Journal of Economic Geography **14** (2014) pp. 179–198 doi:10.1093/jeg/lbt019
Advance Access Published on 9 August 2013

Genetic distance, transportation costs, and trade[1]

*Paola Giuliano**,†, Antonio Spilimbergo** and Giovanni Tonon****

*UCLA, NBER, CEPR, and IZA, Los Angeles, CA, USA
**IMF, CEPR, CREaM, and WDI, Washington, DC, USA
***Dana-Farber Cancer Institute, Harvard University, Cambridge, MA, USA
†Corresponding author: Paola Giuliano, UCLA Anderson School of Management, 110 Westwood Plaza, C517 Entrepreneurs Hall, Los Angeles, CA 90095. *email* <paola.giuliano@anderson.ucla.edu>

Abstract

Genetic distance, geographic proximity, and economic variables are strongly correlated. Disentangling the effects of these factors is crucial for interpreting these correlations. We show that geographic factors that shaped genetic patterns in the past are also relevant for current transportation costs and could explain the correlation between trading flows and genetic distance. After controlling for geography, the impact of genetic distance on trade disappears. We make our point by constructing a database on geographical barriers, by introducing a novel dataset on transportation costs, and by proposing a new classification of goods according to the ease with which they can be transported.

Keywords: Transportation costs, genetics, trade, cultural economics, geography
JEL classifications: Z10, F10
Date submitted: 27 September 2011 **Date accepted:** 31 May 2013

1. Introduction

Do genetic differences play a role in explaining social and economic outcomes? And, if so, why? Social scientists and economists are interested in the study of genetic differences between populations for several reasons. First, contemporary genetic differences reveal the ancient history of populations, which may have an impact on contemporary economic outcomes. Second, genetic distance can be used as a proxy of *vertical* transmission of cultural traits, and the study of genetic distance may be revealing for the study of cultural differences (Cavalli-Sforza et al., 1994; Stone and

1 The views expressed in this paper are those of the authors and do not necessarily represent those of the International Monetary Fund, its board of directors, or the Dana-Farber Cancer Institute. We thank the editor and two anonymous referees for comments that substantially improved the paper. We also thank Francesco Caselli, Luca Cavalli-Sforza, James Fearon, Oded Galor, Luigi Guiso, Peter Howitt, David Hummels, Simon Johnson, Ross Levine, Nuno Limao, Paolo Mauro, Enrico Spolaore, Arvind Subramanian, Antonio Terraciano, Romain Wacziarg, David Weil, Fabrizio Zilibotti, Luigi Zingales, and seminar participants at Brown University, the IMF, the University of California at Berkeley, the NBER workshop on political economy, the 'Economics of Diversity, Migration and Culture' workshop in Bologna, and the NBER Summer Institute on Macroeconomics and Income Distribution. José Romero and Andrea Passalacqua provided excellent research assistance. Paola Giuliano thanks the Russell Sage Foundation for its wonderful hospitality.

Lurquin, 2007). Third, genetic differences across populations may have a direct impact on social outcomes; for instance, there is increasing evidence showing that genetic characteristics are associated with the incidence of some medical conditions.[2]

In a seminal work, Cavalli-Sforza et al. (1994) introduced a measure of genetic distance to address the questions listed above. (For a technical description of this measure, see Section 2.1). After collecting an impressive database on genetic distances among various populations, Cavalli-Sforza et al. reconstructed the history of the spreading of the Neolithic revolution. In their dual-inheritance theory, Cavalli-Sforza and Feldman (1981) also claim that culture is influenced and constrained by genes. In particular, they show that linguistic trees representing the history of language separations and evolution are highly correlated with genetic differences so that the distant history of linguistic evolution can be studied looking at concomitant genetic differences across populations (Cavalli-Sforza and Feldman, 1981; Stone and Lurquin, 2007).

Not surprisingly, economists are starting to recognize the usefulness of the notion of genetic distance to explain economic phenomena. Spolaore and Wacziarg (2009) find that genetic distance has a statistically significant effect on income differences, even after controlling for geographical and cultural differences. In studying the relationship between bilateral trust and bilateral trade, Guiso et al. (2009) use genetic and somatic differences as an instrument for trust. Desmet et al. (2011) show that genetic distance is relevant in explaining the stability of political coalitions in Europe.

The main challenge in the interpretation of the correlation between genetic distance and economic outcomes is to disentangle the relative importance of vertically transmitted characteristics from the importance of geographical barriers. Various scholars have documented a striking association between genetic distance and geographical barriers such as mountains, rivers, and seas. In particular, discontinuous jumps in genetic distance are often associated with geographical barriers (Rosenberg et al., 2005). In Europe, for example, sharp differences in genetic distance correspond to geographical hindrances, including high mountains and seas (Barbujani and Sokal, 1990). A major problem with the use of genetic distance in economics is therefore related to the possibility that the simple measure of distance (as captured by the kilometric distance between the capitals of two countries) is not sufficient to detect the relevance of geographical barriers in shaping genetic distances.

Given the established correlation between geographical barriers and genetic distance, we construct new measures of microgeography to control for the presence of geographical barriers that are not captured by simple distance, and we analyze the relationship between genetic distance, transportation costs, and trade. We argue that the simple geographic distance between two countries, which is the measure commonly used to control for geography, is not adequate because it does not take into account geographical barriers, such as mountains, seas, and rivers, that are crucial determinants of genetic differences. We show that these measures of geography explain satisfactorily both genetic distance and transportation costs. We also show that the explanatory power of genetic distance in explaining trade in a standard gravity framework decreases substantially, or even disappears, once we control for geographical factors.

2 Note that none of these reasons posit any hierarchy of genetic traits. (Pseudo)scientific theories that, sadly, claimed any genetic hierarchy have been disqualified.

We study trade because, among economic outcomes, it is connected both to economic and geographical characteristics and to differences in tastes and cultures, which are possibly correlated to vertically transmitted traits, as captured by genetic distance. In addition, trade can be studied with gravity models, which have become a benchmark in the study of genetic differences. Finally, the richness of trade data allows us to design a set of tests to disentangle the different mechanisms through which genetic distance may operate. We restrict our sample to Europe, because there is a considerable correspondence between genetically defined populations for which genetic distance is available and politically defined countries.[3] In addition, Europe allows us to control more easily for geographical factors.

Genetic distance, geographic barriers, and genealogical proximity of languages are intertwined, generating issues of multicollinearity and hampering the interpretation of the results. To determine whether the correlation between genetic distance and trade is spurious, we scrutinize the links between geography, trade, and genetic distance and propose a strategy to 'unbundle' their correlation.

We show that geographic factors, which shaped genetic patterns in the past but are also relevant for current transportation costs, explain the correlation between trading flows and genetic distance: after controlling for measures of microgeography, the impact of genetic distance on trade disappears. We go further, proposing a new classification of goods according to the ease with which they can be transported: if genetic distance is mostly capturing geographical barriers, bulky goods, which are difficult to transport, should be more affected by genetic distance than easy-to-transport goods. We indeed find that genetic distance, geographical barriers, and transportation costs are significant for bulky goods but less relevant or even insignificant for easy-to-transport goods.

This paper contributes to three lines of research. First, by finding a way of unbundling the correlation between genetic distance and geography, our paper contributes to the rapidly growing literature on the role of genetic differences in economics. Being relatively novel, the use of genetics in economic analysis is still fraught with potential pitfalls; our paper analyzes the challenges and opportunities of the use of genetics in economics and shows that it is crucial to take into account less obvious measures of geography when trying to interpret the relationship between genetic distance and economic outcomes.

Second, our paper contributes to the debate on the role of geography in economic development (see Sachs, 2003; Rodrik et al., 2004). Geography matters for development in ways that are not obvious, affecting, for example, the ethnic composition of a

3 Although the genetic and the political concepts of 'populations' are often used synonymously, they are clearly different. For example, Sardinians are a population genetically very different from the rest of Italians (the genetic distance between Sardinians and continental Italians is 221 while the difference between continental Italians and Swedes is 95; similarly, Lapps are very different from Finns). In this paper, we consider only the populations clearly associated to current countries. We therefore exclude Basques, Lapps, Sardinians, and Scots. Also, Cavalli-Sforza et al. (1994) provide a table with genetic distance for non-European native populations. However, extensive migrations, as well as the fluid, ever-changing boundaries between countries outside Europe, make it impossible to reconstruct reliable genetic distances between non-European countries. As a result, outside Europe, many genetically defined populations (i.e., populations within which there is random interbreeding) span several countries and, conversely, within many countries there are several genetically defined populations. (For a discussion of the meaning of population in genetics, see Section 2.1). Given these caveats, it is practically impossible to precisely define genetic distance between non-European countries.

country. For instance, Acemoglu et al. (2001) show how geography had an impact on settlers' mortality and so on the pattern of colonization. Alesina et al. (2011) show that those countries whose border shape does not reflect natural geographical barriers experienced a lower level of economic development. Our paper provides a further example of the role that geography may play in an indirect, although not less powerful, way on trade.

Third, our paper contributes to the literature on the determinants of trade and transportation costs. Several authors have shown that the simple measure of (log)-distance is only a first approximation for actual transportation costs (Hummels, 1998; Limao and Venables, 2001); in the context of the gravity model, many studies have included geographical variables such as insularity or contiguity to complement the standard crude measure of distance. Building on this tradition, we have shown that high-altitude mountain ranges, seas shared between countries, and ruggedness of the terrain also contribute to transportation costs.

The rest of the paper is organized as follows. Section 2 provides an overview of the available measures of genetic distance, and it shows that genetic distance may explain very well trade between European countries in a standard gravity equation with (log)-distance as a proxy of transportation costs; however, genetic distance loses significance once we control for measures of geographical barriers and/or transportation costs. Section 3 presents the results, separating goods by their degree of 'bulkiness'. Section 4 concludes.

2. Genetic distance, geography, and trade

2.1. How is genetic distance defined?

Several indices have been proposed to quantify the degree of genetic distance between two or more populations. In order to understand these indices, it is useful to review a few concepts in genetics. A gene is a sequence of DNA that encodes a protein. An allele is one of two or more versions of a gene. In some instances, different alleles can result in different observable phenotypic traits (i.e. an individual's observable characteristics), such as different skin or eye color. However, in most cases different alleles result in no observable variation. An allele is 'selectively neutral' if it does not provide any selective advantage to the individual who has it.

Within a population, multiple alleles for each gene are present. Importantly, the frequency of alleles differs across populations, so for example a given allele A for gene X might be present in a population with 75% frequency and in another with 25% frequency.[4] Genetic distance is a synthetic measure of the difference in allelic frequencies across populations. In theory, the allelic frequency of any gene (or set of genes) can be used to measure genetic distance. In practice, only a (limited) subset of alleles that have been proven to be selectively neutral is used. As was already clear to Darwin, selectively neutral alleles are best for reconstructing evolutionary history because, by not providing any selective advantage, they are uncorrelated with

4 Allele frequencies for various genes and for most populations can be found at http://alfred.med.yale.edu/.

environmental characteristics. They reflect only the evolutionary history and do not reflect environmental selective pressure.[5]

Among the genetic-distance indices, one is the F_{ST} distance of Cavalli-Sforza et al. (1994).[6] This index is based on the frequency of 128 alleles related to 45 genes and includes alleles coding for blood groups, lymphocyte antigens, immunoglobulin, hemoglobin, and enzymes.[7] The genes were selected so that they are (1) selectively neutral and (2) easy to collect. F_{ST} takes a value equal to zero if and only if the allele distributions are identical across the two populations, whereas it is positive when the allele distributions differ. A higher F_{ST} is associated with larger differences[8].

Given the choice of genes and alleles used for determining F_{ST}, it is not surprising that the pattern of overall genetic variation among populations differs substantially from traditional racial divisions. Populations that are similar in terms of phenotypes are not necessarily genetically close. The absence of correlation between genetic distance and phenotypic traits, like the color of the skin, is particularly intriguing and would argue against a relationship between 'cultural perception' and overall genetic features. In contrast, the correlation between genetic distance and geographical distance has been widely documented (Cavalli-Sforza et al., 1994; Rosenberg et al., 2005). Crucially, geographical barriers are associated with discontinuity in genetic distance, especially in Europe, where sharp increases in genetic distance correspond to major mountain ranges and seas (Figure 1). We use this observation to construct our measures of microgeography in the following section.

In this paper, we use the genetic-distance measure of Cavalli-Sforza for the European sample. The use of Cavalli-Sforza's measure of genetic distance outside Europe is problematic. Genetic distance is a measure across genetic populations, not across countries, where genetic population is defined as the group of individuals within which there is random interbreeding.[9] *Strictu sensu*, it would not make sense to talk about genetic distance between France and Italy if genetic populations and economic-political populations were not overlapping. In Europe, the two concepts (loosely speaking)

5 To understand why using nonselectively neutral genes can be misleading if the interest is evolutionary history, consider the following example. The allele associated with resistance to malaria is frequent in populations living in areas where malaria was traditionally present, even if these populations are historically very different and live in areas set widely apart.

6 Several methods have been used to measure the genetic composition of a population. Some of these techniques *directly* measure DNA alterations. Classical analysis, instead, measured the *result* of DNA alterations, that is, protein variations. The genetic data in Cavalli-Sforza et al. (1994) have been obtained from classical analysis, and theirs is the most extensive, comprehensive study performed on protein polymorphism. The ongoing Human Genome Diversity Project (HGDP) and (The International HapMap Consortium, 2005) will soon provide a wealth of data and information linked directly to the DNA status, but the results available so far and the analysis performed on these data are not exhaustive. Preliminary analysis, however, supports the notion that the major tenets of the classical protein polymorphism analysis correlate closely with this new, more extensive scrutiny. For this reason, in the present study, we have relied on the data as provided in Cavalli-Sforza et al. (1994).

7 For the list of genes and alleles used in the construction of the F_{ST} measure, see Cavalli-Sforza et al. (1994), pp. 383–468.

8 For a formal definition of the F_{ST} measure of genetic distance, see the Appendix.

9 The term 'population' (or more precisely Mendelian population) in genetics defines a group of individuals who interbreed *randomly*. This definition differs from the common use, especially among economists. For instance, basic genetics textbooks make the example that it is impossible to talk about the genetic population of the city of New York (see, for instance, Stone et al., 2007, p. 95) because groups of different ethnic origin do not interbreed randomly. As a corollary, it does not make sense to talk of (genetically defined) populations in countries with separate groups that do not usually interbreed.

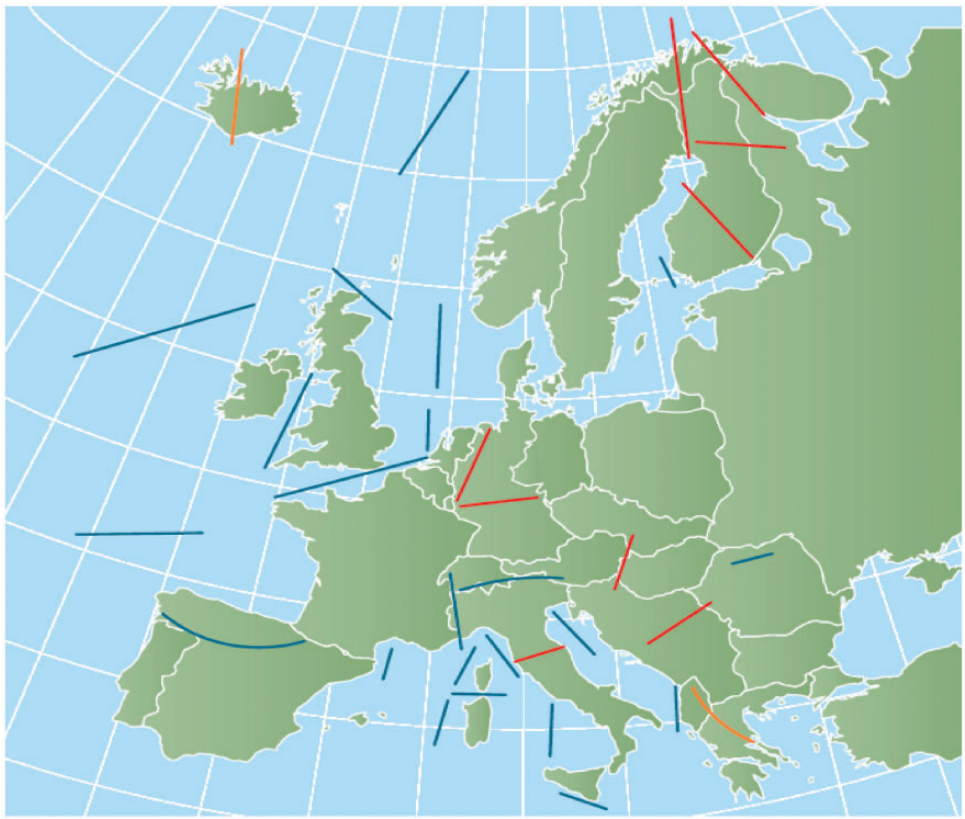

Figure 1. Zones of sharp genetic changes in Europe.
Source: Barbujani and Sokal (1990). Figure 1 shows 33 genetic boundaries in Europe; of these, 22 indicate zone of geographical barriers and nine indicate genetic boundaries, which are also linguistic boundaries but do not correspond to evident geographical barriers. The remaining two are genetic boundaries not associated to either geographical or linguistic boundaries.

overlap, but outside Europe it is clearly not the case. So, many genetic populations straddle several countries, and vice versa. In Asia, for example, there is almost no overlap between countries and populations.[10]

Errors in matching genetic populations to countries could be particularly severe and lead to substantial measurement error. For example, in Latin American countries, it is often difficult to ascertain whether populations are of European or Amerindian descent. This is particularly problematic in countries such as Columbia with large proportions of Mestizos, people of mixed descent. Since the main message of our paper is that genetic distance does not matter for trade, we want to use a sample where measurement error is minimized.

10 For Southeast Asia, see the table on page 236 of Cavalli-Sforza et al. (1994), where the genetic distance for populations refers to such populations as the Khasi, Mon Khmer, Semai, Negrito, Ami, Bunun, and Paiwas; the structure is similar for East and Central Asia, on page 230.

2.2. Data

The index of genetic distance F_{ST} is taken from Cavalli-Sforza et al. (1994). We use only the values in which there is a correspondence between genetic populations and countries.[11] We do not use data on subnational populations with distinct genetic information, including the Sardinians, Basques, and Lapps. The Cavalli-Sforza data cover 231 ($= 22 \times 21/2$) pairs of countries for Europe. Table A1 lists the countries in the sample. The measure of genetic distance for the European sample ranges from 0.009 (between Denmark and the Netherlands) to 0.317 (between the former Yugoslavia and Iceland). Other pairs in the low range include Germany–Switzerland (0.01), Belgium–Netherlands (0.012), and Austria–Switzerland (0.012). Other genetically distant populations include Greece–Ireland (0.289), Greece–Iceland (0.288), and the former Yugoslavia and Ireland (0.272). The average genetic distance is represented by pairs like Germany–Ireland (0.084), Sweden–Finland (0.082), and Sweden–Poland (0.082).

The bilateral export data are obtained from the United Nations' COMTRADE database as revised by Feenstra et al. (2005). The GDP data are obtained from the World Development Indicators of the World Bank; distance between capitals, common official language, and contiguity dummies are obtained from a dataset compiled at the Centre d'Etudes Prospectives et d'Informations Internationals (CEPII).[12] The export and GDP data refer to the 1975–2000 period.

We use a novel measure of transportation costs.[13] This measure (tc_{ij}) is taken from shipping company quotes collected by Import Export Wizard (IEW), a shipping company that provides estimates of transportation costs. IEW calculates the surface freight data based on a survey of intermodal and marine tariffs from carriers around the world. The variable tc_{ij} is the cost in U.S. dollars of transporting a '1000 kg unspecified freight type load (including machinery, chemicals, etc.) with no special handling required, using the optimal combination of going through land and water to transport the goods'. The advantage of this measure is that it represents the actual average transportation costs and not indirect measures or proxies, which are typically plagued by measurement error. The measures of transportation costs provided by IEW are symmetric.

To be sure that our results are not driven by our new measure of transportation costs, we also replicate our main results using an ad valorem measure of transportation costs. The most widely available data on ad valorem transportation costs is the ratio of carriage, insurance, and freight (c.i.f.) to free on board (f.o.b.) values that the IMF reports in its Direction of Trade Statistics for bilateral trade between countries.[14] Specifically, importing countries report the value of imports from partner countries, inclusive of c.i.f., and exporting countries report their value f.o.b., which measures the cost of the imports and all the charges incurred in placing the merchandise aboard a carrier in the exporting port. The ratio of c.i.f./f.o.b. data has also been used by Limao

11 For a description on how genetic-distance measures are defined at the population level, see Section 2.1, above.

12 The data on bilateral exports can be found at http://cid.econ.ucdavis.edu/. The data on distance, language, and contiguity are available at http://www.cepii.fr/anglaisgraph/bdd/distances.htm.

13 For a review of the literature on transportation costs, see Anderson and van Wincoop (2004).

14 We use data from the IMF's Direction of Trade Statistics from 1975 to 2000.

and Venables (2001) as a viable measure of transportation costs.[15] Following Hummels and Lugovskyy (2006), we restrict our analysis to ad valorem transportation costs, which range from 0 to 100%, considered a reasonable range of variation.

We construct a set of measures of geographical barriers using information on seas, mountain chains, and the topographical variability, or ruggedness, of countries. We define a variable (*mountains*), identifying major mountain chains between countries.[16] Following the World Atlas, major mountain chains in Europe are the Alps, the Apennines, the Atlantic Highlands (which include the Kjolen in Norway and Sweden, and the Pennines in the United Kingdom), the Balkan Mountains, the Massif Central, the Meseta, the Pyrenees, the Urals, the Carpathians, and the Caucasus Mountains. We define a dummy *common sea* equal to one if a pair of countries shares the same sea, which can be the Mediterranean Sea, the Atlantic Ocean, or the Northern/Baltic Sea. Of our 231 country pairs, 100 (43% of our sample) share a common sea. The number of mountain chains between countries ranges from zero to four (Table A3). Most of the countries in our sample are separated by zero or one mountain chain (46 and 37%, respectively.) Thirty-seven couples of countries are separated by two or three mountain barriers, whereas only two couples of countries have four mountains between them.[17]

Finally, we construct a variable measuring the topographical variability, or *ruggedness*, of those countries that lie between two trading partners. For instance, for the Germany–Italy pair, this variable measures the ruggedness of Germany, Austria, and Italy. The ruggedness measure has been constructed using data from Nunn and Puga (2012).[18] In addition, we also test the robustness of our results using four other measures constructed by the same authors. The first measure considers the average uphill slope of the countries' surface area. The second measure considers the average

15 Several issues have arisen regarding the use of data on c.i.f./f.o.b. transport costs (see Hummels, 1998). One is that the measure aggregates all commodities imported, so it is biased if trade on high-transport-cost routes systematically involved lower-transportation-cost goods. This suggests that the estimates will eventually underestimate the true magnitude of transportation costs. Another is the presence of measurement error, arising particularly from the fact that exports are not always accurately reported.

16 The distance between two countries is defined as the shortest land distance between major economic areas of the two countries.

17 There are many instances of two countries with no mountains between them, for example Austria and Hungary; two countries separated by one mountain range, for example Austria and Belgium (the Alps); two countries separated by two mountain ranges, for example Austria and Greece (the Dinaric Alps and the Alps), and two countries separated by three mountain ranges, for example Portugal and Greece (Dinaric Alps, Alps, and Pyrenees). The two country pairs separated by four mountain ranges are Macedonia and Portugal, and Macedonia and Spain (Rhodope Mountains, Alps, Dinaric Alps, and Alps).

18 To calculate their ruggedness measure, Nunn and Puga (2012) take a point on the earth's surface and calculate the difference in elevation between that point and the point on the grid 30 arc-seconds (roughly 1 km) north of it. The calculation is performed for each of the eight major directions of the compass (north, northeast, east, southeast, south, southwest, west and northwest). The terrain ruggedness index is given by the square root of the sum of the squared differences in elevation between the central point and the eight adjacent points. Formally, if e_{rc} denotes elevation at the point located in row r and column c of a grid of elevation points, then the terrain ruggedness index of Riley et al. (1999) at that point is calculated as $\sqrt{\sum_{i=r-1}^{r+1}\sum_{i=c-1}^{c+1}\left(e_{i,i}-e_{r,c}\right)}$. They then average across all grid cells in the country not covered by water to obtain the average terrain ruggedness of the country's land area. Since the sea-level surface that corresponds to a 30-by-30 arc-second cell varies in proportion to the cosine of its latitude, when calculating the average terrain ruggedness for each country, they weigh each cell by its latitude-varying sea-level surface. The units for the terrain ruggedness index correspond to those used to measure elevation differences. Therefore, ruggedness is measured in hundreds of meters of elevation difference for grid points 30 arc-seconds apart.

standard deviation of elevation within the same eight-cell neighborhood. The third measure is motivated by the possibility that what matters is having a large-enough area of sufficiently rugged terrain nearby, even if some portions of the country are fairly flat. The fourth measure is a population-weighted measure of ruggedness that allows for the possibility that ruggedness may be more important in areas that are more densely populated today.[19]

Topographical variability is particularly relevant in the determination of transportation costs. The literature on road construction and transportation has long observed that topographical characteristics such as terrain variability can affect the construction and maintenance costs of surface transport networks, as well as the costs to users of those networks.[20] A variety of studies also observe that an increasing road gradient significantly increases fuel consumption. Thus, by shaping the cost of building, maintaining, and using surface transport networks, terrain ruggedness can be a fundamental determinant of transportation costs. Overall, for the same horizontal distance, moving goods across variable terrain requires both more energy and more time. And since these costs are eventually embedded into freight charges, natural terrain variation can induce differences in the transportation infrastructure across countries. Descriptive statistics for all our variables are reported in Table A2.

Table 1 shows the correlation between genetic distance, transportation costs, and several measures of geography. Genetic distance strongly correlates (0.67) not only with the standard measure of geographic distance (the log (distance) between two countries) but also with the topographical variability or ruggedness between countries (0.31) and the number of mountain chains (0.21); the correlation is weaker (and negative) for the presence of a common sea (–0.03). The presence of a sea has an ambiguous effect in determining genetic differences between populations. Ancient migrations often followed sea coasts; having a sea is a unifying factor. At the same time, crossing large seas was relatively complicated, so islands are usually genetically isolated.

A similar degree of correlation is found between transportation costs and all the measures of geographic distances. The correlation is particularly strong for geographical distance (0.91), followed by the presence of mountain chains (0.35), the average ruggedness between countries (0.33) and, to a lesser extent, the presence of a common sea (0.09).

Figure 2 shows the correlation between our measure of transportation costs and genetic distance. The correlation is striking (the coefficient of the regression of transportation cost on genetic distance is equal to 0.45 and significant at the 1% level.) The R-squared of the regression is 0.42, suggesting that genetic distance is probably picking up geographical impediments that are relevant in determining transportation costs. The correlation between genetic distance and transportation costs persists even

19 Our regressions are robust to the inclusion of any of these four alternative measures. Results are presented in the online Appendix, Table B1.

20 For example, 'to reduce construction costs over severe terrain, roadways are often narrower with smaller embankments, limiting the maximum safe vehicular weight and creating congestion. At a 6.5 percent surface gradient, roads can be designed to tolerate a maximum speed of 60 km/h. But a 35 percent increase in the gradient cuts the safe design speed by about 58 percent' (World Bank, 2005). See also Aw (1981), Tsunokawa (1983), and Paterson (1987).

Table 1. Correlations between genetic distance, measures of geography, and transportation costs

	Log (gen. dist.)	Log (transp. costs)	Log (c.i.f./f.o.b. transp. costs)	Log (geogr. dist.)	Contiguity	Mountains	Common sea	Ruggedness
Log (genetic distance)	1							
Log (trans. costs)	0.66***	1						
Log (c.i.f./f.o.b. transp. costs)	0.28***	0.28***	1					
Log (geogr. dist.)	0.67***	0.91***	0.24***	1				
Contiguity	−0.34***	−0.44***	−0.13**	−0.55***	1			
Mountains	0.21***	0.35***	0.11**	0.40***	−0.17***	1		
Common sea	−0.03	0.09*	−0.09*	0.08*	0.05	−0.21***	1	
Ruggedness	0.31***	0.33***	0.21***	0.34***	−0.14***	0.72***	−0.47***	1

Note: *significant at 10%, **significant at 5%, ***significant at 1%.

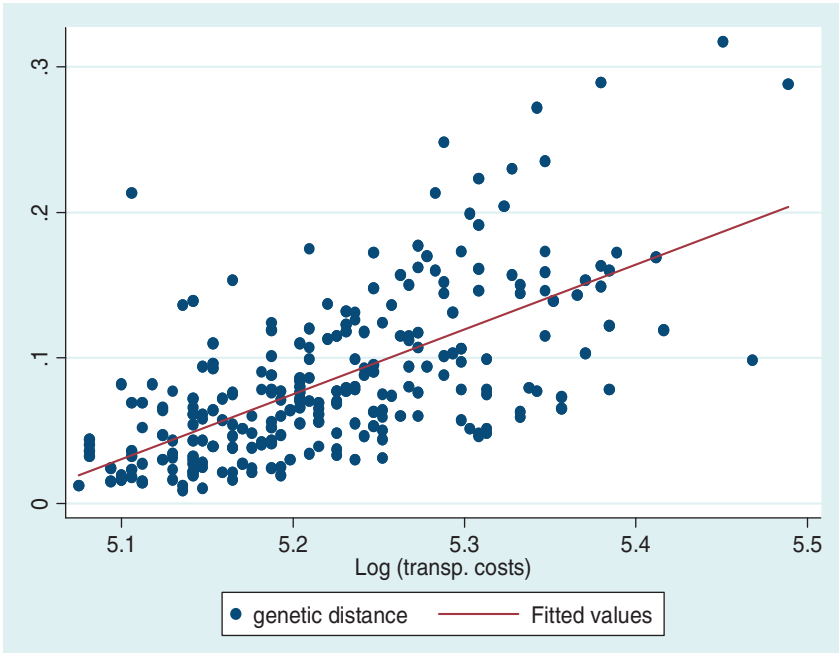

Figure 2. Genetic distance and transportation costs.

after we control for geographical distance, indicating that the simple distance is not enough to take into account geographical features (see Figure 3).[21] This motivates the following subsection, which takes a closer look at how geography affects genetic distance and transportation costs.

21 In this figure, we report the residuals of a regression of genetic distance on log distance and of transportation costs on log distance.

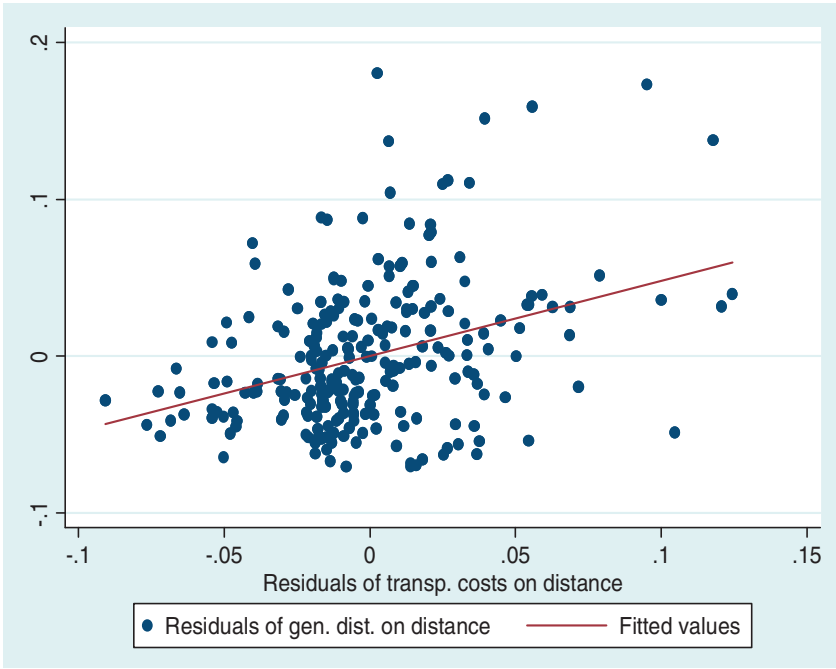

Figure 3. Genetic distance and transportation costs controlling for geographical distance.

2.3. Genetic distance, geography, and transportation costs

This section analyzes the relationship between genetic distance and geography. As discussed in the previous section, geography [including not only the (log) distance between countries, but also the presence of major mountains chains, the topographical variability between countries and the presence of shared seas] plays a fundamental role in explaining genetic distance, either by having determined past migration routes or by having separated populations, thereby contributing to the genetic drift. In this section, we argue that (1) the factors that determined genetic distance in the past also have a strong influence on current transportation costs today; and (2) once we properly control for transportation costs, the impact of genetic distance on trade is sensibly reduced or even disappears, indicating that the correlation between trade and genetic distance may be spurious.

In order to investigate more systematically how geographical factors have shaped genetic distance, we run a regression of (log) genetic distance between populations against several geographical barriers as control variables. As discussed before, the measure of genetic distance is derived from Cavalli-Sforza et al. (1994). We choose the geographical variables following the genetics literature. Our starting point is Barbujani and Sokal (1990), who show that the zones of abrupt genetic change in European populations correspond mostly to geographical boundaries.[22] Our measure of geographical barriers includes distance, number of mountains between countries, the

22 Barbujani and Sokal (1990) have identified 33 boundaries of sharp changes in gene frequencies across Europe. Specifically, the authors have counted 22 physical barriers, of which four are mountainous and 18 are marine boundaries. See Figure 1.

presence of a common sea, and the topographical variability, or ruggedness, between two trading countries (as defined above). We run two regressions on the geographical determinants of genetic distance and transportation costs, one without (Table 2) and one with (Table 3) country-of-origin and country-of-destination fixed effects, which are included to control for country-specific characteristics, such as insularity and remoteness.

The results presented in the first four columns of Table 2 confirm that geographical measures and genetic distance between European countries are indeed correlated. The most important variable (which explains 46% of variation in genetic distance) is simple geographic distance. The other geographical barriers, with the exception of the presence of a shared sea, are also significant and explain about 10% of the overall variation.

The last four columns of Table 2 report the results of regressions of transportation costs against measures of geographic distance. The importance of geography in determining transportation costs is well established; however, the standard measure of geography—(log) distance between (the capitals of) two countries—is considered only a first rough approximation of transportation costs (Hummels, 1998; Limao and Venables, 2001; Eaton and Kortum, 2002). Consistent with the transportation literature reviewed above, we find that geographical distance, number of mountain chains, and topographical variability between countries are positively correlated with transportation costs. The presence of a common sea increases bilateral transportation costs, but it is not significant. Geographical distance appears to explain a large part of the variation in transportation costs (the R-squared is 0.83), the topographical variability between countries explains roughly 11% of the variation in transportation costs, and the presence of mountain chains explains 10% of the variation; the presence of a common sea, however, explains almost zero percent of the variability. The results are robust to the inclusion of country-of-origin and country-of-destination fixed effects for both genetic distance and transportation costs (Table 3).[23]

The regressions reported in Tables 2 and 3 therefore show that geography (including the distance between countries, the topographical variability of the terrain, the presence of major mountains chains, and common seas) plays a big role in explaining genetic distance, either by having determined past migration routes or by having separated populations, thereby contributing to the genetic drift.

Given the strong correlation between geography and genetic distance, we hypothesize that (1) geography affects both genetic distance and, via transportation costs, trade flows and (2) that the correlation between trade and genetic distance is largely spurious. In the next section, we test this conjecture.

2.4. Genetic distance, transportation costs, and trade

To make our point, we estimate various versions of a standard gravity equation:

$$\ln(X_{ijt}) = \beta_0 + \beta_1 \ln(\text{gen. dist}_{ij}) + \beta_2 \ln(Y_{it}) + \beta_3 \ln(Y_{jt}) + \beta_4 \ln(D_{ij}) + \beta_5 C_{ij} + \beta_6 L_{ij} + \beta_7 E_{ijt} + \varepsilon_{ijt}$$

23 The presence of a common sea becomes negative and significant for both genetic distance and transportation costs, while the significance of mountain chains disappears in the regression of genetic distance.

Table 2. Geographic determinants of (log) genetic distance and transportation costs (without country-of-origin and country-of-destination fixed effects)

	Log (gen. dist.)	Log (gen. dist.)	Log (gen. dist.)	Log (gen. dist.)	Log (tr. costs)	Log (tr. costs)	Log (tr. costs)	Log (tr. costs)
Log distance	0.801				0.125			
	(0.061)***				(0.006)***			
Number of mountain chains		0.0264				0.029		
		(0.041)***				(0.005)***		
Common sea			−0.016				0.015	
			(0.095)				(0.011)	
Ruggedness				0.712				0.084
				(0.128)***				(0.015)***
Observations	231	231	231	231	231	231	231	231
R-squared	0.46	0.11	0.01	0.10	0.83	0.10	0.01	0.11

Note: *significant at 10%, **significant at 5%, ***significant at 1%. Robust standard errors are in parentheses.

Table 3. Geographic determinants of (log) genetic distance and transportation costs (with country-of-origin and country-of-destination fixed effects)

	Log (gen. dist.)	Log (gen. dist.)	Log (gen. dist.)	Log (gen. dist.)	Log (tr. costs)	Log (tr. costs)	Log (tr. costs)	Log (tr. costs)
Log distance	0.384				0.107			
	(0.054)***				(0.007)***			
Number of mountain chains		0.016				0.027		
		(0.039)				(0.008)***		
Common sea			−0.462				−0.073	
			(0.082)***				(0.010)***	
Average elevation b/w countries				0.125				0.148
				(0.024)***				(0.031)***
Observations	231	231	231	231	231	231	231	231
R-squared	0.87	0.83	0.86	0.84	0.92	0.71	0.72	0.73

Note: *significant at 10%, **significant at 5%, ***significant at 1%. Robust standard errors are in parentheses.

where X_{ijt} is the value of annual exports from country i to country j in year t, $\ln(\text{gen. dist}_{ij})$ is the log of the genetic-distance measure between country i and j as defined by Cavalli-Sforza et al. (1994), Y_{it} is the real GDP of country i, D_{ij} is the geographical distance between the capitals of countries i and j, C_{ij} is a dummy variable for geographic contiguity between country i and j, L_{ij} is a dummy variable for common language between countries i and j, E_{ijt} is a dummy equal to one if country i and country j both use the Euro at time t, and ε_{ijt} is the error term. We use a yearly panel from 1975 to 2000, controlling for year, country-of-origin fixed effects, country-of-destination

fixed effects, and country-specific trends.[24] We cluster the standard errors at the bilateral-country-pair level.

As a first pass in our analysis (Table 4, column 1), we simply show the impact of genetic distance on European trade in a standard gravity model without including any measure of geographic distance. The impact of genetic distance on trade is significant at the 1% level and with an elasticity of –0.850 (significant at the 1% level). As a second step (column 2) we introduce log (distance) and contiguity as proxies for geographical impediments to trade. The impact of genetic distance on trade is now strongly reduced, both in terms of significance (the coefficient becomes insignificant) and magnitude (the elasticity is 6.5 times smaller compared to a version of the gravity equation when we do not include any measure of geography among the controls).[25] As a final step, we run two specifications: one (column 3) in which we include all measures of geographical impediments discussed above (ruggedness between countries, number of mountain chains, and the presence of a common sea) and the other (column 4) in which we also control for transportation costs.[26] Both give similar results: after controlling for measures of geographical barriers or transportation costs, genetic distance does not have any significant effect on trade. The topographical variability of the terrain substantially reduces trade, whereas the presence of a common sea increases it. The number of mountain chains is not significant (this could be driven by the fact that the topographical variability of the terrain, a more precisely constructed measure, is capturing part of the effect). Transportation costs substantially reduce trade.[27]

Overall, our results indicate that genetic distance does not have an impact on trade once we properly control for those geographical impediments that shaped migration among populations and are very likely to still be relevant for transportation costs today.[28]

24 The country-specific dummies are meant to capture the multilateral resistance terms (see Anderson and van Wincoop, 2003).

25 Contiguity, distance, and language have all the expected sign, while the Euro variable is not significant.

26 To be sure that our results are not driven by our new measure of transportation costs, we replicate our main results using the ad valorem measure of transportation costs, calculated as the ratio of c.i.f. to f.o.b. values (see description in the data section above). The results (reported in Table B2 of the online Appendix) are consistent with the main findings of the paper: the impact of genetic distance disappears once we introduce measures of geographical barriers and transportation costs. The elasticity of trade to this measure of transportation costs is however low, this could be explained by the fact that measures of transportation costs based on the c.i.f./f.o.b. ratio cannot be used as a valid measure of transportation costs (see Hummels and Lugovskyy, 2006). However, the authors also show that these measures appear to be helpful when considering fitted values of c.i.f. and f.o.b. against plausible correlates revealing true transportation costs, such as geographic distance. When we run the gravity equation using the fitted value of transportation costs (obtained when considering all our measures of geography as correlates), we find an elasticity of –2.72 (with a standard error of 0.559). This value is very close to what was reported in Limao and Venables (2001).

27 Overall, we find that the presence of a common sea facilitates trade. But the sea could also be a proxy for a geographical impediment. To disentangle the opposing effect of the 'common sea variable', we interacted its effect with the variable indicating the presence of a common land border (contiguity). The interaction is not significant (the coefficient is –0.010, with a standard error of 0.157); at least for our sample, the presence of a common sea simply facilitates trade.

28 In the online Appendix, we show that the results do not depend on the specific functional form of genetic distance by running a more flexible functional form, which uses quintile dummies to study the impact of genetic distance on trade (Table B3 in the online Appendix). Similarly to our main specification, we do find that genetic distance matters when geographical controls are not included. However, its impact disappears once we properly control for geography.

Table 4. Genetic distance, transportation costs, and trade (dependent variable: log total exports)

	(1)	(2)	(3)	(4)
Log (genetic distance)	−0.850*** (0.105)	−0.131 (0.097)	0.009 (0.095)	0.064 (0.094)
Log (distance)		−0.919*** (0.095)	−0.819*** (0.087)	−0.298** (0.123)
Contiguity		0.240** (0.108)	0.280*** (0.104)	0.332*** (0.100)
Ruggedness			−0.615** (0.257)	−0.536** (0.245)
Number of mountain chains			0.041 (0.064)	0.040 (0.064)
Common sea			0.411*** (0.110)	0.307*** (0.100)
Log (transp. costs)				−5.734*** (1.220)
Common language	0.711*** (0.203)	0.578*** (0.185)	0.684*** (0.169)	0.765*** (0.143)
Euro	0.070 (0.070)	0.044 (0.062)	0.025 (0.062)	0.013 (0.060)
Log (GDP importer)	0.273 (0.362)	0.317 (0.360)	0.309 (0.361)	0.320 (0.359)
Log (GDP exporter)	0.432* (0.244)	0.467* (0.238)	0.462* (0.238)	0.464* (0.236)
Observations	9,200	9,200	9,200	9,200
R-squared	0.851	0.873	0.877	0.880

Note: *significant at 10%, **significant at 5%, ***significant at 1%. Errors are clustered at the bilateral country-pair level. Each regression controls for country and year fixed effects and country-specific trends.

3. Further probing the relationship between trade, genetic distance, and geography

In the previous section, we showed that the impact of genetic distance on trade disappears with the inclusion of measures of geographical barriers. We therefore concluded that genetic distance could be simply a more precise proxy of geographical impediments but not cultural differences in the determination of trade in Europe. To dig deeper into the relationship between genetic distance and trade, we propose a test based on the ease of transporting goods. If geography explains the association between trade and genetic distance, the elasticity of trade flows to genetic distance should be higher for bulky goods.[29] Similarly, there is no reason as to why genetic distance should favor trade in goods that are easy to move.

3.1. Bulky vs. non-bulky goods

We construct the bulkiness index by looking at the freight-to-value ratio for U.S. imports from Mexico and Canada at four SITC digits.[30] We classify as 'bulky' all goods whose freight-to-value ratio is higher than the median; we classify the other half of the goods as 'easy to transport'. Table 5 reports the results of our baseline specification for easy-to-transport (first four columns) and bulky goods (last four columns).

When we run a regression of exports for easy-to-transport and bulky goods on only genetic distance, the elasticity of this measure is larger for bulky than for easy-to-transport goods (columns 1 and 5). The coefficient on genetic distance when controlling

29 We thank David Hummels for useful discussion on this particular index and for providing the data to construct our test.

30 We use U.S. data because it gives detailed information on freight rate and values at the four-digit level. We chose imports from Mexico and Canada because as they are contiguous to the United States, all modes of transportation are used: sea, land, and air. We assume that the ranking of the freight-to-value ratio is the same in North America as in Europe.

Table 5. Genetic distance, transportation costs, and trade (dependent variable: log of bulky and easy-to-transport goods)

	(1) Easy	(2) Easy	(3) Easy	(4) Easy	(5) Bulky	(6) Bulky	(7) Bulky	(8) Bulky
Log (genetic distance)	−0.728*** (0.111)	−0.074 (0.110)	0.035 (0.114)	0.106 (0.112)	−0.920*** (0.108)	−0.183* (0.102)	−0.031 (0.098)	0.019 (0.097)
Log (distance)		−0.848*** (0.099)	−0.806*** (0.100)	−0.126 (0.149)		−0.936*** (0.100)	−0.806*** (0.091)	−0.316** (0.126)
Contiguity		0.210 (0.127)	0.234* (0.126)	0.289** (0.122)		0.281** (0.111)	0.330*** (0.107)	0.377*** (0.103)
Ruggedness			−0.619** (0.314)	−0.488 (0.299)			−0.678*** (0.255)	−0.597** (0.246)
Number of mountain chains			0.092 (0.076)	0.091 (0.074)			0.020 (0.066)	0.019 (0.066)
Common sea			0.230* (0.133)	0.110 (0.127)			0.467*** (0.112)	0.371*** (0.104)
Log (transp. costs)				−7.518*** (1.318)				−5.375*** (1.202)
Common language	0.574*** (0.206)	0.448** (0.186)	0.537*** (0.179)	0.650*** (0.149)	0.779*** (0.209)	0.615*** (0.194)	0.727*** (0.175)	0.805*** (0.151)
Euro	0.091 (0.083)	0.065 (0.079)	0.051 (0.079)	0.036 (0.078)	0.056 (0.073)	0.027 (0.063)	0.006 (0.062)	−0.006 (0.060)
Log (GDP importer)	0.033 (0.294)	0.091 (0.295)	0.089 (0.295)	0.107 (0.294)	0.352 (0.244)	0.402* (0.244)	0.395 (0.245)	0.414* (0.246)
Log (GDP exporter)	0.696* (0.382)	0.734* (0.380)	0.737* (0.379)	0.736* (0.381)	0.032 (0.295)	0.089 (0.294)	0.082 (0.295)	0.100 (0.296)
Observations	9,084	9,084	9,084	9,084	9,290	9,290	9,290	9,290
R-squared	0.874	0.892	0.893	0.899	0.842	0.868	0.872	0.875

Note: *significant at 10%, **significant at 5%, ***significant at 1%. Standard errors are clustered at the bilateral country-pair level. Each regression controls for country and year fixed effects and country-specific trends.

for the standard measure of geography (log distance and contiguity) is still high and significant for bulky goods, but it becomes much smaller (and not significant) for easy-to-transport goods, an indication that genetic distance is still proxying for other geographical characteristics for bulky goods. Finally, when we include our measures of microgeography, the coefficient on genetic distance becomes small and not significant for both types of goods. Also, while the measures of microgeography are not relevant for the case of easy-to-transport goods, geographical barriers, as expected, tend to play a much more important role for bulky goods. When geography is not relevant, as for the case of easy-to-transport goods, genetic distance plays no role. On the other hand, the inclusion of measures of geographical barriers and transportation costs makes the effects of genetic distance on trade disappear for bulky goods, for which transportation costs and geographical barriers are very relevant.

In conclusion, classifying goods according to their ease of transport allows us to perform a better test on the importance of geography in the determination of trade. Using our measure of bulkiness, we find that genetic distance matters only for bulky goods and not for easy-to-transport goods when we do not control for more sophisticated measures of geography. The effect also disappears for bulky goods once geography is properly controlled for. We take this as a strong indication that, even though genetic distance, geographical barriers, and cultural differences are correlated, trade is explained mostly by transportation costs.

4. Conclusions

Our paper is motivated by the finding that genetic distance and trade flows are correlated. Differences in genetic makeup between populations could be correlated with trade because genetic distance could be a proxy for cultural differences between countries.

In this paper, we show that, at least for the case of trade flows, genetic distance is proxying for the same geographical factors that shaped genetic differences across populations, mostly in the Neolithic Period, and that are also responsible for influencing transportation costs nowadays. The standard measures of geography used in a gravity framework (such as log distance and contiguity) are not sufficient to pick up the relevance of geography on trade. Other features (including ruggedness of the terrain, the presence of mountain chains between countries or the presence of a common sea) appear to have a relevant effect. That additional effect is what genetic distance seems to capture.

We make our point by constructing a database on geographical impediments, by introducing a novel database on transportation costs, and by proposing a new classification of goods according to the ease with which they can be transported. Genetic distance could therefore be largely a proxy for transportation costs and geographical impediments, and economists should be careful when using it as a proxy for vertically transmitted characteristics.

This paper makes several contributions. First, it proposes a methodology to 'unbundle' the correlation between genetic distance and geographical barriers. Second, it contributes to the trade literature by introducing two new databases on transportation costs and geographical barriers. Third, it contributes to the study of the role of long-term cultural factors and geography in economic development.

Supplementary material

Supplementary data for this paper are available at *Journal of Economic Geography* online.

References

Acemoglu, D., Johnson, S., Robinson, J. A. (2001) The colonial origins of comparative development: and empirical investigation. *American Economic Review*, 91: 1369–1401.

Alesina, A., Easterly, W., Matuszeski, J. (2011) Artificial states. *Journal of the European Economic Association*, 9: 246–277.

Anderson, J., van Wincoop, E. (2003) Gravity with gravitas: a solution to the border puzzle. *American Economic Review*, 93: 170–192.

Anderson, J., van Wincoop, E. (2004) Trade costs. *Journal of Economic Literature*, 42: 691–751.

Aw, W. B. (1981) *Highway Construction Cost Model for Sector Planning in Developing Countries*, Mimeo. MA: MIT.

Barbujani, G., Sokal, R. R. (1990) Zones of sharp genetic change in Europe are also linguistic boundaries. *Proceedings of the National Academy of Sciences USA*, 87: 1816–1819.

Cavalli-Sforza, L. L., Feldman, M. W. (1981) *Cultural Transmission and Evolution*. Princeton, NJ: Princeton University Press.

Cavalli-Sforza, L., Menozzi, P., Piazza, A. (1994) *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.

Desmet, K., Le Breton, M., Ortuño-Ortín, I., Weber, S. (2011) The stability and break up of nations: a quantitative analysis. *Journal of Economic Growth*, 16: 183–213.

Eaton, J., Kortum, S. (2002) Technology, geography and trade. *Econometrica*, 70: 1741–1779.

Feenstra, R., Lipsey, R., Deng, H., Ma, A., Mo, H. (2005) *World Trade Flows: 1960-2000*. NBER Working Paper 11040. Cambridge, MA: National Bureau of Economic Research.

Guiso, L., Sapienza, P., Zingales, L. (2009) Cultural biases in economic exchange. *The Quarterly Journal of Economics*, 124: 1095–1131.

Hummels, D. (1998) *Towards a Geography of Transport Costs*. Mimeo. Chicago: University of Chicago, Department of Economics.

Hummels, D., Lugovskyy, V. (2006) Are matched partner trade statistics a usable measure of transportation costs? *Review of International Economics*, 14: 69–86.

Limao, N., Venables, A. (2001) infrastructure, geographical disadvantage, transport costs and trade. *World Bank Economic Review*, 15: 451–479.

Nunn, N., Puga, D. (2012) Ruggedness: the blessing of bad geography in Africa. *Review of Economics and Statistics*, 94: 20–36.

Paterson, W. D. O. (1987) *The Highway Design and Maintenance Standards Model (HDM-III), Volume III, Road Deterioration and Maintenance Effects: Models for Planning and Management*. Washington: World Bank, Transportation Department.

Riley, S. J., De Gloria, S. D., Elliot, R. (1999) A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Science*, 5: 1–4, 23–27.

Rodrik, D., Subramanian, A., Trebbi, F. (2004) Institutions rule: the primacy of institutions over geography and integration in economic development. *Journal of Economic Growth*, 9: 131–165.

Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K., Feldman, M. W. (2005) Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genetics*, 1: 70.

Sachs, J. (2003) *Institutions Don't Rule: Direct Effects of Geography on Per Capita Income*, NBER Working Paper 9490. Cambridge, MA: National Bureau of Economic Research.

Spolaore, E., Wacziarg, R. (2009) The diffusion of development. *The Quarterly Journal of Economics*, 124: 469–529.

Stone, L., Lurquin, P. F. (2007) *Genes, Culture, and Human Evolution: A Synthesis*. Malden, MA: Wiley-Blackwell.

The International HapMap Consortium. (2005) A haplotype map of the human genome. *Nature*, 437: 1299–1320.

Tsunokawa, K. (1983) *Evaluation and Improvement of Road Construction Cost Models* (unpublished draft working paper, Washington: World Bank).
World Bank. (2005) Low cost design standards for rural road projects. Document # 4654RO/B.1./3a/3.5/010. Washington: World Bank.

Appendix

Definition of Genetic Distance

In this paper, we use a F_{ST} measure of genetic distance (Cavalli-Sforza et al., 1994). The F_{ST} measure is based on indices of heterozygosity, the probability that two alleles at a given locus selected at random from two populations will be different. F_{ST} takes a value equal to zero if and only if the allele distributions are identical across the two populations, whereas it is positive when the allele distributions differ. A higher F_{ST} is associated with larger differences.

We describe here the construction of genetic distance for the case of two populations (a and b) of equal size and one gene that can take only two forms (allele 1 and allele 2).[31] Homozigosity and heterozigosity are defined, respectively, as the probability that two randomly selected alleles at the given locus are identical and different. Calling, p_a and $q_a = 1 - p_a$ the gene frequencies of allele 1 and allele 2 in population a, homozigosity is given by $p_a^2 + q_a^2$; whereas heterozigosity is given by $1 - p_a^2 - q_a^2$. Since $p_a + q_a = 1$, one can also write $(p_a + q_a)^2 = 1$, which implies that $1 - p_a^2 - q_a^2 = 2p_a q_a$.

The average gene frequencies of allele 1 and 2 in the two populations are, respectively, $\bar{p} = (p_a + p_b)/2$ and $\bar{q} = (q_a + q_b)/2$.

We can define heterozigosity in the sum of the two populations as $h_m = (h_a + h_b)/2$ and average heterozigosity as $h = 2\overline{pq}$.

F_{ST} measures the variation in the gene frequencies of populations, obtained by comparing h and h_m:

$$F_{ST} = 1 - \frac{h_m}{h} = 1 - \frac{p_a q_a + p_b q_b}{2\overline{pq}} = (p_a - p_b)^2 / 4\bar{p}(1 - \bar{p}).$$

If the two populations have identical allele frequencies ($p_a = p_b$), F_{ST} is zero. On the other hand, if the two populations are completely different at the given locus ($p_a = 1$ and $p_b = 0$, or $p_a = 0$ and $p_b = 1$), F_{ST} takes the value of 1. In addition, the higher the variation in the allele frequencies across the two populations, the higher is their F_{ST} distance.

31 For a generalization to L alleles, S populations and different population sizes see Cavalli-Sforza et al., 1994.

Table A1. Countries included in the sample

Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Macedonia, Netherlands, Norway, Poland, Portugal, Russia, Spain, Sweden, Switzerland, United Kingdom

Table A2. Summary statistics

	Mean	Std. Dev.	Observations
Log (distance)	7.160	0.624	231
Contiguity	0.121	0.327	231
Common language	0.021	0.146	231
Log (genetic distance)	4.209	0.729	231
Number of mountain chains	0.771	0.881	231
Ruggedness	0.899	0.327	231
Common sea	0.433	0.432	231
Log (GDP exporter)	25.97	1.41	9,200
Log (GDP importer)	25.97	1.41	9,200
Log (total exports)	12.90	2.36	9,200
Log (easy-to-transport goods)	11.69	2.49	9,084
Log (bulky goods)	12.50	2.31	9,290
Log (transport costs)	5.22	0.082	9,200
Euro	0.024	0.15	9,200

Table A3. Distribution of shared mountains and seas

	Frequency	Percentage
Sea		
0	131	56.71
1	100	43.29
Mountains		
0	106	45.89
1	86	37.23
2	27	11.69
3	10	4.33
4	2	0.87