# Subtle cultural boundaries reinforce genetic structure in England

Yakov Pichkar[1,2] and Nicole Creanza[1,2]

[1] Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA

[2] Evolution Studies Initiative, Vanderbilt University, Nashville, TN, USA

Correspondence: nicole.creanza@vanderbilt.edu

## Abstract

Genes and languages both maintain signatures of human history. The evolution of genetics and of culture both have features that can track population movements and demographic history. Further, cultural traits may themselves impact these movements and demography. In particular, while speaking a different language appears to act as a barrier to gene flow, it is not clear whether more subtle dialect-level linguistic differences within a language can influence mating preferences and thus affect genetic population structure. We examine the strength of cultural barriers and of association within England using the spatial similarities between rates of linguistic and genetic change. We find that genes and dialect markers have similar spatial distributions at all geographic scales, though these similarities are more pronounced at larger scales. This covariation, in the absence of geographic barriers to coordinate linguistic and genetic differentiation, suggests that some cultural boundaries have maintained genetic population structure in England.

## Introduction

For as long as evolution has been studied, people have worked to understand the relationship between biological traits and socially learned behaviors (1–4). There are striking similarities between genetics and culture, both of which spread and change as groups of people move across the world. When people move into a region, their genes and culture can begin diverging from their previous population and mixing with those of nearby populations (5). The spatial patterns that form can mirror these demographic histories (6). Unlike vertically transmitted genetic traits, cultural traits are often inherited from people other than parents, so the transmission of cultural and biological traits can deviate at the scale of individuals (7). However, if similar individuals preferentially interact with one another – a tendency termed homophily – the population becomes structured and this structure limits the transmission of traits (8–10). These assortative interactions may be caused by genetics, as among many animals to limit interspecific mating, but humans in particular can learn to bias their interactions in complex and subtle ways, making homophily a partly learned behavior (11–13). When culture creates barriers to contact and interaction, it potentially drives the maintenance or the creation of covariation between genes and culture at relatively small scales. In this way, humans learn behaviors and preferences that act as cultural barriers to gene flow, altering the course of evolution.

Previous studies have investigated the distribution and the causes of this gene-culture covariation. In some cases, the evolution of cultural traits provides the selective pressures that drive genetic variation. For example, the consumption of unfermented dairy products (14) and of high-starch diets (15) altered the selective pressures on genes related to metabolism. Other traits do not affect fitness directly, but can be used as markers that drive cultural homophily. One such trait is the specific languages or dialects a person speaks, since languages contain clues about the history of populations (16–20) and have been found to co-vary with genetic variation on a worldwide scale (16, 21). Finer-scale analyses have found this linguistic-genetic covariation in smaller regions as well, including the Caucasus (22, 23), the Levant (24), the Amazon (25), and Africa (26, 27). Other studies have directly compared the rates of linguistic and genetic change, as these are expected to be highest near geographic or behavioral boundaries to human interaction (28). These borders between languages and regions with lower rates of gene flow have been found to co-localize throughout Europe (28), and even within countries (29).

However, these studies that observed evidence of gene-language covariation have not been able to analyze its causes. They tended to compare people speaking different languages, and few used quantitative measures of language variation (22, 28, 29). These limits were due, in part, to the limited availability of quantitative data describing cultural variation and the low

spatial resolution of available genetic data. These limits only allow the comparison of genes and language at the scale of ethnic groups and large regions, where the effects of population movement and cultural homophily cannot be disentangled. Therefore, the impact and spatial extent of these forces has not been thoroughly studied at smaller scales, such as within a geographic region where individuals generally speak the same language.

To clarify these causes of gene-language covariation, we present a joint analysis of genes and language in England. We use fine-scale linguistic and genetic data to measure the intensity, scale, and location of covariation. To compare genetics and language, we made use of two separate sets of data: the Survey of English Dialects (SED) (30) and the People of the British Isles (PoBI) project (31). Each dataset has a high geographic density, has been used previously to study the history of the region, and independently captures spatially meaningful linguistic and genetic variation (31–33). Leslie et al. (31) used the PoBI dataset to generate genetic similarities and clustered individuals into groups based on these similarities; these clusters retain evidence of population movements within the British Isles and from different parts of mainland Europe.

By using these genetic and language data in concert, we measure the relationship between genes and culture that formed over thousands of years in Great Britain. Comparing the rates of linguistic and genetic change throughout England, we find similarities between their spatial distribution at all scales of analysis, suggesting that subtle cultural boundaries could reinforce genetic population structure.

## Results

We quantified linguistic variation by measuring the rates of change of all phonetic features in a region (Fig. 1). Regions in which many features are changing suggest a border between dialects and potentially a difference in the underlying cultural identities associated with these dialects. We observed patterns of linguistic variation similar to those found in other analyses of the SED (30, 34). We found the high rates of language change near the boundaries of several dialect areas, namely in Northern England near the Scottish Border, the southern Welsh Marches, and between the West Country and the rest of South England.

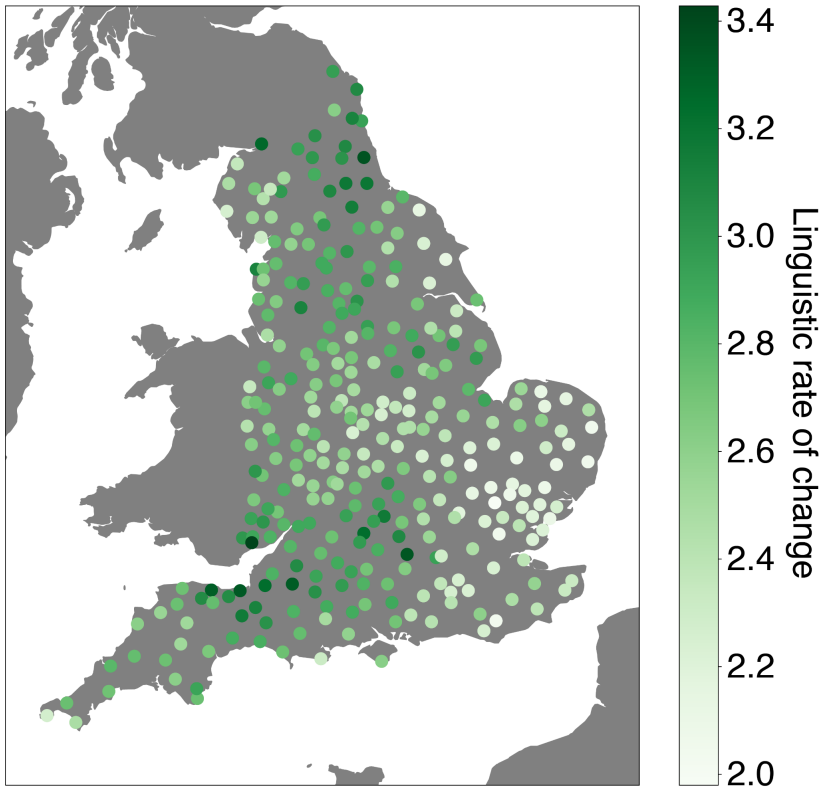


Fig. 1 **Linguistic rates of change in England.** For each location with linguistic data, estimated the linguistic rates of change by using the mean linguistic distance between that location and those within 100 km of it, adjusted for the distances to those locations.

We used fineSTRUCTURE to group 1,667 individuals from England into six geographically distinct hierarchical clusters. These were very similar to those clusters found in England by Leslie et al. (31). However, which individuals are assigned to each cluster vary between these two studies because of the stochastic nature of fineSTRUCTURE. Of the six clusters we found, one was dominant in South England and the Midlands, containing nearly half of all sampled individuals. Other clusters were found near the Welsh Marches, Scottish Marches, Cornwall, Northern England (centered on Yorkshire), and Somerset.

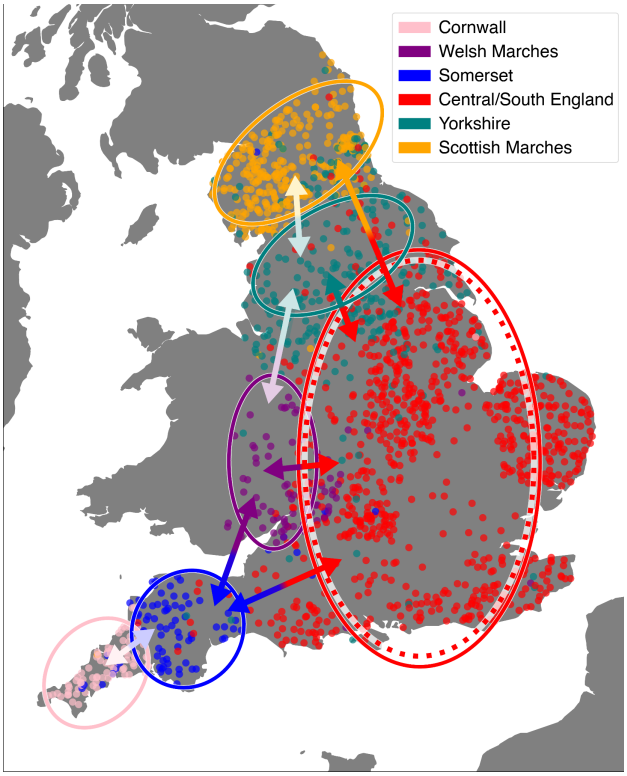


Fig. 2 **Relationship between rates of genetic and linguistic rates of change.** We measured whether regions with high rates of genetic change, particularly cluster boundaries, had higher rates of linguistic change, representing dialectal and other cultural boundaries. For between-cluster boundaries (double-sided arrows), a significant gene-language relationship is shown by a dark arrow and a non-significant relationship is shown by a light arrow. We draw a dotted ellipse in the cluster that represents much of the Midlands and South England, where there is a significant correlation between the spatial distributions of genes (of individuals in that cluster) and language (in the region where the cluster is found). No other clusters have this relationship (see Table 1). In this figure, the hierarchical clustering was set to K=6.

Table 1 **Genetic-linguistic covariation within clusters.** For each cluster identified at K=6 hierarchical clusters, we compared genetic and linguistic rates of change using linear regression. The only cluster in which genetic and linguistic rates of change were significantly correlated encompassed most of South and Central England (represented in red in Fig. 2). All of the individuals in the Cornwall genetic cluster were in a single locale, so we could not compare their rates of genetic change. Bold values are those significant after Bonferroni correction. We include the number of individuals in each cluster (*N*).

| Cluster (*N*) | P-value | Cluster (*N*) | P-value |
|---|---|---|---|
| Scottish Marches (295) | 1 | Somerset (82) | 0.977 |
| Cornwall (82) | NA | South and Central England (836) | **$3.14\times10^{-11}$** |
| Welsh Marches (86) | $3.66\times10^{-2}$ | Yorkshire and NW England (286) | $5.91\times10^{-2}$ |

Table 2 **Genetic-linguistic covariation between clusters.** For each set of adjacent clusters at K=6 hierarchical clusters, we measured whether their cluster boundary was found in regions with high linguistic rates of change using Mann-Whitney U tests. We also include the number of locations used for the tests. Bold values indicate those significant after Bonferroni correction.

| Cluster 1 | Cluster 2 | $N_1$, $N_2$ | $U$-value | $P$-value |
|---|---|---|---|---|
| Somerset | Cornwall | 24, 7 | 131.0 | 0.99 |
| Somerset | Welsh Marches | 62, 15 | 185.0 | $\mathbf{1.6\times10^{-4}}$ |
| Yorkshire | Scottish Marches | 82, 48 | 1531.0 | $1.76\times10^{-2}$ |
| South/Central England | Scottish Marches | 213, 66 | 4724.0 | $\mathbf{3\times10^{-5}}$ |
| South/Central England | Yorkshire | 208, 64 | 4581.0 | $\mathbf{8\times10^{-5}}$ |
| Yorkshire | Welsh Marches | 126, 29 | 2471.0 | 0.999 |
| South/Central England | Welsh Marches | 221, 62 | 3356.0 | **0.00** |
| South/Central England | Somerset | 240, 25 | 1368.0 | **0.00** |

We find that for most pairs of adjacent or overlapping clusters, their boundary lies in a region with higher linguistic rates of change than the core of these clusters (Fig. 2 and Table 2). Notably, regions with long-lasting cultural differences – Welsh and Scottish borders – both have co-occurring genetic and linguistic boundaries. Between the clusters in these regions and those in the interior of England, the colocalization of genetic and cultural barriers is also consistently strong (Fig. 2).

We find some covariation of language and genetics between clusters within the interior of England, farther from the borders and in regions that have been historically English, as opposed to Scottish or Welsh. Near both Yorkshire and the West Country, we find an association between cluster boundaries and linguistic rates of change. We do not find such an association in Cornwall, which is unexpected given the region's ethnic, dialectical, and genetic distinctiveness. However, the region's unique dialect patterns are most prominent in the western-most end of the county, whereas our genetic data has at most a county-scale resolution.

In addition to the association between language and genetic cluster boundaries, we measured whether genetics within relatively well-mixed clusters co-varies with language. For each level of hierarchical clustering, we compared the rates of linguistic and genetic change within the boundaries of each cluster (Fig. 2, Table 1). We find that for a single cluster, there is a spatial relationship between genetic and linguistic variation that was not captured through our

hierarchical clustering. For this cluster, made up of most individuals in the Midlands to East England, the spatial distribution of genetic rates of change is correlated with those of linguistics (Fig. 2). In this region, a similar pattern is found at all scales of the hierarchical clustering (Fig. S1-S5), pointing to spatially heterogeneous genetic variation that is not captured by the clustering.

## Discussion

Independently of one another, our analyses of genetic and linguistic data identify geographic patterns also found in previous studies. We measured the spatial distribution for rates of language change – the rates being greater at boundaries between dialects – throughout England, including near various cultural boundaries (30, 34). We found high rates of language change at the borders of England and between previously identified dialect clusters, suggesting that our measure is a good proxy for cultural variation. Our genetic analysis also identifies spatial patterns similar to those found previously. These patterns of genetic clusters reflect waves of migration to Great Britain and of later movements of people (31). This includes up to eight clusters ranging in size from 82 to 836 individuals; these clusters spanned regions as small as a single county (Cornwall) to large enough to encompass most of the Midlands and South England. Attempting to cluster this genetic data into more than six clusters with fineSTRUCTURE did not reveal more geographic population structure (Fig. S5).

We find that linguistic and genetic variation parallel one another throughout much of England. The boundaries of genetic clusters almost always appear in regions with high rates of cultural change. This correspondence between genetic boundaries – which are associated with both historical migrations and distributions of ethnicities – and linguistic boundaries – associated with dialect and other cultural differences – demonstrate that language and genetics have similar geographic patterns in England. These gene-language patterns at the the English-Welsh and the English-Scottish borders support previous studies that identify the covariation of culture and genetics at ethnic boundaries (22, 24, 26–29). We find that, in addition to more prominent cultural borders, this covariation extends to less dramatic distinctions, including those between South East and South West England and between North and South England. These clusters are not only geographically significant, but also maintain variation associated with various admixture events from mainland Europe, including the Anglo-Saxons, Danes, and Normans (31). These results suggest that cultural distinctions have influenced how people move and behave by maintaining genetic boundaries; however, these results cannot fully distinguish between the influence of cultural homophily and that of other demographic events, both of which may have contributed to the existing patterns of genes and language.

To find evidence of cultural barriers to gene flow, we also analyzed the genetic variation existing within genetic clusters, rather than considering only their boundaries. We tested whether regions with higher rates of language change within the bounds of a single cluster corresponded to higher rates of genetic change within that cluster. We find that genetic differences do not appear to covary with rates of language change within any of the smaller clusters, but do so in the largest cluster, which encompasses much of the Midlands and South England. Despite this cluster appearing relatively well-mixed (it cannot be divided into more geographically meaningful clusters) higher rates of genetic change coincide with dialect boundaries. These data suggest that dialect boundaries – or related cultural differences – contribute to genetic differentiation in that region. The lack of major geographic boundaries in Central and South England imply that the population should be relatively well-mixed; however, the varied history of the Midlands and South England (31) and existing covariation between genes and language within this region strongly suggest that cultural boundaries have limited gene flow here.

These results demonstrate that genes and culture in England have close relationships at two scales: at long-standing cultural borders and in a large region with few geographic barriers.. The relationship between clusters may be the product of large movements in the past. The longevity of these geographically significant distinctions, however, suggests that culture had a role in maintaining this variation. In addition to common genetic and linguistic boundaries between clusters, their colocalization within them is evidence of culture influencing population structure. For genetic variation within clusters to covary with culture, either a third force (such as geography) limited their diffusion, or cultural differences contributed to the formation of groups. Given the many roads and navigable rivers in the Midlands and South England (35), the covariation of genes and culture in the region favors cultural homophily as the primary barrier to gene flow.

Our analysis of the interplay between genetics and culture has a greater spatial resolution and more quantitative data than those of others (23, 27–29), but we are still limited by the resolution of our data and our use of separate sets of data. Our genetic data was restricted to spatial resolutions as fine as counties, which prevented the comparison of genetics to intra-county linguistic variation. Since we used separate genetic and linguistic datasets, we could only comment on processes that formed patterns over many generations, rather than a direct analysis of assortative mating based on language or cultural homophily, which could be possible with linguistic information from a set of genotyped individuals. However, previous work has identified widespread endogamous and geographically restricted marriages in England (36). Such homophily likely had profound effects on both the cultural and genetic landscapes (37–39). Given the movement and cultural change that has occurred since the

Industrial Revolution, these results are more informative about our human past than predictive of culture-based homophily today. Nevertheless, we suggest that cultural distinctions have shaped the population structure of individuals within England. Our analysis reveals a cultural process that has had consequences on both genetic and cultural diversity. Similar processes may have taken place around the world, molding families, ideas, and the history of peoples.

## Methods

*Processing linguistic data*

Linguistic data was collected by the Survey of English Dialects (SED) (30), which gathered word choice and pronunciation in England, the Isle of Man, and parts of Wales. The SED, conducted in the 1950's, prioritized people who were elderly, from rural areas, and who had agricultural backgrounds. This survey took place across 313 locales, of which we excluded two on the Isle of Man.

We selected and digitized 225 binary elements that describe whether a particular sound is present in each word, expanding our previous digitized SED data of 45 elements (32). For example, an element may be whether "hammer" is pronounced with or without an initial [h] sound.

We took these 225 binary elements and combined those that were relevant into 55 categories representing the frequencies of specific linguistic features, such as the presence of syllable-initial [h] (the mean occurrence of syllable-initial [h] in hammer, halter, harvest, etc.). If a linguistic feature contained multiple binary elements, each of the 311 localities was assigned a frequency of occurrence for that feature; if a linguistic feature was represented by only a single element, the data were binary for that category.

*Processing genetic data*

We used genetic data collected for the People of the British Isles (PoBI) project, which included the genotyping of people from the British Isles. These people were mostly from rural backgrounds and all had grandparents born within 80 km of one another; these grandparents had a mean birth year of 1885 (s.d.=18yrs.). The genetic data consists of over 500,000 single nucleotide polymorphisms (SNPs) for each of 2,039 people, after quality control. Each of these individuals was placed into one of 36 locations in Great Britain and Northern Ireland to protect their anonymity, and 29 of these are in England. Previous work by Leslie et al. found that these individuals could be placed into 17 geographically cohesive clusters based on their genetics.

We replicated the quality control, phasing, and clustering procedure from Leslie et al. Since our linguistic data was limited to England, we used a subset of 1,667 individuals by keeping those whose grandparents were born in England. This differs from the data used by Leslie et al., as

they included all 2,039 individuals with grandparents from England, Wales, Scotland, and Northern Ireland (31).

After these filters, we used fineSTRUCTURE to hierarchically cluster these individuals into geographically cohesive groups based on genetics. Beginning with two clusters, we increased the number of clusters until we found six such clusters; past that point, new clusters lost geographic cohesiveness. These six clusters correspond to the six English clusters found by Leslie et al. within England.

*Gene-language processing*

Since the genetic and linguistic datasets include information from different individuals, we conducted our analyses based on the geographic sampling locations of the individuals in each dataset. We assigned each of the 311 locales with linguistic data to one of the 29 county-sized regions with genetic data. Each locale of these 311 locations was treated as including individuals with genetic gata from that region.

We found the Euclidean distance between the linguistic features for all pairs of locations within 100 km of one another. We then found the average rate of linguistic change at that location by averaging the linguistic distances to that location, weighed by the inverse of spatial distances to that location.

To find genetic distances, we used the output produced by fineSTRUCTURE, termed a genetic copying matrix, which represents the total recombination map distance of haplotypes donated from one individual to another, according to fineSTRUCTURE's model (40). Distances between copying vectors have been used previously to find the genetic distances between groups of individuals (31). We used this method to calculate distances between each pair for individuals, which we used to find the mean distance between clusters and between regions with genetic data. We did this for every pair of adjacent genetic clusters from fineSTRUCTURE, and within every cluster found in more than one location. For every set of clusters being compared, for each pair of locales within 100 km of one another, we found the mean genetic distance between individuals from different clusters (or within the same cluster for within-cluster comparisons) and from the genetic regions associated with those locales. We found rates of genetic change at each location using genetic distances, identically to rates of linguistic change.

We repeated the steps to calculate genetic rates of change another 500 times, randomizing the order of the genetic relatedness matrix each time. We used these rates of change on permuted data to standardize the rates of genetic change to make them comparable and to decrease sampling biases.

To compare the genetic and linguistic rates of change, we used two methods. Within clusters, we measured whether a positive relationship existed via linear regression of the linguistic and genetic rates of change. This was done for each cluster with individuals spanning multiple sampling locations. To measure the overlap of genetic cluster boundaries and dialect boundaries, we used the Mann-Whitney U test. Specifically, we tested whether linguistic rates of change were greater at cluster boundaries than they were in other regions in which the clusters appeared.

## References


1. C. Darwin, *The Descent of man* (1871).

2. L. L. Cavalli-Sforza, M. W. Feldman, Cultural transmission and evolution: a quantitative approach. *Monogr. Popul. Biol.* **16**, 1–388 (1981).

3. C. J. Lumsden, E. O. Wilson, GENES, MIND, AND IDEOLOGY. *The Sciences* **21**, 6–8 (1981).

4. R. Boyd, P. J. Richerson, *Culture and the Evolutionary Process* (University of Chicago Press, 1985).

5. N. Creanza, M. W. Feldman, Worldwide genetic and cultural change in human evolution. *Curr. Opin. Genet. Dev.* **41**, 85–92 (2016).

6. M. J. Hamilton, B. Buchanan, Spatial gradients in Clovis-age radiocarbon dates across North America suggest rapid colonization from the north. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 15625–15630 (2007).

7. C. R. Guglielmino, C. Viganotti, B. Hewlett, L. L. Cavalli-Sforza, Cultural variation in Africa: role of mechanisms of transmission and adaptation. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7585–7589 (1995).

8. N. Creanza, M. W. Feldman, Complexity in models of cultural niche construction with selection and homophily. *Proc. Natl. Acad. Sci. U. S. A.* **111 Suppl 3**, 10830–10837 (2014).

9. E. Katsnelson, A. Lotem, M. W. Feldman, Assortative social learning and its implications for human (and animal?) societies. *Evolution* **68**, 1894–1906 (2014).

10. F. Fu, M. A. Nowak, N. A. Christakis, J. H. Fowler, The evolution of homophily. *Sci. Rep.* **2**, 845 (2012).

11. I. Eshel, L. L. Cavalli-Sforza, Assortment of encounters and evolution of cooperativeness. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 1331–1335 (1982).

12. A. Abdellaoui, *et al.*, Association between autozygosity and major depression: stratification due to religious assortment. *Behav. Genet.* **43**, 455–467 (2013).

13. A. Tenesa, K. Rawlik, P. Navarro, O. Canela-Xandri, Genetic determination of height-mediated mate choice. *Genome Biol.* **16** (2015).

14. A. Beja-Pereira, *et al.*, Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nat. Genet.* **35**, 311–313 (2003).

15. R. C. Iskow, O. Gokcumen, C. Lee, Exploring the role of copy number variants in human adaptation. *Trends Genet.* **28**, 245–257 (2012).

16. 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).

17. C. Barbieri, A. Butthof, K. Bostoen, B. Pakendorf, Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia. *Eur. J. Hum. Genet.* **21**, 430–436 (2013).

18. C. Barbieri, *et al.*, Between Andes and Amazon: the genetic profile of the Arawak-speaking Yanesha. *Am. J. Phys. Anthropol.* **155**, 600–609 (2014).

19. P. Verdu, E. M. Jewett, T. J. Pemberton, N. A. Rosenberg, M. Baptista, Parallel Trajectories of Genetic and Linguistic Admixture in a Genetically Admixed Creole Population. *Current Biology* **27**, 2529–2535.e3 (2017).

20. R. D. Gray, A. J. Drummond, S. J. Greenhill, Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* **323**, 479–483 (2009).

21. L. L. Cavalli-Sforza, E. Minch, J. L. Mountain, Coevolution of genes and languages revisited. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 5620–5624 (1992).

22. T. M. Karafet, *et al.*, Coevolution of genes and languages and high levels of population structure among the highland populations of Daghestan. *J. Hum. Genet.* **61**, 181–191 (2016).

23. O. Balanovsky, *et al.*, Parallel evolution of genes and languages in the Caucasus region. *Mol. Biol. Evol.* **28**, 2905–2920 (2011).

24. M. Haber, *et al.*, Genome-wide diversity in the levant reveals recent structuring by culture. *PLoS Genet.* **9**, e1003316 (2013).

25. T. Di Corcia, *et al.*, East of the Andes: The genetic profile of the Peruvian Amazon populations. *Am. J. Phys. Anthropol.* **163**, 328–338 (2017).

26. E. G. Atkinson, *et al.*, Genetic structure correlates with ethnolinguistic diversity in eastern and southern Africa. *Am. J. Hum. Genet.* **109**, 1667–1679 (2022).

27. D. Sengupta, *et al.*, Genetic substructure and complex demographic history of South African Bantu speakers. *Nat. Commun.* **12**, 2080 (2021).

28. G. Barbujani, R. R. Sokal, Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 1816–1819 (1990).

29. G. Barbujani, R. R. Sokal, Genetic population structure of Italy. II. Physical and cultural barriers to gene flow. *Am. J. Hum. Genet.* **48**, 398–411 (1991).

30. H. Orton, E. Dieth, *Survey of English Dialects* (1963).

31. S. Leslie, *et al.*, The fine-scale genetic structure of the British population. *Nature* **519**, 309–314 (2015).

32. A. C. Sherriah, H. Devonish, E. A. C. Thomas, N. Creanza, Using features of a Creole language to reconstruct population history and cultural evolution: tracing the English origins of Sranan. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **373** (2018).

33. C. Upton, J. D. A. Widdowson, *An Atlas of English Dialects: Region and Dialect* (Routledge, 2013).

34. W. Viereck, Dialectal speech areas in England: Orton's phonetic and grammatical evidence. *J. Eng. Linguist.* **19**, 240–257 (1986).

35. J. F. Edwards, B. P. Hindle, The transportation system of medieval England and Wales. *J. Hist. Geogr.* **17**, 123–134 (1991).

36. K. D. Snell, English rural societies and geographical marital endogamy, 1700-1837. *Econ. Hist. Rev.* **55**, 262–298 (2002).

37. B. W. Domingue, J. Fletcher, D. Conley, J. D. Boardman, Genetic and educational assortative mating among US adults. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7996–8000 (2014).

38. N. Creanza, O. Kolodny, M. W. Feldman, Cultural evolutionary theory: How culture evolves and why it matters. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 7782–7789 (2017).

39. V. Labeyrie, M. Thomas, Z. K. Muthamia, C. Leclerc, Seed exchange networks, ethnicity, and sorghum diversity. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 98–103 (2016).

40. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).