Supplementary Materials for

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

Serena Aneli[1,2], Massimo Mezzavilla[1,3], Eugenio Bortolini[4,5], Nicola Pirastu[6], Giorgia Girotto[3,7], Beatrice Spedicati[7], Paola Berchialla[8], Paolo Gasparini[3,7], Luca Pagani[1,9]

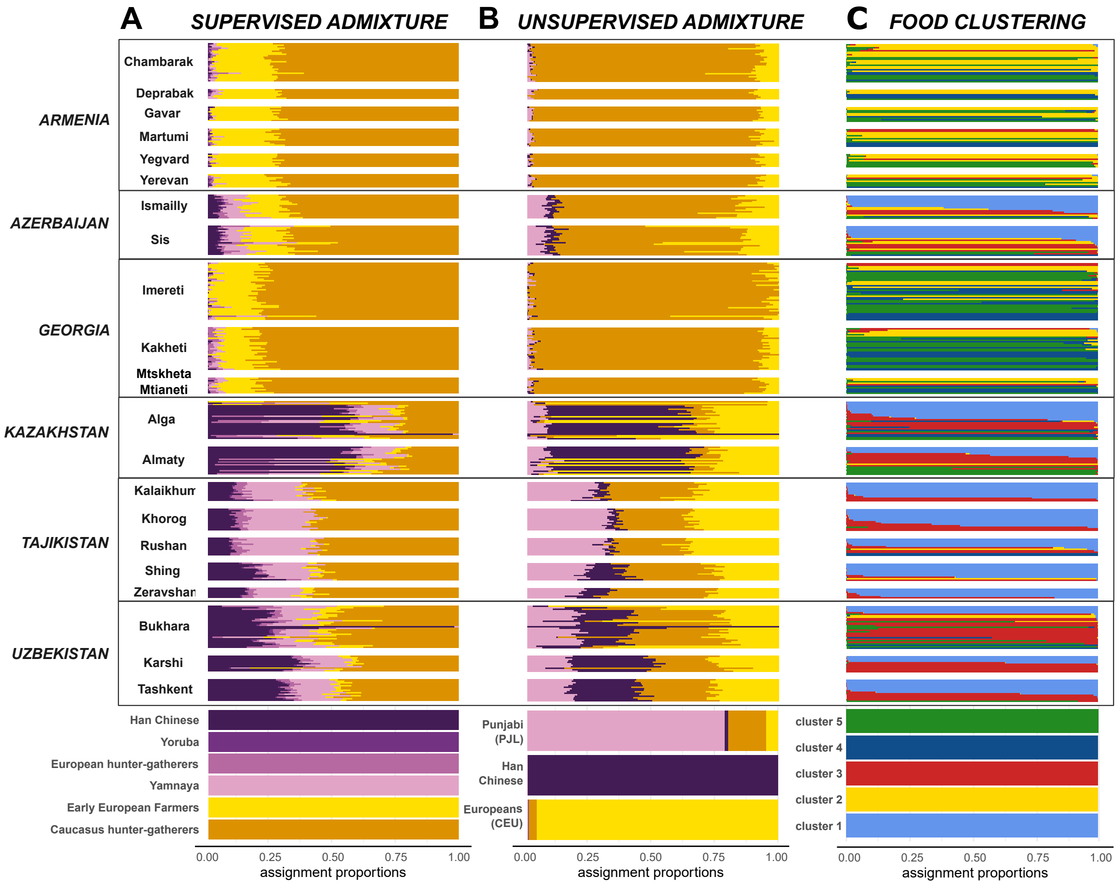
1 Department of Biology, University of Padova, Via Ugo Bassi, 58/B, 35131, Padua, Italy
2 Department of Public Health Sciences and Pediatrics, University of Turin, Turin, Italy
3 Institute for Maternal and Child Health IRCCS Burlo Garofolo, Via dell'Istria 65/1, 34137, Trieste, Italy
4 Human Ecology and Archaeology (HUMANE), IMF, CSIC, C/Egipciaques 15, 08001 Barcelona, Spain
5 Department of Cultural Heritage, University of Bologna, Via degli Ariani 1, 48121 Ravenna, Italy
6 Human Technopole, Milan, Italy
7 Department of Medicine, Surgery and Health Sciences, University of Trieste, 34149 Trieste, Italy
8 Department of Clinical and Biological Sciences, University of Torino, 10100 Torino, Italy
9 Estonian Biocentre, Institute of Genomics, University of Tartu, Riia 23b, 51010 Tartu, Estonia

This PDF file includes:

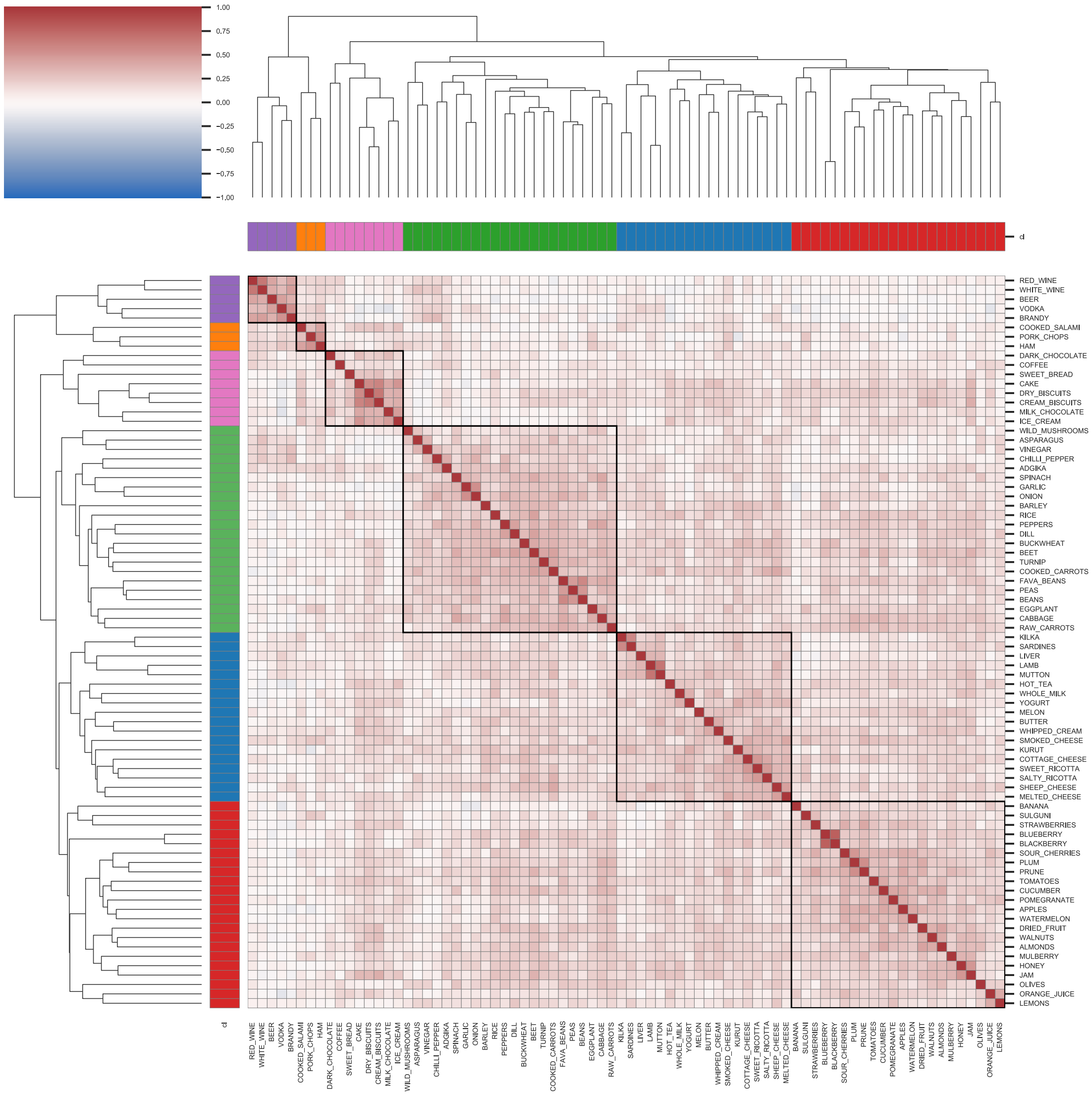
- Supplementary Figures 1 to 14.
- Supplementary Tables 1 to 3.

Other Supplementary Materials for this manuscript include:

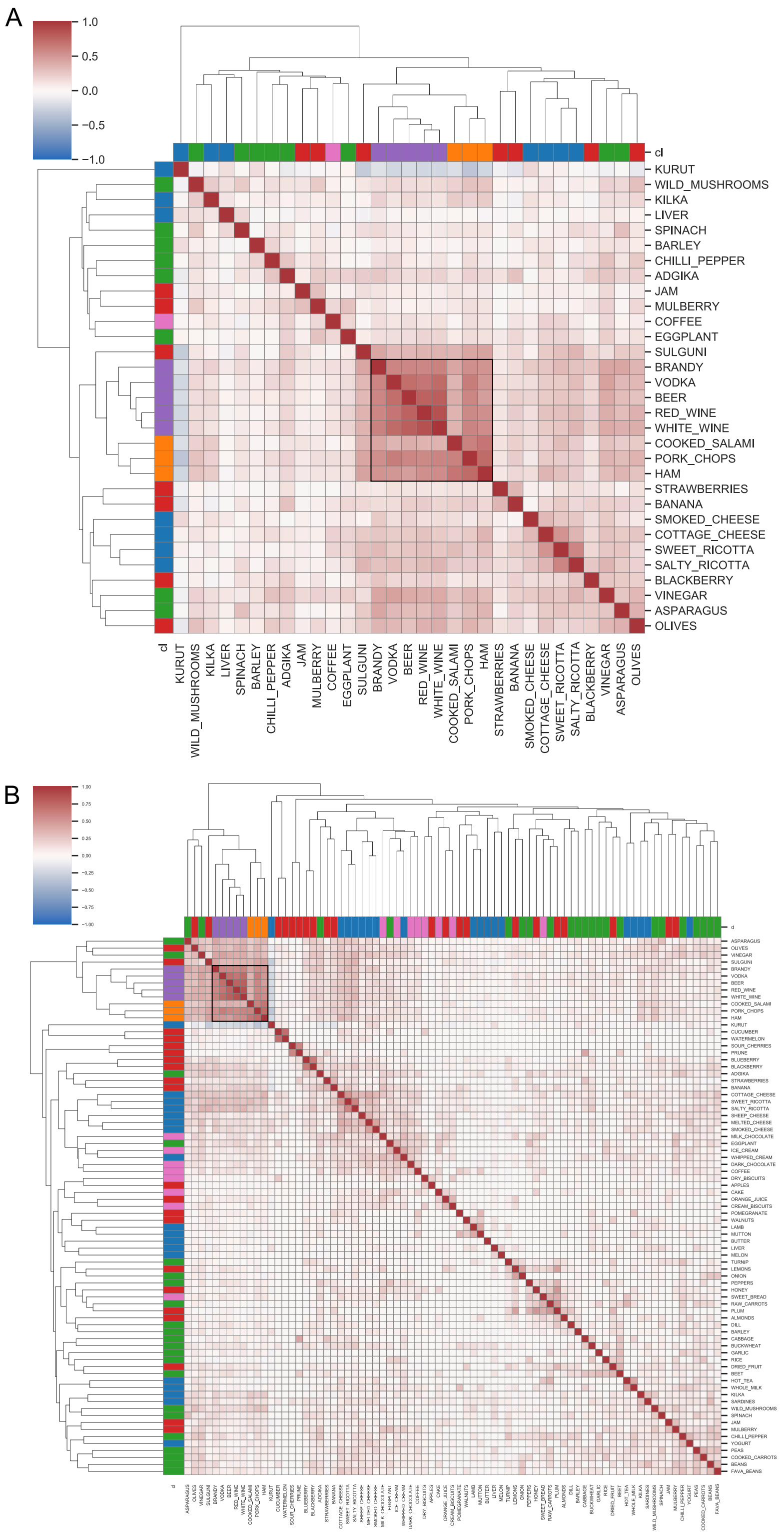
- Dataset S1. Population labels and individuals used in the supervised ADMIXTURE analysis as reference.
- DAPC results (individual assignment probabilities, supervised admixture and dimensionality reduction results), as well as the code used in this work, are available on GitHub (`https://github.com/serena-aneli/silk_road_cultural_admixture/tree/main/data`).



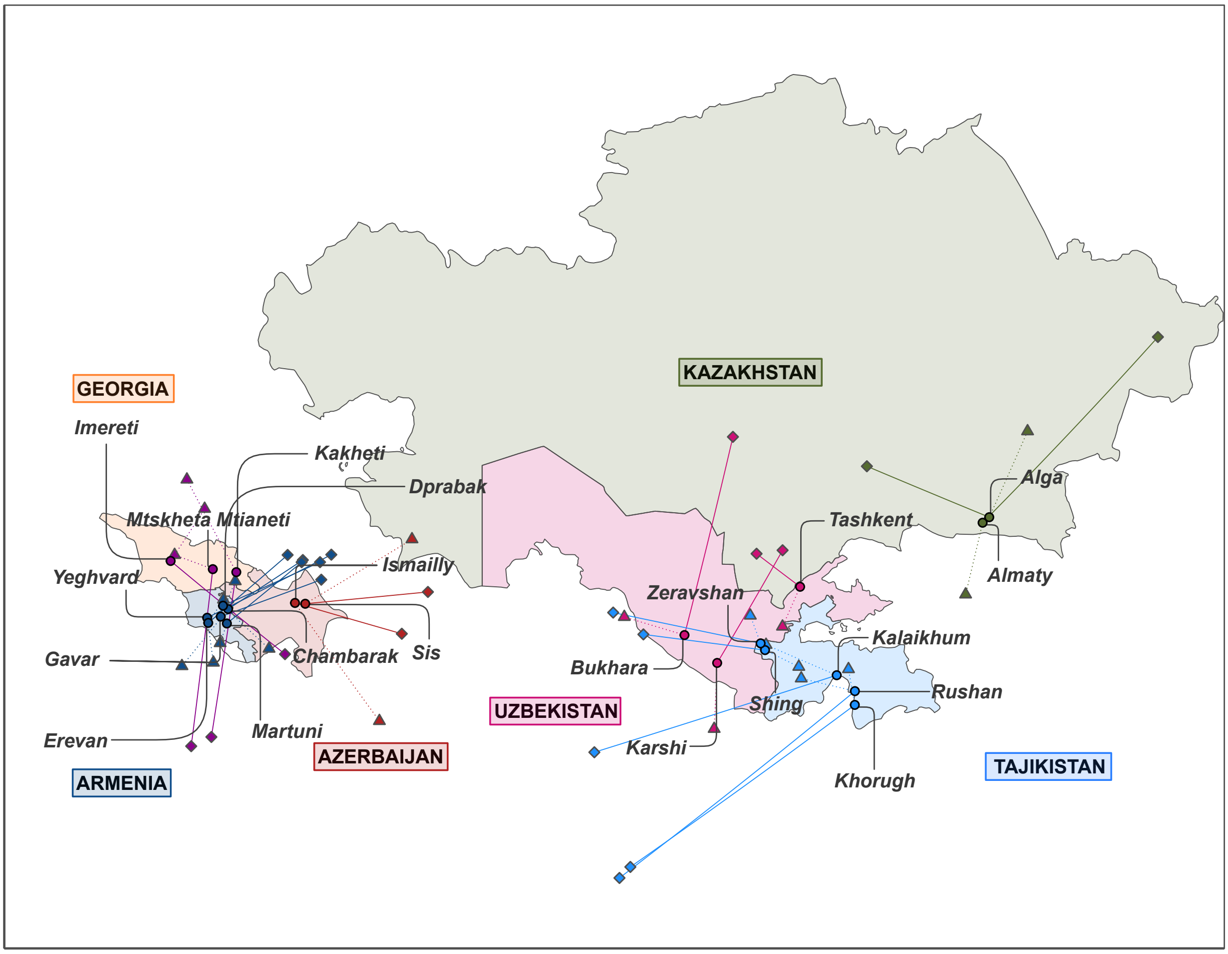
Supplementary Figure 1. Genetic and dietary structure. All panels report the clusters' proportions inferred from genetic (A and B) and food preferences data (C). In all panels, each row represents an individual in the same order. A) Supervised ADMIXTURE analysis of populations from Armenia, Azerbaijan, Georgia, Kazakhstan, Tajikistan and Uzbekistan. Two modern (Han Chinese and the African Yoruba population) and four ancient populations (European hunter-gatherers, Early European Farmers, Caucasus hunter-gatherers, and Yamnaya) are used as references. B) Unsupervised ADMIXTURE analysis of individuals from the six Silk Road countries and reference populations from 1000 Genomes Project (CEU: Utah residents (CEPH) with Northern and Western European ancestry; Han Chinese in Beijing, China; PJL: Punjabi in Lahore, Pakistan). C) DAPC clustering analyses, reporting the results for 5 clusters.



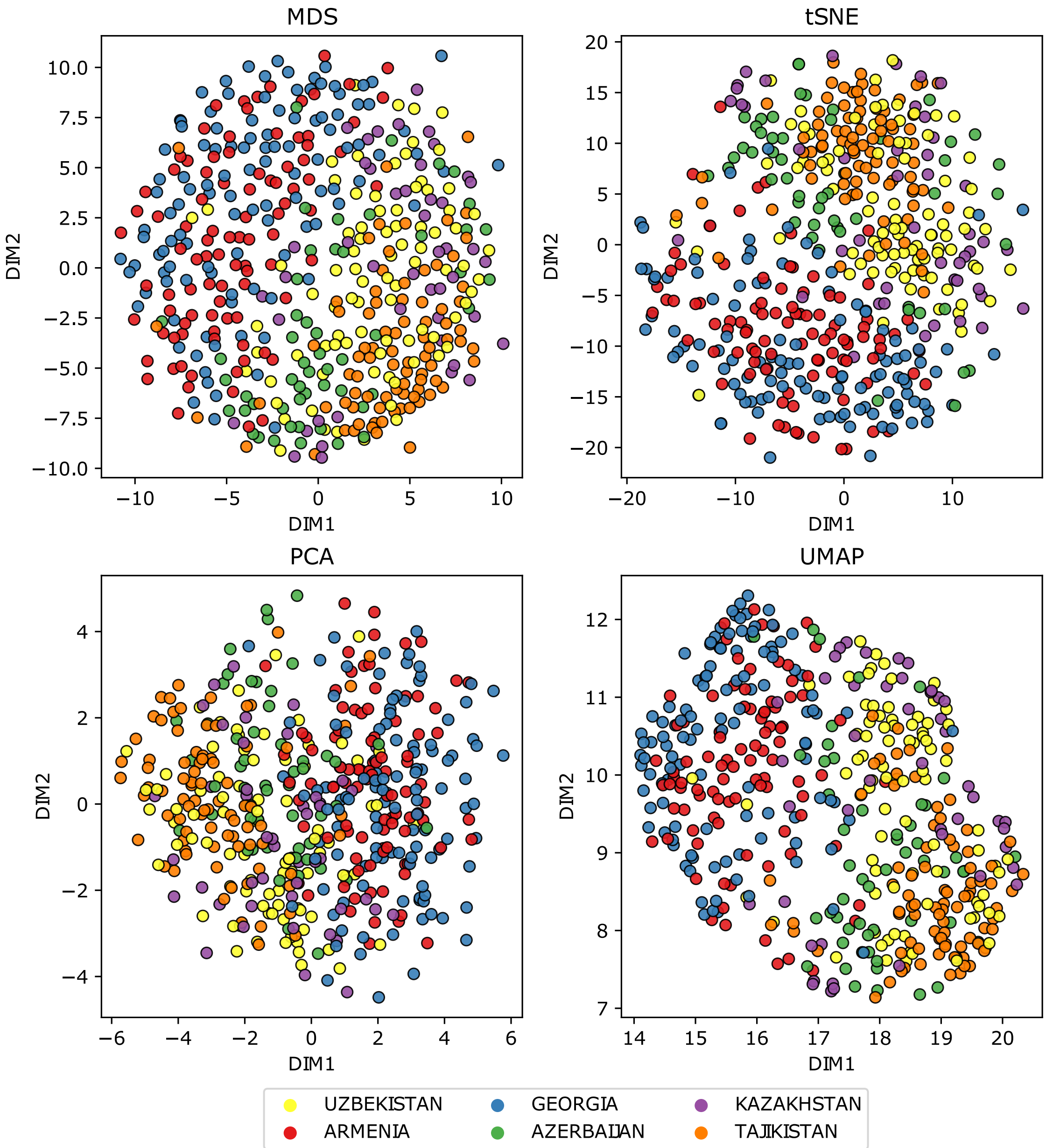
Supplementary Figure 2. Hierarchically clustered heatmap of the foods preferences using the UPGMA algorithm. Colored boxed at the edge of the heatmap refer to the inferred food groups (“alcoholic beverages” in purple, “pork” in orange, “sweets” in pink, “vegetables” in green, “animal proteins” in blue and “fruits” in red), while the colorbar indicates the correlation values.



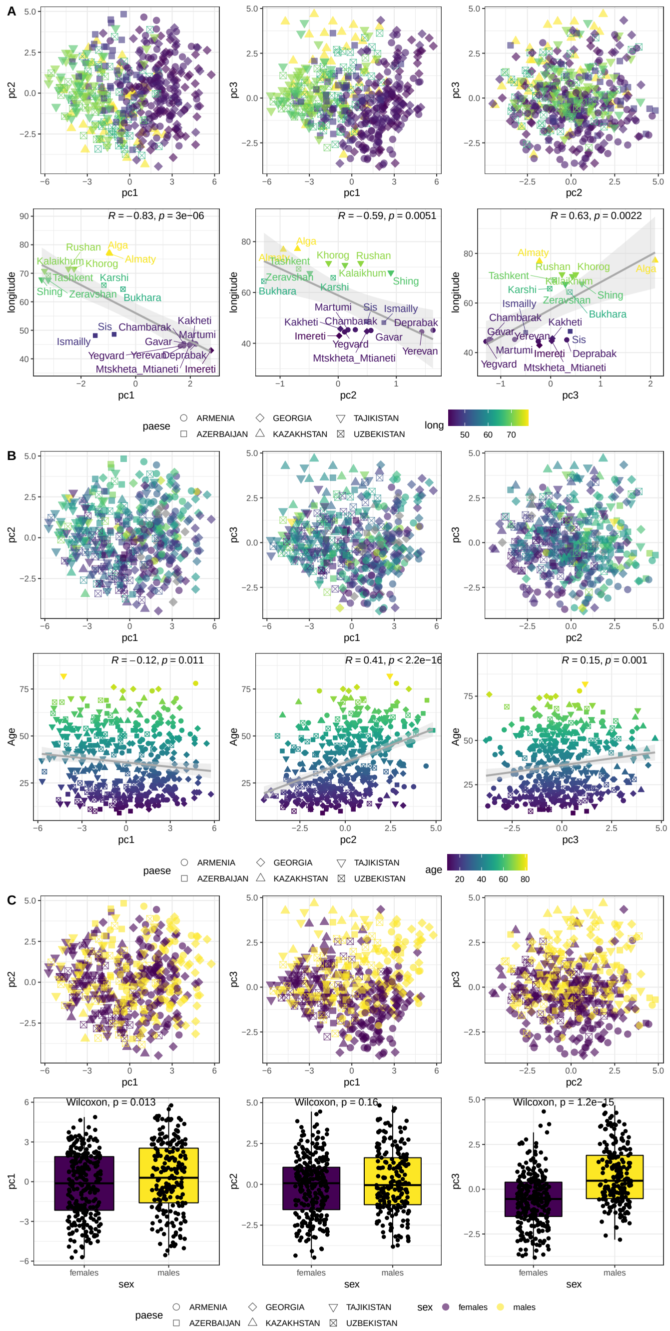
Supplementary Figure 3. Hierarchically clustered heatmap of the correlation matrix computed on the "never tasted"/missing answers to the questionnaire. Only the foods whose missing counts were significantly different among countries ("country-associated missing foods") have been considered in A, while B reports all foods.



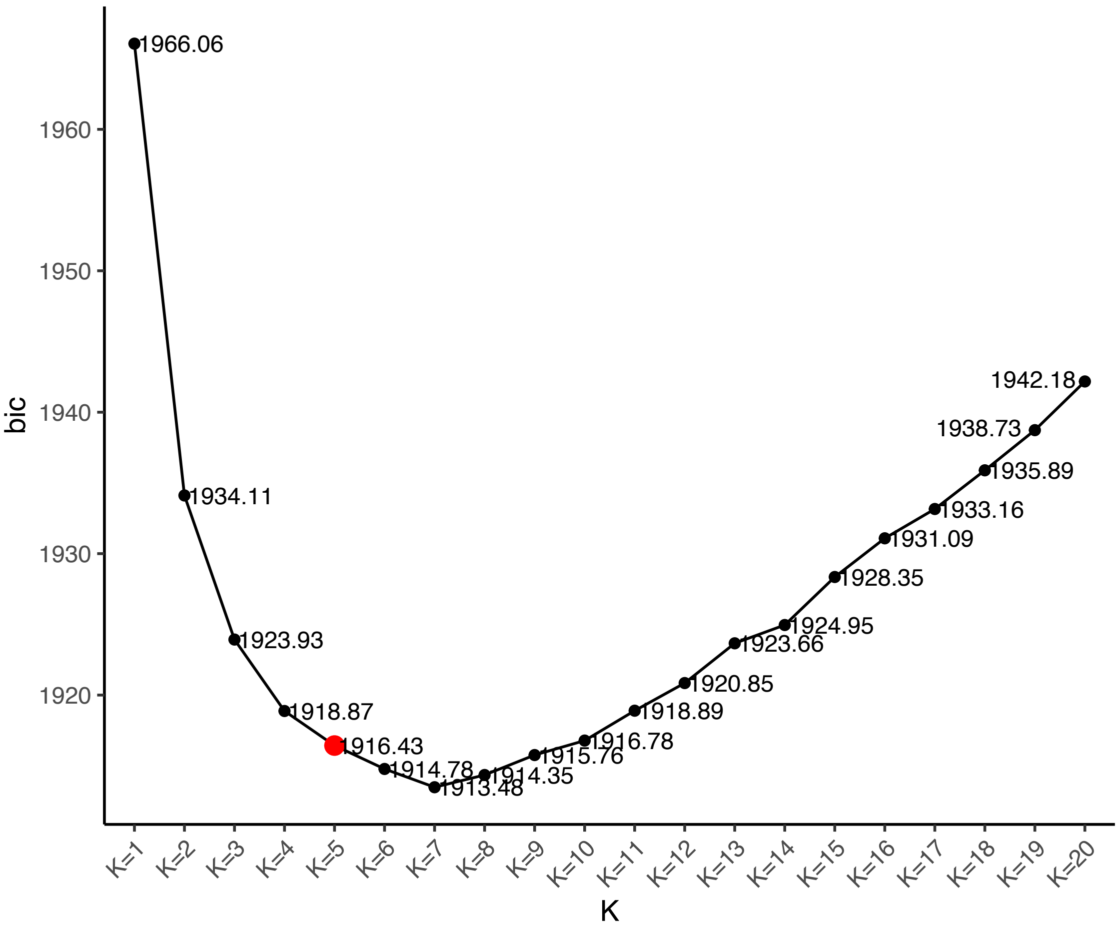
Supplementary Figure 4. Detailed representation of Figure 1. Map showing the geographical location of the cities (dots) from the six Silk Road countries sampled in this study (dots are coloured according to the country); pseudo-coordinates of each city from genetic (diamonds) and food preferences distances (triangles) are also shown.



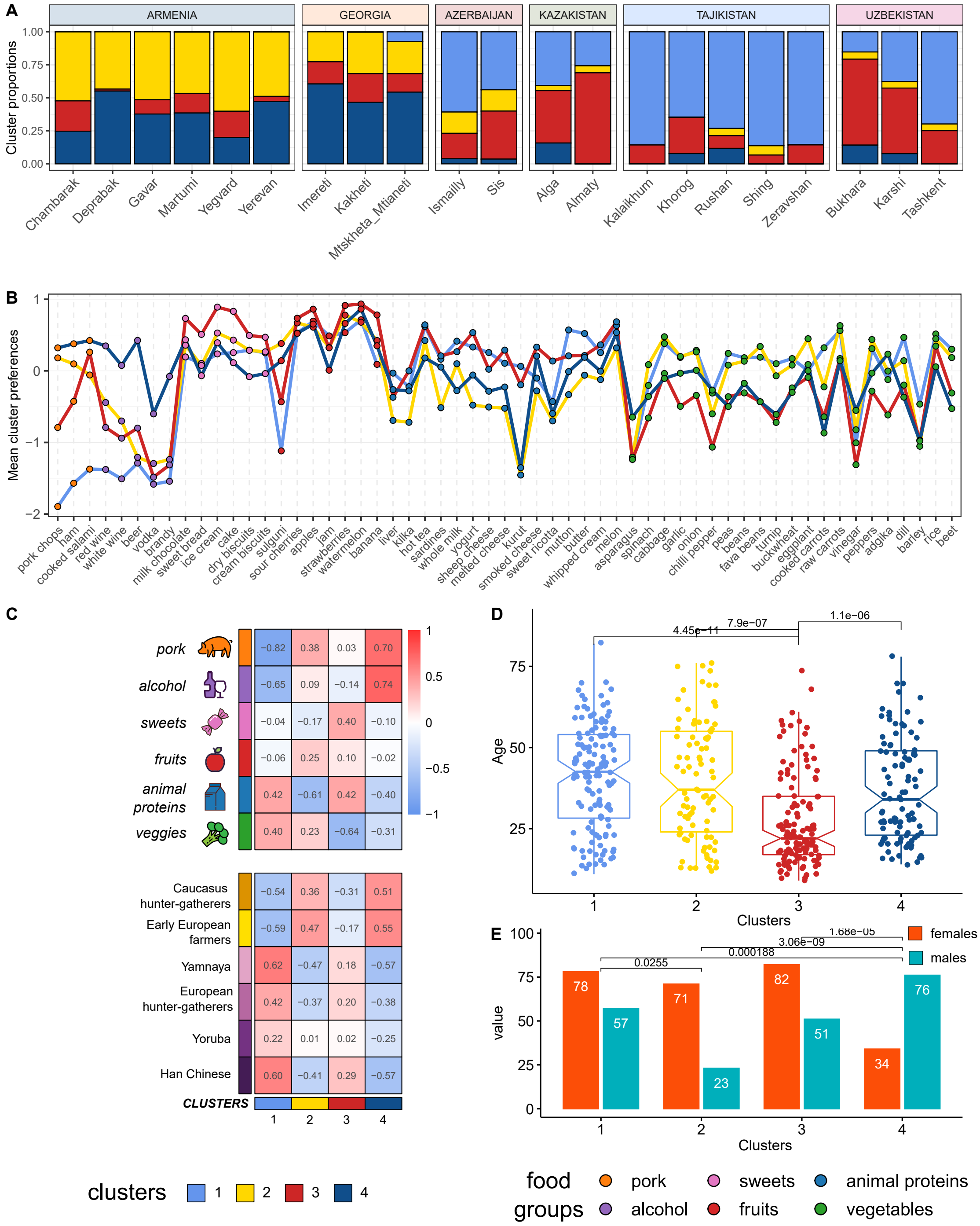
Supplementary Figure 5. Dimensionality reduction analyses on the "processed" dataset using four different techniques: multidimensional scaling (MDS), principal component analysis (PCA, first and second principal components explained 10.59% and 5.83% of the total variance), t-distributed stochastic neighbour embedding (tSNE) and Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP). Individuals are colored according to their country.



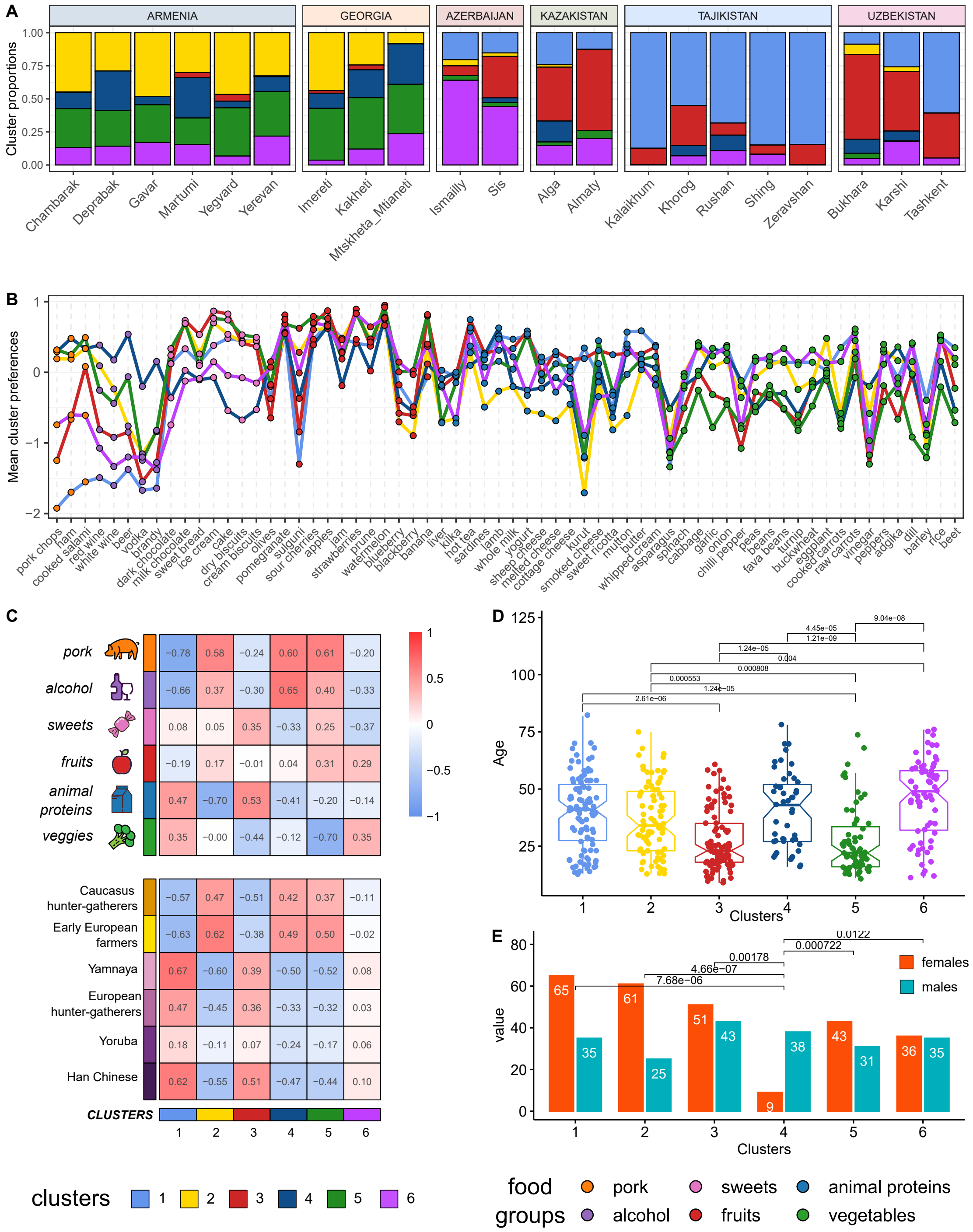
Supplementary Figure 6. Principal Components correlations with longitude, age and sex. A) First three principal components plot (top row) and Pearson correlation between component values averaged at the city levels and city longitudes (bottom row). Points colors and shapes refer to longitude and country, respectively. B) First three principal components plot (top row) and Pearson correlation between component values and age (bottom row). Points colors and shapes refer to age and country, respectively. C) First three principal components plot (top row) and boxplots representing the distribution of males and females along each principal component (bottom row). Points colors and shapes refer to sex and country, respectively. Statistical tests results have been added: Person correlations for panels A and B and Wilcoxon Rank Sum test for panel C.



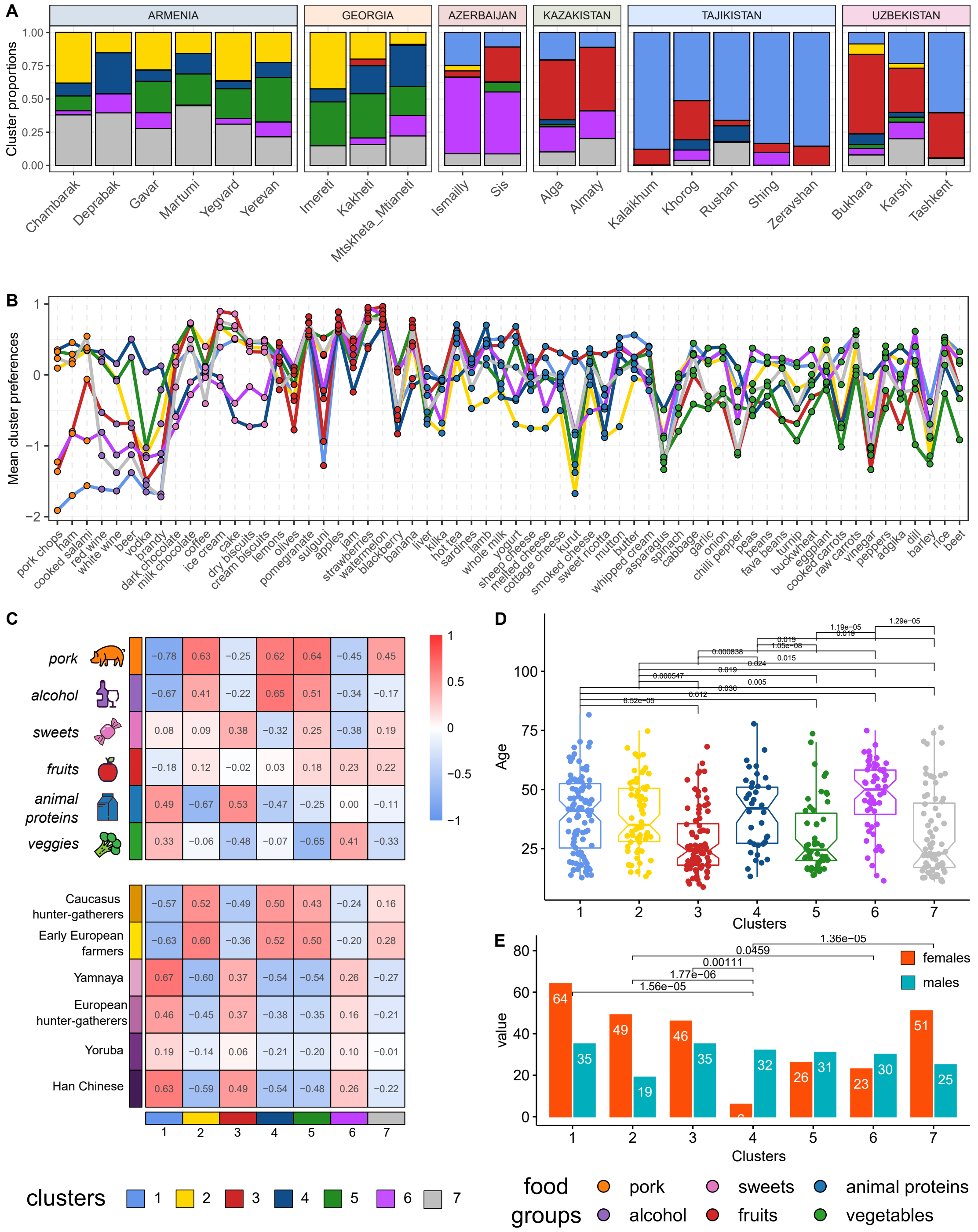
Supplementary Figure 7. Bayesian Information Criterion (BIC) value for each number of clusters as estimated by DAPC. The chosen number of clusters is marked in red.



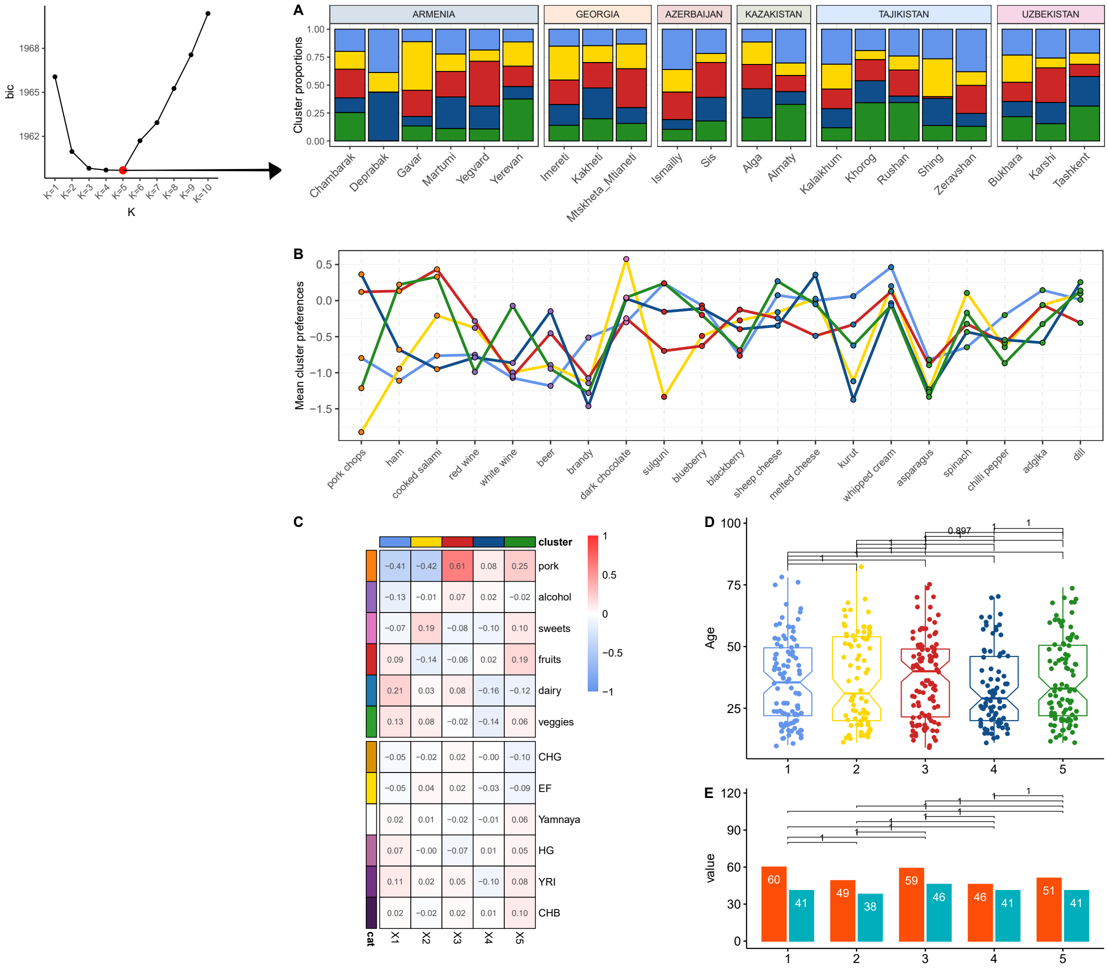
Supplementary Figure 8. DAPC clustering analyses with 4 clusters. A) DAPC clustering analyses with K=4 reporting the inferred clusters' probabilities of assignment averaged on the individual city of origin. B) Mean food preference values of individuals attributed to each cluster (each colour refers to a different cluster). Only the foods whose mean preferences were significantly different among the clusters are shown. C, D, E) Statistical analyses exploring the putative associations of the inferred clustering patterns with known individual information: food groups and genetic ancestral components (C), age (D) and sex (E). Bonferroni adjusted p-values are shown in panels D and E.



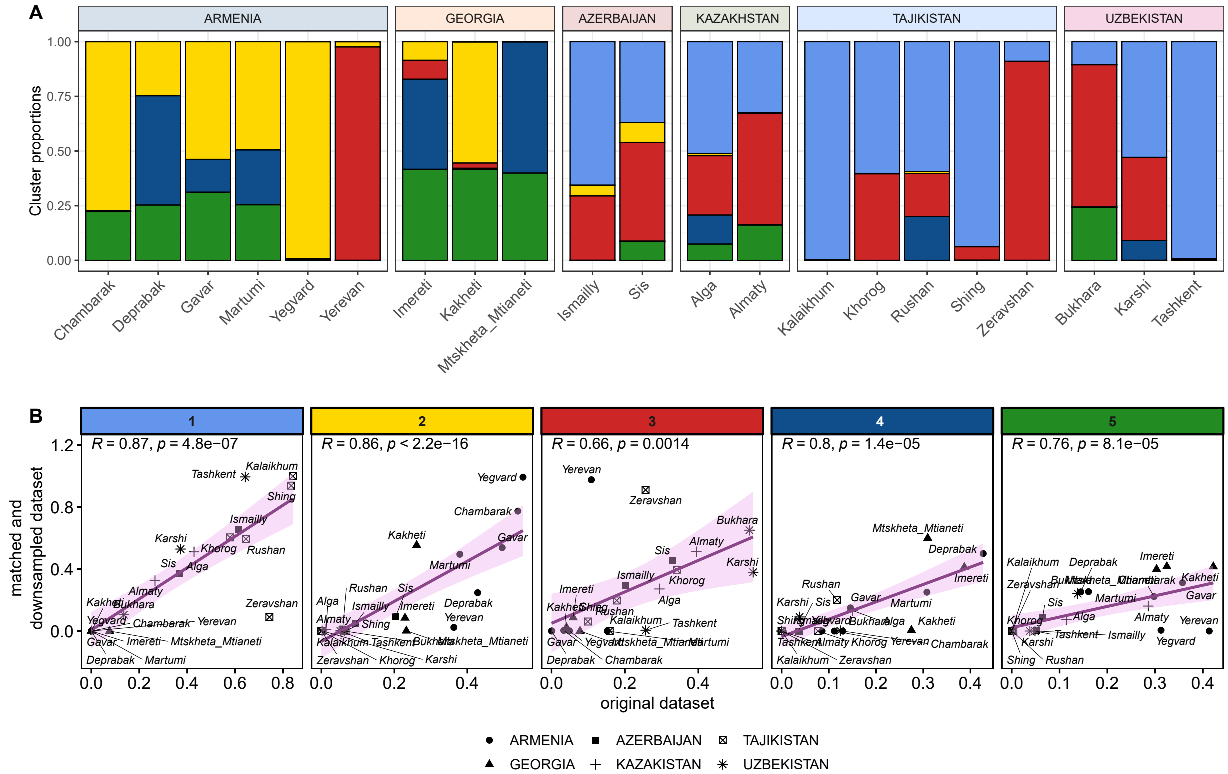
Supplementary Figure 9. DAPC clustering analyses with 6 clusters. A) DAPC clustering analyses with K=6 reporting the inferred clusters' probabilities of assignment averaged on the individual city of origin. B) Mean food preference values of individuals attributed to each cluster (each colour refers to a different cluster). Only the foods whose mean food preference were significantly different among the clusters are shown. C, D, E) Statistical analyses exploring the putative associations of the inferred clustering patterns with known individual information: food groups and genetic ancestral components (C), age (D) and sex (E). Bonferroni adjusted p-values are shown in panels D and E.



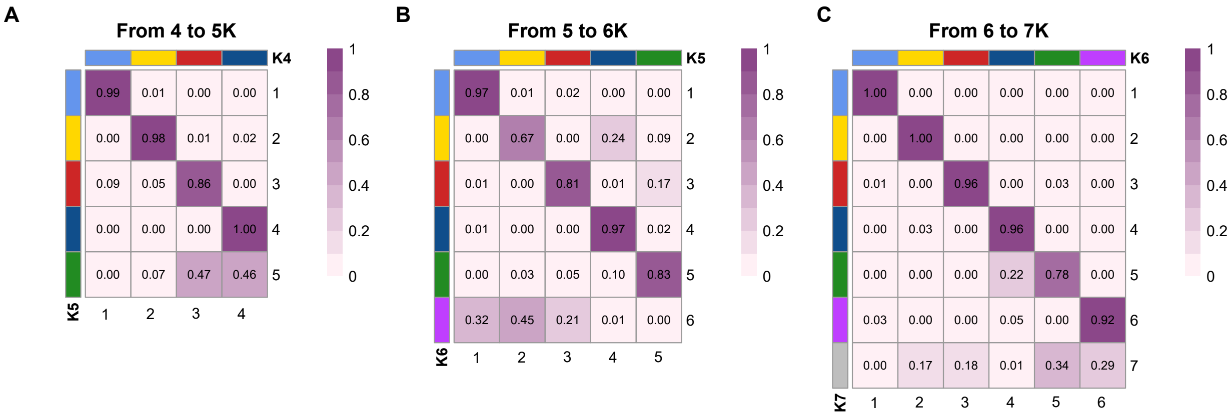
Supplementary Figure 10. DAPC clustering analyses with 7 clusters. A) DAPC clustering analyses with K=7 reporting the inferred clusters' probabilities of assignment averaged on the individual city of origin. B) Mean food preference values of individuals attributed to each cluster (each colour refers to a different cluster). Only the foods whose mean food preference were significantly different among the clusters are shown. C, D, E) Statistical analyses exploring the putative associations of the inferred clustering patterns with known individual information: food groups and genetic ancestral components (C), age (D) and sex (E). Bonferroni adjusted p-values are shown in panels D and E.



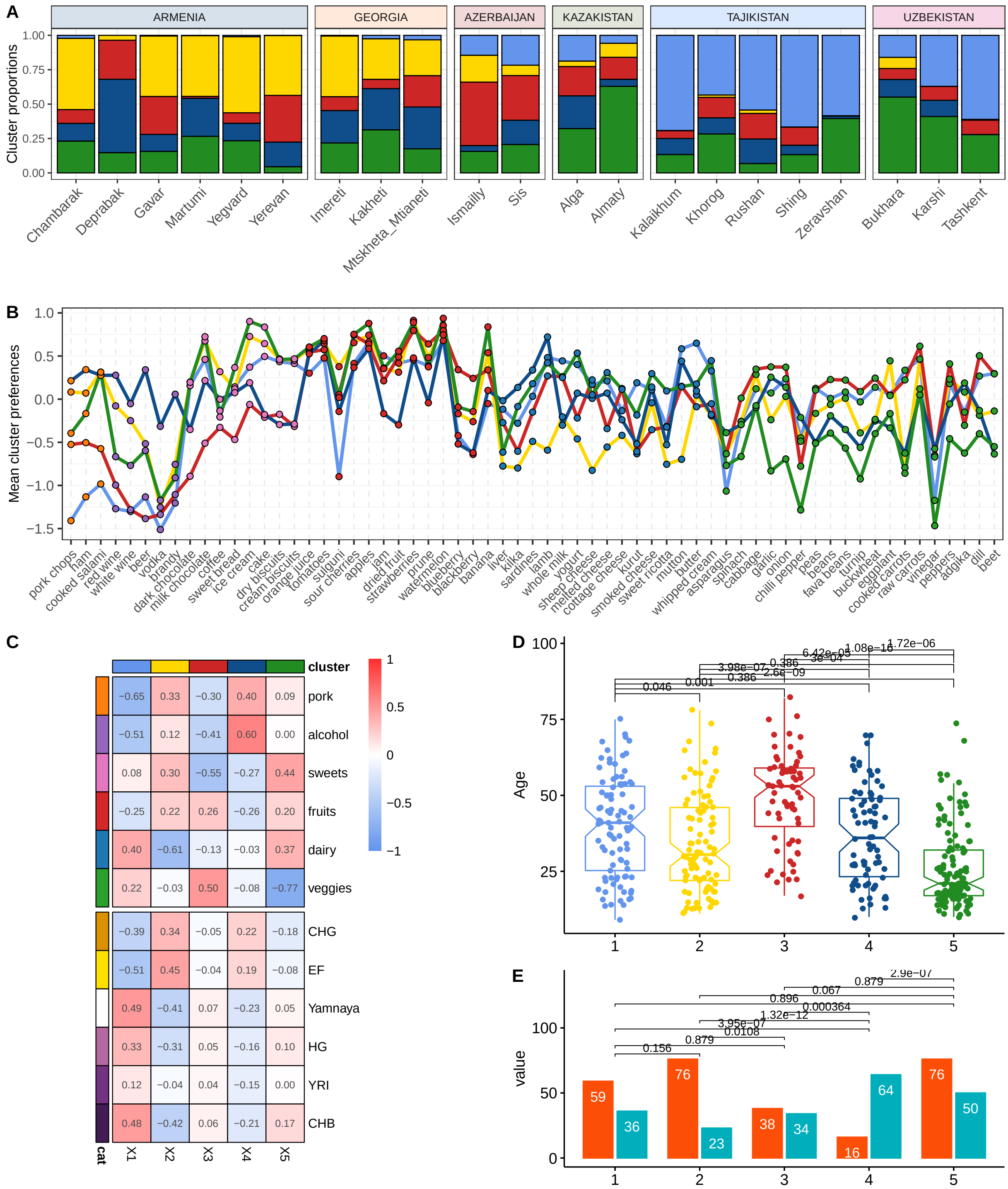
Supplementary Figure 11. DAPC clustering analyses with 5 clusters (best number of clusters according to BIC values) on the permutated dataset. A) DAPC clustering analyses with K=5 reporting the inferred clusters' probabilities of assignment averaged on the individual city of origin. B) Mean food preference values of individuals attributed to each cluster (each colour refers to a different cluster). Only the foods whose mean food preference were significantly different among the clusters are shown. C, D, E) Statistical analyses exploring the putative associations of the inferred clustering patterns with known individual information: food groups and genetic ancestral components (C), age (D) and sex (E). Bonferroni adjusted p-values are shown in panels D and E.



Supplementary Figure 12. A) DAPC results plotted for the downsampled and sex/age matched dataset. B) Correlation of DAPC results averaged by city between the original (Fig. 3A) and the downsampled and sex/age matched dataset.



Supplementary Figure 13. Non-negative least square function (NNLS) applied on the individual clustering assignments in order to decompose each cluster assignment vector of the run with $k + 1$ clusters as a combination of clustering results obtained with k clusters.



Supplementary Figure 14. DAPC clustering analyses with 5 clusters on the dataset where we replaced the "never tasted" answers with the value 3. A) DAPC clustering analyses with K=5 reporting the inferred clusters' probabilities of assignment averaged on the individual city of origin. B) Mean food preference values of individuals attributed to each cluster (each colour refers to a different cluster). Only the foods whose mean food preference were significantly different among the clusters are shown. C, D, E) Statistical analyses exploring the putative associations of the inferred clustering patterns with known individual information: food groups and genetic ancestral components (C), age (D) and sex (E). Bonferroni adjusted p-values are shown in panels D and E.

Descriptives

	country	N	Mean	95% Confidence Interval	
				Lower	Upper
cluster 1	UZBEKISTAN	85	0.33271	0.23501	0.43041
	ARMENIA	95	7.25e-4	4.68e-5	0.00140
	GEORGIA	114	0.00920	-0.00793	0.02633
	AZERBAIJAN	53	0.47946	0.35116	0.60776
	KAZAKHSTAN	46	0.36151	0.23094	0.49209
	TAJIKISTAN	84	0.70967	0.61875	0.80059
cluster 2	UZBEKISTAN	85	0.03958	2.09e-4	0.07895
	ARMENIA	95	0.48311	0.38578	0.58044
	GEORGIA	114	0.24087	0.16516	0.31658
	AZERBAIJAN	53	0.14809	0.05890	0.23727
	KAZAKHSTAN	46	0.02963	-0.01351	0.07277
	TAJIKISTAN	84	0.02371	-0.00800	0.05543
cluster 3	UZBEKISTAN	85	0.47855	0.37760	0.57949
	ARMENIA	95	0.05243	0.00900	0.09586
	GEORGIA	114	0.05479	0.01708	0.09249
	AZERBAIJAN	53	0.28014	0.16488	0.39540
	KAZAKHSTAN	46	0.33559	0.21264	0.45854
	TAJIKISTAN	84	0.21770	0.13654	0.29886
cluster 4	UZBEKISTAN	85	0.06612	0.01467	0.11757
	ARMENIA	95	0.17515	0.10140	0.24890
	GEORGIA	114	0.33886	0.25497	0.42274
	AZERBAIJAN	53	0.03552	-0.01309	0.08414
	KAZAKHSTAN	46	0.08815	0.01058	0.16572
	TAJIKISTAN	84	0.04769	0.00189	0.09350
cluster 5	UZBEKISTAN	85	0.08304	0.03003	0.13605
	ARMENIA	95	0.28859	0.20145	0.37572
	GEORGIA	114	0.35629	0.27280	0.43978
	AZERBAIJAN	53	0.05679	2.14e-4	0.11336
	KAZAKHSTAN	46	0.18512	0.07581	0.29442
	TAJIKISTAN	84	0.00123	-7.96e−4	0.00325

Supplementary Table 1. Descriptive table of cluster assignments per country.

food cluster	Cronbach's alpha
pork	0.6870
alcohol	0.7968
vegetables	0.8876
sweets	0.7729
animal proteins	0.8613
fruits	0.8752

Supplementary Table 2. Cronbach's alpha values for the food clusters.

Food	adj-pvalue
PORK_CHOPS	2.79e-43
SULGUNI	3.36e-40
KURUT	1.07e-35
HAM	2.97e-35
BRANDY	6.50e-32
RED_WINE	1.70e-27
COOKED_SALAMI	2.09e-27
WHITE_WINE	4.39e-27
VODKA	8.67e-26
BEER	2.20e-23
ASPARAGUS	9.22e-15
VINEGAR	9.42e-11
OLIVES	9.58e-11
BLACKBERRY	2.24e-10
BANANA	3.75e-09
ADGIKA	4.31e-08
JAM	4.12e-07
SALTY_RICOTTA	5.66e-07
SWEET_RICOTTA	3.00e-06
MULBERRY	5.30e-06
SPINACH	1.15e-05
BARLEY	1.29e-05
COTTAGE_CHEESE	1.77e-04
STRAWBERRIES	3.43e-04
CHILLI_PEPPER	1.71e-03
LIVER	3.29e-03
WILD_MUSHROOMS	4.33e-03
EGGPLANT	7.06e-03
SMOKED_CHEESE	1.25e-02
KILKA	2.00e-02
COFFEE	3.88e-02

Supplementary Table 3. Bonferroni adjusted p-value of chi-square test on missing counts among countries for each food. Just the significantly different foods are reported.