

On the other side, clusters 2, 4, and 5 are represented mainly in the Georgian and Armenian cities. Notably, being clear of religious-based food restrictions, they exhibit higher interindividual variability (Fig. 2). For instance, cluster 2 is significantly enriched with women (Fig. 3D) who prefer wine but dislike beer or spirits (e.g., brandy and vodka). Moreover, belonging to these clusters seems barely correlated with sweets and vegetable preferences and strongly anticorrelated with animal protein foods (Fig. 3C). Conversely, cluster 4 is composed mainly of men (Fig. 3D), for which we report a strong correlation with alcohol ($R = 0.72$) and pork ($R = 0.64$) and a reduced preference for the other food groups. The last cluster (cluster 5) is again represented by younger subjects, who, differently from cluster 3, prefer pork, alcoholic beverages (with lower scores for spirits), sweets, and fruits but strongly dislike vegetables ($R = -0.70$, Fig. 3C). Notably, a DAPC analysis of the permuted dataset did not lead to either country-specific distributions or strong or significant associations with food groups, age, and sex of the subjects (SI Appendix, Fig. S11), thus supporting our approach. As a partial control for sampling bias, we report that the DAPC results at the city level obtained for the downsampled and sex- and age-matched dataset highly correlated with those of the original dataset (SI Appendix, Fig. S12).

Finally, we compared the DAPC results obtained for increasing number of clusters by using the nonnegative least square function which decomposes each cluster's probabilities of assignment for the run with $k + 1$ clusters as a combination of the results obtained with k clusters (SI Appendix, Fig. S13), thus highlighting the relationships between the various DAPC iterations. In particular, we noticed that the addition of each cluster offers an increasingly fine-grained view of our results without impairing the previous DAPC run (SI Appendix, Fig. S13). For instance, cluster 5 originates mainly from clusters 3 (young,

mostly female subjects in eastern countries) and 4 (mostly male subjects in western countries) of the previous run (SI Appendix, Fig. S8), with no contributions from clusters 1 and 2. Moreover, we also confirm that the religious-driven food preferences are strong drivers of the clustering process, with proportions almost unaffected by the growing k .

Discussion

With our work, we summarized 78 cultural variables to extract individual probabilities of assignments to discrete cultural clusters. When we looked at the individual level, many subjects were probabilistically assigned to more than one cluster. The soft clustering approach may then be interpreted as a mixed membership, thus hinting at complex cultural dynamics regarding dietary preferences. Such admixed assignments mirror what is observed from a genetic point of view, where we confirmed the previously described (56) complex history of migrations and admixtures characterizing these populations. The same patterns of demographic movements have also brought together different cultural habits from every corner of the Silk Road. Taking both genetic and cultural variables into account, our approach aimed at describing population structure and admixture at both levels, as well as highlighting dietary, social, and genetic patterns shared across these lands.

More specifically, we found different layers of cultural stratification, possibly reflecting religious dietary restrictions, age-related personal preferences, and group traditions. For instance, the first and second cultural clusters (“pork” and “alcohol”) are mirrored by a geographic and genetic west–east cline, with the exception of Azerbaijan (Fig. 3A). In this case, we propose that the consumption of pork and alcoholic drinks is determined by recent cultural shifts, driven by the extensive Islamization of Central Asia starting from 751 AD (55), not matched by a genetic signature. This layer of complexity is further highlighted by the reduced interindividual variability (for cluster 1 when compared to other clusters and genetics; Fig. 2), as well as the presence of outlier cluster components within the country at the city level (Fig. 3A). For example, the city of Bukhara in Uzbekistan showed a different level of clustering than other villages in the same country, possibly reflecting its history as an important economic and cultural center along the Silk Road, which attracted people from far and wide. Similar behavior can be seen in Alga and Almaty in Kazakhstan, a country hosting a wide variety of ethnic groups with different religions. Indeed, while Islam is the most commonly practiced religion, about 26% of the Kazakh population identifies as Christian (64), with the city of Almaty itself hosting the main Russian Orthodox Church in Kazakhstan, the Ascension Cathedral. This shows how cultural influences could sharply differentiate these variables in multicultural populations even if genetic variation does not immediately reflect such dynamics. Nevertheless, although this pattern seems supported even when we plotted the downsampled and sex- and age-matched by country dataset (SI Appendix, Fig. S12), we stress here that such compositions may also reflect a sampling bias among countries and cities. Both clusters 1 and 3 showed also a greater preference for foods rich in dairy and meat proteins, a pattern that may be influenced by deep (Bronze Age–related introduction of pastoralism) and more recent historical influences over the sampled populations (Fig. 3C). This link could also hint at a more profound relationship, whose roots may be intertwined with genetics and the environment, as the lactose tolerance allele was brought to this population from the steppe during

the Bronze Age (65). We stress here that, while such results seem specific to the geographic areas analyzed in this work with roots that, at least for the genetic ancestral components, go back to ancient times (66, 67), the formation of these dietary patterns may also be the result of historical and modern events.

With religion-based dietary restrictions fading in the remaining clusters, other features of the participants, partially related to culture, emerged, such as age and sex (Fig. 3 *D* and *E*). For instance, the role of biological sex and its social perception may guide a person's dietary choices by generating specific food preference patterns, such as the high correlation with alcoholic beverages in cluster 4, which consists mainly of men.

In conclusion, our study provides a perspective on the intertwined relationship between cultural and genetic structures and admixture by quantitatively summarizing an individual culture in vectors of probabilities, similarly to what is done in genetic ancestry analyses. In particular, it proved to be a valuable resource to highlight the effects on cultural patterns of peculiar events in recent timescale, whose light footprints genetics fails to retrace, and to provide a compact and tractable means of including cultural variables as a covariate in future association studies.

Materials and Methods

Sample Study. This study analyzed the food preference questionnaires for 79 foods of 543 individuals (310 female and 233 male subjects ranging from 9 to 82 y). Such data have been collected as part of the scientific expedition Marcopolo 2010 (www.marcopolo2010.it), undertaken with the aim of analyzing individual and regional-specific variations in the perception of the human senses (e.g., taste, smell, hearing, and vision) across six countries (Georgia, Armenia, Azerbaijan, Uzbekistan, Kazakhstan, and Tajikistan) along the Silk Road (57, 68). Genetic samples were also collected with lifestyle and habit information for a subset of the individuals participating in the study. More information on the project and technical details about the sampling and genotyping steps can be found in (56). The data that support the findings of this study are available upon request.

Each participant in the study provided written informed consent. The research protocol had been approved by the ethical committee of IRCCS-Burlo Garofolo Hospital.

Data Preprocessing and Food Clustering. The dataset we worked on was initially designed to explore the main genetic factors leading to individual food preferences. The participants in the study were asked to rate 79 foods on a 5-point scale, which ranges from 1 ("dislike extremely") to 5 ("like extremely"). The option "never tasted" was also included as an alternative answer (68).

We removed five individuals whose food preferences profiles had an SD equal to zero (they expressed the same rate to all the foods) or they were completely missing, and the food "artichokes" because more than half of the subjects did not provide an answer for it, thus resulting in a dataset of 78 food preferences over 538 individuals. We will refer to this dataset as the "raw" dataset.

We built the correlation matrix of "raw" food preferences, and we used the `scipy.cluster.hierarchy.linkage()` function on this matrix (method = "average", metric = "euclidean") to hierarchically cluster the food items (the "average" method is also known as the unweighted pair group method with arithmetic mean algorithm, UPGMA). With the function `scipy.cluster.hierarchy.fcluster()` we selected a maximum of six clusters (criterion = "maxclust"), which, according to their internal composition were called "alcohol," "pork," "sweets," "vegetables," "animal proteins," and "fruits."

Finally, we computed the Cronbach's alpha for each of the food clusters to measure their internal consistency with the formula

$$\alpha_{st} = \frac{N \cdot \bar{r}}{1 + (N - 1) \cdot \bar{r}},$$

where N is the number of items (foods) and r is their mean correlation. The Cronbach's alpha of the inferred food clusters was higher than 0.75, except for the

smallest one ("pork," which consisted of three foods with a Cronbach's alpha of 0.69; *SI Appendix, Table S2*), thus demonstrating the internal consistency of the clusters themselves.

Missing Data Processing. We reasoned that the "never tasted" options (coded in the original dataset as missing values) could point toward meaningful signals related to systematic differences in dietary choices between countries.

To experimentally confirm that, we performed a χ^2 test on the "missing" counts of each food in each country, with the idea of spotting the putative cultural signatures of shared dietary choices. Indeed, a χ^2 test on "missing" counts by countries pointed toward a significant dependence of "missing" hits from countries ($P = 1.22e-185$). By repeating a χ^2 test on "missing" counts for each food, we found 31 foods whose "missing" count was significantly different between countries ("country-associated missing foods," adjusted $P < 0.05$; *SI Appendix, Table S3*). Then, we investigated whether the distributions of "missing" hits across different foods were somewhat correlated together. By answering the question "Are there foods which are never tasted or skipped by the same individuals?," we wanted to pinpoint foods whose "missing" values could connect people. In order to do that, we built a correlation matrix on "missing" counts by using just the "country-associated missing foods" (*SI Appendix, Fig. S3A*) and the entire dataset (*SI Appendix, Fig. S3B*). We noticed that "missing" answers for alcoholic beverages (e.g., "brandy," "vodka," "beer," "red wine," and "white wine") and pork (e.g., "cooked salami," "pork chops," and "ham") preferences colocalized, which may suggestively indicate a cultural influence such as the adhesion to religious precepts.

Given that the "never tasted" answers seem to carry cultural information, we could not impute them with the mean of the feature, thus ignoring this information. Instead, we decided to give them the minimum value of the questionnaire, which is 1, because they indicate mainly individuals who never ate or experienced a particular food. Nevertheless, to get rid of the possibly present technical "missing" hits, we excluded 62 individuals with a "missing" ratio higher than 0.10 computed on the foods whose "missing" count was not associated with the countries (i.e., the foods outside the list of the 31 country-associated ones; *SI Appendix, Table S3*). Then, we normalized the individual profiles with the `StandardScaler()` function of the python library `scikit-learn`. We will refer to this dataset as the "processed" dataset.

As an additional control, we replaced the "never tasted" answers with the mid food preference value (3) and repeated the clustering analyses (*SI Appendix, Fig. S14*).

SpaceMix. We ran SpaceMix (60) to provide a pseudocoordinate relationship between the geographic, genetic, and dietary habits of the sampled individuals. We used as input the sampling longitudes and latitudes of each town, as well as the allele frequency of each single-nucleotide polymorphism (SNP) and the average food preferences. In particular, for the genetic allele frequencies, we selected the 130,000 SNPs with minor allele frequency (MAF) above 5% and genotyping rate of 100%. The average food preferences were computed for each food, at the city level, after first quantile-cutting into three levels (0, 1, and 2), the full range of answers (thus mirroring allelic dosage patterns) using the quantile-based discretization function `pandas.qcut`. The genetic and dietary pseudocoordinates outputted by SpaceMix were then plotted along with the actual geographic ones.

Finally, we computed the Euclidean distances between each geographic location and its genetic or dietary pseudocoordinates by using the R function `dist` (method = "euclidean", pairwise = T), and we compared the dietary and the genetic distances via the Wilcoxon rank sum test.

Dimensionality Reduction. We used the following dimensionality reduction techniques in order to spot general patterns of variability within the "processed" dataset by using the Python `scikit-learn` library (69): multidimensional scaling, PCA, t-distributed stochastic neighbor embedding (with `learning_rate = 50`), and Uniform Manifold Approximation and Projection for Dimension Reduction (with default parameters). Finally, we checked the overall consistency of results across the different techniques via the Procrustes analysis, which tests the similarity of two datasets by transforming the second one through scaling, dilation, rotations, and reflections. In detail, we computed the correlations in a symmetric Procrustes rotation by using the "protest" function within the `vegan` R package (70) on the first two dimensions of each dataset obtained through the different

techniques, and we reported the mean value across the six combinations of techniques.

Cultural Clustering. We used the DAPC technique from the *adegenet* R package (61) to infer cultural clusters of individuals on the “processed” datasets (where each row describes an individual and each column a food). DAPC is a multivariate technique initially designed to identify genetic clusters among individuals and describe the relationships between such clusters, although it can be used on any quantitative data (61). This method sequentially runs the algorithm *k*-means with increasing values of *k* on the principal components computed on the dataset for identifying a given *k* number of clusters. Then, discriminant analysis is performed to seek synthetic variables, called discriminant functions, which maximize the variance between clusters while minimizing variation within clusters.

As a first step, we used the *adegenet* function *find.clusters* on the “processed” dataset in order to infer the best number of clusters, k_s , describing it by retaining all the principal components, which represent the overall variability in our dataset. We analyzed the trend of the BIC, with the number of clusters ranging from 1 to 20. We chose to focus within the main text on the results by using five clusters, since further increases in *k* led to BIC improvements of less than 2 units. However, we also conducted the analyses exploring $k = 4$, $k = 6$, and $k = 7$, the latter being the *k* minimizing the BIC. Then, we used the inferred attributions to find the posterior cluster assignments with the *dapc* function. In order to compare the results obtained across different *k*s, we used the software *Distrupt for many K's* implemented in CLUMPAK (71) to align the DAPC across the different *k*s and the nonnegative least square function implemented in the R package *npls* (72) to model the run with $k + 1$ clusters as a combination of that with *k* groups.

Finally, in order to partially control for potential technical or sampling biases, we adopted the following approaches. First, we permuted the dataset by shuffling independently the columns, each containing a food preference distribution, and we repeated the DAPC analysis on the best supported number of clusters (five; see the BIC plot of *SI Appendix, Fig. S11*). Second, we selected the same number of subjects matched for sex and age (three categories: 9–24, 24–45, and 45–82 y old) from each country (the final dataset contained 144 individuals), and we plotted the DAPC results by using the downsampled and sex-/age-matched dataset.

Analysis of Variance and AMOVA. Total variance of each distribution of cluster assignments was partitioned into “Between countries,” “Between cities,” “Within cities,” and “Among individuals.” In order to do so, we performed a linear mixed-effects analysis for each distribution of cluster assignments by using the function *lmer* implemented within the R package *lme4* (73).

We used an AMOVA (62) to quantify the variation between countries, cities, and individuals at the genetic level with the R package *poppr* (74) on the 130,000 SNPs with MAF above 5% and genotyping rate of 100%.

Genetic Analyses. Genome-wide data for 841 individuals (whom we will call the “Silk-Road dataset”) were produced as part of the above-mentioned Marcopolo 2010 scientific expedition (56, 58). We removed related individuals by excluding one of each individual pair showing an identity by descent value higher than 0.25 with PLINK 1.9 (75). After genotype quality controls and removal of individuals and variants with missing call rate exceeding 1% (–mind

and –geno 0.01 with PLINK), the final dataset consisted of 697 individuals, of which 350 were also sampled for their food preference, and 598,612 SNPs. We merged the Silk Road dataset with the populations Utah Residents with Northern and Western European Ancestry (CEU); Han Chinese in Beijing, China (CHB); and Punjabi from Lahore, Pakistan (PJI), from the 1000 Genomes Project (76), after removing second-degree and higher relatives among them (identity by descent > 0.25). We further excluded genetic variants on sexual chromosomes or with MAF lower than 1% and those exhibiting extreme deviations from Hardy-Weinberg equilibrium (–hwe 0.00001).

Finally, we pruned the variants for linkage disequilibrium (–indep-pairwise 200 50 0.4), and we evaluated the patterns of shared ancestries among populations with the ADMIXTURE software via a cross-validation approach to infer the best number of ancestral components.

In order to time-transsect the inferred ancestral components, we further merged the dataset with samples broadly related to hunter-gatherers, early farmers, Iranian Neolithic/Caucasus hunter-gatherers, and Yamnaya groups from the 1,240,000 dataset from David Reich laboratory, which comprises present-day and ancient DNA data (compiled from published papers up to March 2020). Samples codes and label correspondences are shown in *Dataset S1*. Also, we added the Yoruba in Ibadan, Nigeria (YRI) population from 1000 Genomes Project for further comparison. We removed individuals with more than 90% missing rate and the final merged dataset had a total of 38,069 SNPs with genotyping rate of 90%. We used this last dataset to perform a supervised admixture analysis using hunter-gatherers; early farmers; Iranian Neolithic/Caucasus hunter-gatherers; Yamnaya; CHB; and YRI populations as references.

Data, Materials, and Software Availability. DAPC results (individual assignment probabilities, supervised admixture, and dimensionality reduction results) data have been deposited in GitHub [https://github.com/serena-aneli/silk_road_cultural_admixture (77); https://github.com/serena-aneli/silk_road_cultural_admixture/tree/main/data (78)]. Previously published data were used for this work (56, 68).

ACKNOWLEDGMENTS. The work was supported by STARs@UniPD 2019 (S.A., L.P.).

Data Sharing. The data supporting the findings of this study have been generated in previous publications [genetic SNP array data from (56); food preferences from (68)]. Both datasets were obtained and are available from P.G. (paolo.gasparini@burlo.trieste.it). Codes for this work can be found at the link: https://github.com/serena-aneli/silk_road_cultural_admixture. DAPC results (individual assignment probabilities, supervised admixture, and dimensionality reduction results) are available on GitHub at the link: https://github.com/serena-aneli/silk_road_cultural_admixture/tree/main/data.

Author affiliations: ^aDepartment of Biology, University of Padova, 35121 Padua, Italy; ^bDepartment of Public Health Sciences and Pediatrics, University of Turin, 10126, Turin, Italy; ^cInstitute for Maternal and Child Health IRCCS Burlo Garofolo, 34137 Trieste, Italy; ^dHuman Ecology and Archaeology (HUMANE), IMF, CSIC, 08001 Barcelona, Spain; ^eDepartment of Cultural Heritage, University of Bologna, 48121 Ravenna, Italy; ^fHuman Technopole, 20157 Milan, Italy; ^gDepartment of Medicine, Surgery and Health Sciences, University of Trieste, 34149 Trieste, Italy; ^hDepartment of Clinical and Biological Sciences, University of Torino, 10100 Torino, Italy; and ⁱEstonian Biocentre, Institute of Genomics, University of Tartu, 50090 Tartu, Estonia

1. L. L. Cavalli-Sforza, M. W. Feldman, *Cultural Transmission and Evolution. A Quantitative Approach* (Princeton University Press, 1981).
2. R. Boyd, P. J. Richerson, *Culture and the Evolutionary Process* (University of Chicago Press, 1985).
3. J. Henrich, R. McElreath, The evolution of cultural evolution. *Evol. Anthropol.* **12**, 123–135 (2003).
4. A. Mesoudi, *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences* (University of Chicago Press, 2011).
5. L. Fogarty, A. Kandler, The fundamentals of cultural adaptation: Implications for human adaptation. *Sci. Rep.* **10**, 14318 (2020).
6. S. Carrignon, R. A. Bentley, D. Ruck, Modelling rapid online cultural transmission: Evaluating neutral models on Twitter data with approximate Bayesian computation. *Palgrave Commun.* **5**, 83 (2019).
7. A. Kandler, E. R. Crema, “Analysing cultural frequency data: Neutral theory and beyond” in *Handbook of Evolutionary Research in Archaeology*, A. Prentiss, Ed. (Springer, Cham, 2019), pp. 83–108.
8. D. Ruck, R. Alexander Bentley, A. Acerbi, P. Garnett, D. J. Hruschka, Role of neutral evolution in word turnover during centuries of English word popularity. *Adv. Complex Syst.* **20**, 1750012 (2017).
9. E. R. Crema, A. Kandler, S. Shennan, Revealing patterns of cultural transmission from frequency data: Equilibrium and non-equilibrium assumptions. *Sci. Rep.* **6**, 39122 (2016).
10. P. E. Smaldino, R. McElreath, The natural selection of bad science. *R. Soc. Open Sci.* **3**, 160384 (2016).
11. S. J. Lycett, Cultural evolutionary approaches to artifact variation over time and space: Basis, progress, and prospects. *J. Archaeol. Sci.* **56**, 21–31 (2015).
12. P. E. Savage, S. Brown, E. Sakai, T. E. Currie, Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8987–8992 (2015).
13. A. Acerbi, A. R. Bentley, Biases in cultural transmission shape the turnover of popular traits. *Evol. Hum. Behav.* **35**, 228–236 (2014).
14. R. A. Bentley, C. C. S. Caiado, P. Ormerod, Effects of memory on spatial heterogeneity in neutrally transmitted culture. *Evol. Hum. Behav.* **35**, 257–263 (2014).
15. L. S. Premo, Cultural transmission and diversity in time-averaged assemblages. *Curr. Anthropol.* **55**, 105–114 (2014).
16. J. Steele, C. Glatz, A. Kandler, Ceramic diversity, random copying, and tests for selectivity in ceramic production. *J. Archaeol. Sci.* **37**, 1348–1358 (2010).
17. A. R. Bentley, S. J. Shennan, Cultural transmission and stochastic network growth. *Am. Antiq.* **68**, 459–485 (2003).

18. C. O. Brand, A. Mesoudi, Prestige and dominance-based hierarchies exist in naturally occurring human groups, but are unrelated to task-specific knowledge. *R. Soc. Open Sci.* **6**, 181621 (2019).
19. P. Turchin *et al.*, Quantitative historical analysis uncovers a single dimension of complexity that structures global variation in human social organization. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E144–E151 (2018).
20. A. B. Migliano *et al.*, Characterization of hunter-gatherer networks and implications for cumulative culture. *Nat. Human Behav.* **1**, 0043 (2017).
21. D. Smith *et al.*, Cooperation and the evolution of hunter-gatherer storytelling. *Nat. Commun.* **8**, 1853 (2017).
22. T. E. Currie, S. J. Greenhill, R. D. Gray, T. Hasegawa, R. Mace, Rise and fall of political complexity in island South-East Asia and the Pacific. *Nature* **467**, 801–804 (2010).
23. J. Fort, Synthesis between demic and cultural diffusion in the Neolithic transition in Europe109. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18669–18673 (2012).
24. J. Steele, "Innovation Diffusion and Travelling Waves" in *Pattern and Process in Cultural Evolution*, S. J. Shennan, Ed. (University of California Press, London, 2009), pp. 163–174.
25. K. Basava, H. Zhang, R. Mace, A phylogenetic analysis of revolution and afterlife beliefs. *Nat. Hum. Behav.* **5**, 604–611 (2021).
26. H. Zhang, T. Ji, M. Pagel, R. Mace, Dated phylogeny suggests early Neolithic origin of Sino-Tibetan languages. *Sci. Rep.* **10**, 20792 (2020).
27. A. Mesoudi, Migration, acculturation, and the maintenance of between-group cultural variation. *PLoS One* **13**, e0205573 (2018).
28. E. R. Crema, T. Kerig, S. Shennan, Culture, space, and metapopulation: A simulation-based study for evaluating signals of blending and branching. *J. Archaeol. Sci.* **43**, 289–298 (2014).
29. L. S. Premo, Local extinctions, connectedness, and cultural evolution in structured populations. *Adv. Complex Syst.* **15**, 1150002–1150020 (2012).
30. R. D. Gray, D. Bryant, S. J. Greenhill, On the shape and fabric of human history. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 3923–3933 (2010).
31. G. J. Ackland, M. Sznitzler, K. Stratford, M. H. Cohen, Cultural hitchhiking on the wave of advance of beneficial technologies. *Proc. Natl. Acad. Sci.* **104**, 8714–8719 (2007).
32. M. Collard, S. J. Shennan, J. Tehrani, Branching, blending and the evolution of cultural similarities and differences among human populations. *Evol. Hum. Behav.* **27**, 169–184 (2006).
33. A. Acerbi, J. Kendal, J. J. Tehrani, Cultural complexity and demography: The case of folktales. *Evol. Hum. Behav.* **38**, 474–480 (2017).
34. J. Henrich *et al.*, Understanding cumulative cultural evolution. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E6724–E6725 (2016).
35. L. S. Premo, Effective population size and the effects of demography on cultural diversity and technological complexity. *Am. Antiq.* **81**, 605–622 (2016).
36. A. Powell, S. Shennan, M. G. Thomas, Late Pleistocene demography and the appearance of modern human behavior. *Science* **324**, 1298–1301 (2009).
37. J. Henrich, Demography and cultural evolution: Why adaptive cultural processes produced maladaptive losses in Tasmania. *Am. Antiq.* **69**, 197–121 (2004).
38. S. J. Shennan, A. R. Bentley, "Style, Interaction and Demography among the Earliest Farmers of Central Europe" in *Cultural Transmission and Archaeology. Issues and Case Studies*, M. J. O'Brien, Ed. (The SAA Press, Washington, 2008), pp. 164–177.
39. S. Shennan, Demography and cultural innovation: A model and its implications for the emergence of modern human culture. *Camb. Archaeol. J.* **11**, 5–16 (2001).
40. S. J. Greenhill, Do languages and genes share cultural evolutionary history? *Sci. Adv.* **7**, eabm2472 (2021).
41. H. Matsumae *et al.*, Exploring correlations in genetic and cultural variation across language families in northeast Asia. *Sci. Adv.* **7**, eabd9223 (2021).
42. A. J. M. Key, S. J. Lycett, Biometric variables predict stone tool functional performance more effectively than tool-form attributes: A case study in handaxe loading capabilities. *Archaeometry* **61**, 539–555 (2019).
43. F. Riede, C. Hoggard, S. Shennan, Reconciling material cultures in archaeology with genetic data requires robust cultural evolutionary taxonomies. *Palgrave Commun.* **5**, 55 (2019).
44. C. Posth *et al.*, Language continuity despite population replacement in Remote Oceania. *Nat. Ecol. Evol.* **2**, 731–740 (2018).
45. E. Bertolini *et al.*, Inferring patterns of folktale diffusion using genomic data. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 9140–9145 (2017).
46. N. Creanza *et al.*, A comparison of worldwide phonemic and genetic variation in human populations. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1265–1272 (2015).
47. S. Brown *et al.*, Correlations in the population structure of music, genes and language. *Proc. Biol. Sci.* **281**, 20132072 (2013).
48. Y. Itan, A. Powell, M. A. Beaumont, J. Burger, M. G. Thomas, The origins of lactase persistence in Europe. *PLoS Comput. Biol.* **5**, e1000491 (2009).
49. A. V. Bell, P. J. Richerson, R. McElreath, Culture rather than genes provides greater scope for the evolution of large-scale human prosociality. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17671–17674 (2009).
50. R. Pinhasi, N. von Cramon-Taubadel, Craniometric data supports demic diffusion model for the spread of agriculture into Europe. *PLoS One* **4**, e6747 (2009).
51. E. Gallagher, S. Shennan, M. G. Thomas, Food income and the evolution of forager mobility. *Sci. Rep.* **9**, 5438 (2019).
52. J. Henrich, N. Henrich, The evolution of cultural adaptations: Fijian food taboos protect against dangerous marine toxins. *Proc. Biol. Sci.* **277**, 3715–3724 (2010).
53. G. Eriksson *et al.*, Same Island, different diet: Cultural evolution of food practice on Öland, Sweden, from the Mesolithic to the Roman Period. *J. Anthropol. Archaeol.* **27**, 520–543 (2008).
54. P. Corvo, *Food Culture, Consumption and Society* (Springer, 2016).
55. R. N. Spengler, *Fruit from the Sands: The Silk Road Origins of the Foods We Eat* (University of California Press, 2020).
56. M. Mezzavilla *et al.*, Genetic landscape of populations along the Silk Road: Admixture and migration patterns. *BMC Genet.* **15**, 131 (2014).
57. A. Robino *et al.*, A population-based approach to study the impact of PROP perception on food liking in populations along the Silk Road. *PLoS One* **9**, e91716 (2014).
58. S. Ulivi, M. Mezzavilla, P. Gasparini, Genetics of eye colours in different rural populations on the Silk Road. *Eur. J. Hum. Genet.* **21**, 1320–1323 (2013).
59. M. Haber *et al.*, Genetic evidence for an origin of the Armenians from Bronze Age mixing of multiple populations. *Eur. J. Hum. Genet.* **24**, 931–936 (2016).
60. G. S. Bradburd, P. L. Ralph, G. M. Coop, Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution* **67**, 3258–3273 (2013).
61. T. Jombart, S. Devillard, F. Balloux, Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94 (2010).
62. L. Excoffier, P. E. Smouse, J. M. Quattro, Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
63. C. P. Hackett and B. J. Grim, Pew Forum on Religion & Public Life and Pew-Templeton Global Religious Futures Project. *The Global Religious Landscape: A Report on the Size and Distribution of the World's Major Religious Groups as of 2010/Pew Forum on Religion & Public Life* (Pew Research Center, Pew Forum on Religion & Public Life, 2012).
64. A. A. Smailov, *Results of the 2009 National Population Census of the Republic of Kazakhstan: Analytical Report* (Agency on Statistics of the Republic of Kazakhstan, 2011).
65. N. S. Enattah *et al.*, Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am. J. Hum. Genet.* **82**, 57–72 (2008).
66. L. Papani *et al.*, Genomic analyses inform on migration events during the peopling of Eurasia. *Nature* **538**, 238–242 (2016).
67. V. M. Narasimhan *et al.*, The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).
68. N. Pirastu *et al.*, Genetics of food preferences: A first view from Silk Road populations. *J. Food Sci.* **77**, S413–S418 (2012).
69. V. Pedregosa, Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
70. J. Oksanen, *et al.*, vegan: Community Ecology Package. *R package version 2.6–2* (2022).
71. N. M. Kopelman, J. Mayzel, M. Jakobsson, N. A. Rosenberg, I. Mayrose, Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191 (2015).
72. C. L. Lawson, R. J. Hanson, Back Matter. *Solving Least Squares Problems*, 312–337 (1995).
73. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using **lme4**. *J. Stat. Softw.* **67**, 1–48 (2015).
74. Z. N. Kamvar, J. F. Tabima, N. J. Grünwald, Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**, e281 (2014).
75. C. C. Chang *et al.*, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
76. 1000 Genomes Project Consortium *et al.*, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
77. S. Aneli, Impact of cultural and genetic structure on food choices along the Silk Road. GitHub. https://github.com/serena-aneli/silk_road_cultural_admixture. Deposited 5 August 2022.
78. S. Aneli, Impact of cultural and genetic structure on food choices along the Silk Road. GitHub. https://github.com/serena-aneli/silk_road_cultural_admixture/tree/main/data. Deposited 5 August 2022.