

Genetic, Cultural and Geographical Distances¹

April 2006
Revised: May 2007

Paola Giuliano
Research Department, IMF; Harvard University; IZA

Antonio Spilimbergo
Research Department, IMF; CEPR; WDI

Giovanni Tonon
Dana-Farber Cancer Institute, Harvard University

Abstract

Genetic distance and trade flows are strongly correlated. This paper argues that this surprising correlation is largely explained by geography because the same geographical barriers which shaped genetic patterns play a role in contemporary transportation costs. Using a newly-collected database on transport costs and detailed measures of geography within Europe, we show that i) geography explains both genetic distance and transportation costs between European countries, and ii) the role of genetic distance in explaining trade between European countries disappears once we control for geography. We also restrict the sample to goods with an organized market (using Rauch classification) and to “bulky” and “non-bulky” goods (proposing a new classification) to probe further the relationships between genetic distance, geography, and culture. As a corollary, we show that genetic distance can be used as an instrument for transportation costs.

JEL Classification Numbers: Z10, F10

Keywords: transport costs, genetics, trade, cultural economics, geography.

¹Authors' E-mails: giuliano@fas.harvard.edu; aspilimbergo@imf.org; Giovanni_Tonon@dfci.harvard.edu. The views expressed in this paper are those of the authors and do not necessarily represent those of the IMF, its Board of Directors, or the views of the Dana Farber Cancer Institute. We thank Alberto Alesina, Andrea Bianchi, Francesco Caselli, James Fearon, Oded Galor, Luigi Guiso, Peter Howitt, David Hummels, Simon Johnson, Ross Levine, Nuno Limao, Paolo Mauro, Enrico Spolaore, Arvind Subramanian, Antonio Terraciano, Alessandro Vanzetti, Romain Wacziarg, David Weil, Luigi Zingales, seminar participants at Brown University, the International Monetary Fund, the University of California at Berkeley, the April 2006 NBER workshop on political economy and the workshop on the “Economics of Diversity, Migration and Culture” in Bologna, and José Romero for excellent research assistantship.

I. INTRODUCTION

Genetic distance and economic outcomes are strongly correlated. This surprising and strong correlation cannot be dismissed as a statistical fluke and calls for an explanation.² This paper has three goals: 1) to document further the robustness of the correlation between genetic distance and trade flows; 2) to probe how geography may explain this correlation; and 3) to explore uses of genetic distance in economics in a way which is consistent with these findings. Before moving to these issues, an introduction to the concept of genetic distance is in order.

What is genetic distance? Genetic distance is an index that measures the degree of genetic differentiation among two or more populations using allelic frequencies of a set of genes in different populations. The particular measure of genetic distance used in this paper was introduced by Cavalli-Sforza and his collaborators more than a decade ago. After collecting an impressive database on genetic distances among various populations, Cavalli-Sforza has argued that there is a strong correlation between genetic patterns and ancient migrations taking place especially in Neolithic times (Cavalli-Sforza, Menozzi, and Piazza, 1994).³ Cavalli-Sforza's successful and convincing use of genetics in anthropology and historical linguistics has attracted the attentions of scholars from other social sciences, including economists, who have been looking for quantifiable measures of differences among populations.

The use of genetic distance in economics is new and promises the same revolution as it did in other social sciences. Indeed, three intriguing papers have proposed to use genetic distance as a proxy for "vertically transmitted traits" (Spolaore and Wacziarg, 2006), as an instrument for bilateral trust between populations (Guiso, Sapienza, and Zingales, 2005), or broadly as a proxy for cultural values (Desmet, Le Breton, Ortuño-Ortín, and Weber, 2006). Spolaore and Wacziarg interpret the Cavalli-Sforza's index as a measure of "vertically transmitted characteristics," reflecting different historical paths of populations over the long run, and argue that bilateral differences in per-capita income levels are strongly correlated with genetic distance. Guiso, Sapienza, and Zingales (2005) argue that degree of bilateral trust between countries is an important determinant of trade within Europe. Since trust is obviously

² Spolaore and Wacziarg, 2006, document a correlation between genetic distance and differences in income per capita.

³ Section 2 below defines the measure of genetic distance.

endogenous, they propose genetic distance as an instrument for it. Being pre-determined, genetic distance is unlikely neither to be related to current economic activity nor to be influenced by it (Guiso, Sapienza, and Zingales, 2005). Desmet, Le Breton, Ortuño-Ortín, and Weber (2006) provide a detailed analysis of the correlation between genetic distance and World Values Surveys and conclude that genetic distance could be used as a proxy for cultural differences in studying formation of nations.

Our paper adds to this literature by exploring more deeply the correlation between genetic distance and trade and arguing that the strong correlation between these two variables is largely to ascribe to the geographic distance between countries, as determined by more accurate measures of geography. As corollary, we also argue that genetic distance could be used as an exogenous measure of transportation costs.

Our paper stems from the striking observation that genetic distance and economic outcomes are strongly correlated. Indeed, trade flows between European countries are strongly correlated with genetic distance: the more two countries are genetically similar, the more these countries trade between themselves. After reviewing both the genetic and anthropological literature, we hypothesize that geography might provide an explanation for this correlation. However, we argue that the simple geographic distance between two countries, which is the commonly used measure for controlling for geography, is not adequate because it does not take into account geographical barriers, such as the presence of mountains, seas, and rivers, that appear to be a crucial determinant of genetic differentiation. 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 economic outcomes decreases substantially or even disappears once we control for geographical factors.

Genetic distance, geography, language and any type of “vertically transmitted characteristics” are intertwined, so the previous finding per se does not necessarily mean that cultural factors, as captured by genetic distance, are not linked to economic outcomes. To determine whether the correlation between genetic distance and economic outcomes is spurious we look more carefully at the links between geography, economic outcomes and genetic distance as a generic proxy for cultural similarities and propose a strategy to “unbundle” their correlation.

Our strategy is based on the following assumption. Take two populations separated by a tall mountain chain. The trade between the two populations may be scarce because of direct

geographical barriers and/or, indirectly, because geographic barriers have caused the two cultures to diverge, and ultimately, to have developed some cultural characteristics that did not facilitate trade. Total trade is affected, but different goods are sensitive in a different way to transportation costs –direct effect of geography – or cultural differences – indirect effect of geography through cultural variables. Based on this intuition, we propose a test based on the characteristics of different goods. We classify the goods using Rauch’s classification (Rauch, 1999) and propose a new classification of “bulkiness”. If geography is important, bulky goods that are difficult to transport should be more affected by geography than non-bulky goods. On the other hand, if broadly defined cultural factors have a predominant role in affecting trade independently from geographic barriers, both bulky and non-bulky goods should be equally affected by geographic barriers. By splitting the sample between goods with organized markets or according to the easiness to transport, leads to the same result: transport costs are significant for bulky goods and/or homogeneous goods while genetic distance is never significant where we expected it to be significant, i.e. in goods for which cultural differences may be relevant.

Our findings caution that the correlation between trade and genetic distance should be interpreted with care, since genetic distance is likely a proxy for geographic distance and barriers among populations. In particular, our findings call for a detailed control of geographical features in regressions in which genetic distance is used to explain economic outcomes. However, this is not to say that cultural variables *per se* do not exert any role in explaining economic outcomes, but that any cultural traits considered, are most likely not closely captured by genetic distance.

To make our point we considered trade among European countries as a main economic outcome. We chose Europe because there is a considerable overlap between genetically defined populations and politically defined countries.⁴ We chose trade because, among other economic outcomes, it is clearly connected to cultural traits, including trust and similarity of tastes. In addition, trade allows us interesting robustness tests selecting only groups of goods for which

⁴Although the biological and the political concepts of “populations” are often used as synonymous, they are clearly different. For example, within Italy Sardinians are a population genetically very different from the rest of Italians (the genetic distance between Sardinians and Italians is 221 while the difference between Italians and Swedes is 95); similarly the Lapps are very different from the Finns. In our paper, we consider only the populations associated to political countries, this implies that from Cavalli-Sforza’s specification, we exclude Basque, Lapp, Sardinian and Scottish. Spolaore and Wazziarg (2006) solve this problem by mapping data for 42 genetic populations into more than 120 countries in the world.

the effect of cultural traits should be more relevant. Finally, gravity equations provide an established benchmark to test hypotheses in trade.

The present paper contributes to three lines of research. One is the study of the importance of culture on economic outcomes. Cultural factors have a strong influence on economic and social phenomena. Although this proposition has a long intellectual history⁵ only recently there has been a systematic interest in studying the impact of culture on economic outcomes from a quantitative point of view. To this tradition belong two recent papers by Barro and McCleary (2003) and Tabellini (2005). The first shows that the growth rates across countries are significantly affected by the religiosity of a country.⁶ The second argues that cultural traits such as social capital, a sense of generalized morality, and the types of values parents transmit to their children are relevant to explain economic development in the regions of Europe.⁷ Culture is a very elusive concept and it is very challenging to prove an unequivocal causal relationship between culture and economic outcomes. One of the main challenges is indeed to certify that apparently exogenous measures are not capturing omitted variables. One contribution of this paper is to show that it is necessary to control carefully for geography when interpreting genetic distance as a broad proxy for cultural variables.

Second, our paper contributes to the debate on the role of geography in economic development (see Rodrik, 2002, or Sachs, 2003, for a summary). Geography matters for development in ways that are not obvious, including by influencing the ethnic composition of a country. For instance, Acemoglu, Johnson and Robinson (2001) show how geography had an impact on settlers' mortality and so on the pattern of colonization; Alesina et al. (2006) 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, but not less powerful way, on economic development.

⁵ Social scientists have argued that cultural innovations have led to the development of capitalism (Weber, 1958) or that different historical and culture experiences in Italian regions have led to different development paths (Putnam, 1993 and Banskfield, 1958).

⁶ Barro and McCleary (2003) moreover show that beliefs rather than church attendance are what matters for growth.

⁷ For a review of papers looking at microeconomic evidence on the importance of culture on economic outcomes, see Guiso, Sapienza and Zingales (2006).

Finally, our paper contributes to the literature on the determinants of transportation costs. The paper provides an original contribution to this literature in two ways. Several authors have shown that the simple measures of (log)-distance is only a first approximation for true transport costs (Hummels, 1998; Limao and Venables, 2001); in the context of the gravity models, 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 major mountains, common seas and countries elevation also contribute to transport costs. Second, we use a new dataset on transportation costs. The currently used measures of transportation costs are indirect measures, plagued by measurement errors; our measure represents the actual transport costs, allowing us to study more reliably the importance of transportation costs on trade.

The rest of the paper is organized as follows. Section II provides an overview of the available measures of genetic distance, highlighting how they are calculated, what they measure and their relationship with physical anthropology data, including anthropometric characters like stature or qualitative traits such as eye color or skin pigmentation. Section III shows that genetic distance may explain very well trade between European countries in a standard gravity equation with (log)-distance as a proxy of transport costs, but loses significance once transport costs, properly instrumented, are introduced. Section IV presents the results separating bulky goods from goods in which networks are important. Section V discusses alternative uses of genetic distance in economics; section VI concludes.

II. WHAT IS IN THE GENES?

Population genetics studies populations' genetic composition and their changes over time, focusing on genes that are present in at least two different forms (alleles) in the population. In its simplest form, the fundamental measurement in population genetics is the frequency at which alleles are found at any specific *gene locus* (allele frequency).⁸

⁸ With the term *gene locus* (or, for simplicity, gene) we intend a sequence of DNA that encodes for a protein, and with the term *allele* we consider a particular form of a specific gene. Often alleles are distinguished for their effects on the phenotype (e.g., morphological, physiological or biochemical characteristics of an individual – or group of like individuals – that differ in this respect from other individuals), or, simply, for differences in pair sequence. Several different methods have been used to

(continued)

Although not all alleles occur in all human populations, differences in alleles within local human populations are much greater than among different populations. Specifically, 93% of total human variability is found within local populations. The remaining 7% is found between populations (Rosenberg et al., 2002). As noted by Lewontin, “if everyone on earth becomes extinct except for the Kikuyu of East Africa, about 90% of all human variability would still be present in the reconstituted species (Lewontin, 1984). The genes that are used in population genetics for determining genetic distance belong to this second, small group.

The central tenet of human population genetics is that phenotypic characteristics (including anthropometric characters like stature or qualitative traits such as eye color, hair form, facial form or skin pigmentation; in a word, why a person from Finland looks different from a person from Sardinia) are not related to the overall genetic structure of human populations, as it is measured in this study or, for that matter, by any other index used in human population genetics. Through history, the genomes of human populations have been shaped by forces affecting all genes in the same time, whereas natural selection acts on phenotypes (like the color of the skin) that are controlled by a very small subset of specific genes, largely unknown and almost certainly not included in the limited number of genes used in population genetics to determine the genetic distance between populations (Cavalli-Sforza 1966; Lewontin and Krakauer 1973). Recent analyses using the HapMap data agree that the percentage of genes that have been under natural selection in human populations – and therefore linked to phenotypes under strong selective pressure like the color of the skin -- is between 1.6 and 2.6% of all genes (Carlson et al. (2005); Voight et al. (2006), Kelley et al. (2006), Wang et al. (2006)). Therefore, the genes linked for phenotypes upon which the characterization of different races is based, as for

measure the genetic composition of a population. Some of these techniques are directly linked to DNA alterations. Classical analysis instead measured the result of DNA alterations, that is, protein variation. The most extensive and comprehensive studies on variants have been performed on protein polymorphism. The ongoing Human Genome Diversity Project (HGDP) and the International HapMap Project 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, as presented in its more comprehensive form by Cavalli-Sforza et al. in 1994, correlates 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. However, the correspondence between genetic data as collected by Cavalli-Sforza and modern countries is available mostly for Europe.

example the form of the hair, the facial form and the skin color are not a random sample of the overall genetic structure of a population, and most importantly, as mentioned, such genes are not included in the alleles used for measuring the genetic distance between populations.

In contrast, likely all the genes used to compute genetic distance are selectively neutral, that is, they lack selective advantage (see “Neutral Theory of Evolution”, Kimura, 1968). As already clear to Darwin, neutral characters are best for reconstructing evolutionary history. If the genes used for the analysis show intercorrelated responses to the various environments in which human evolution has occurred, the measured genetic distance would be a reflection of the environment rather than of evolutionary history.

Subsets of the specific group of genes that varies between populations and that are selectively neutral are used to reconstruct the evolutionary history of populations. Genetic variation among human populations derives mainly from gradations in allele frequencies of subset of genes rather than from distinctive alleles present in specific populations. It is only through the accumulation of small allele-frequency differences across many loci that the genetic structure of a population, that is, the distinctive combination of allele frequencies, could emerge.

Several indices have been proposed to quantify the degree of genetic differentiation among two or more populations using series of gene frequencies. One such index is the F_{ST} distance, which measures the genetic variance between populations as a fraction of the total genetic variance. The F_{ST} index by Cavalli-Sforza is based on the frequency of 128 alleles related to 45 genes, and includes alleles coding for blood groups, lymphocyte antigens, immunoglobulins, hemoglobins, and enzymes. The choice of the genes sampled depended by 1) the desire to sample only genes which are clearly selectively neutral and by 2) the simplicity of collecting data. By construction, F_{ST} ranges between 0 and 1; the closer F_{ST} is to 1, the higher is the genetic distance between two populations. This index has shown a high degree of correlation with other measures of genetic distances and since the data provided by Cavalli-Sforza are expressed in F_{ST} this index will be used in this study.

Given the choice of genes and alleles used for determining F_{ST} , it comes at no surprise that the pattern of overall genetic variation among populations differs substantially from traditional racial divisions, as driven for example by the color of the skin (Figure 1). Indeed, recent reports have suggested how skin pigmentation correlates with polymorphisms affecting single genes (Lamason et al., 2005; Soejima et al., 2006), of course not included in the list of

genes used for calculating genetic distance. The absence of correlation between genetic distance and the color of the skin is particularly intriguing and would argue against a relationship between “cultural perception” and overall genetic features, as measured by Cavalli-Sforza et al., as well as by classical human population studies.

In contrast, genetic distance and geography are strongly correlated. Without using prior information about individual sampling locations, a clustering algorithm applied to multilocus genotypes from worldwide human populations produced genetic clusters largely coincident with major geographic regions (Rosenberg et al., 2005). For populations that are geographically close, genetic and geographic distances are often highly correlated, with genetic distance reaching an asymptote at about 1000-2600 miles on average (Figure 2).⁹ Moreover, small discontinuous jumps in genetic distance are present for most population pairs on opposite sides of geographic barriers (Rosenberg et al., 2005). This is also true for Europe, where sharp increases in genetic distance correspond to geographical impediments, including major mountains and seas (Barbujani and Sokal, 1990; see Figure 3).

In conclusion, the Cavalli-Sforza’s measure of genetic distance, which has been used in economics, is poorly correlated with some external traits, which determine social perception of “races”, including skin pigmentation and heights, while it is well correlated with geographical variables.

A. Genes, Culture, and Geography in Europe

While the correlation between genetic distance and geography is unambiguously accepted, the correlation between genetic distance and cultural variables is still under study. Even the relationship with the most obvious measure of culture, language, is subject of debate. While Cavalli-Sforza has convincingly argued that linguistic families are correlated with ancient migrations and genetic patterns (Cavalli-Sforza et al., 1994), other studies, one for Northern European populations (Zerjal et al. 2000) and one for the entire Europe (Rosser et al., 2000) show that populations in Europe are related mainly on the basis of geography and not on the

⁹ This issue is analyzed further in the empirical section, see in particular Table 1 for the geographic determinants of genetic distance.

basis of linguistic affinity.¹⁰ For example, Northern Europe shows linguistic and cultural diversity.¹¹ At the same time, the Scandinavian Peninsula is separated from Finland and the Baltic countries by the Baltic Sea. Zerjal et al. (2000), using Y-chromosomal data, conclude that the major genetic difference in Northern Europe is geographical, distinguishing populations living in the Western and Eastern side of the Baltic. Language plays a less but still important part in the determination of genetic differences (they found that Latvians showed greater genetic similarity to the Lithuanians than to the Estonians). Using Y-chromosome data and extending the sample to 47 European countries¹², Rosser et al. (2000) also find a strong and highly significant partial correlation between genetics and geography but a low and non-significant partial correlation between genetics and language.

Even if some correlation between language families and genetic distance is established, the application of this finding in economics is still problematic. For instance, if genetic distance is a proxy for cultural variables such as religion and languages, the inclusion of linguistic and/or religious differences (the proximate measure of cultural difference) in explaining economic outcomes should reduce greatly the magnitude and the significance of the coefficients on genetic distance in regressions with economic outcomes as dependent variables. However, Fearon (2006) has shown that this is not the case. In other words, genetic distance in economics seems to capture something which is not correlated with the usual variables capturing cultural factors, such as language and religion.¹³

In addition, the common belonging to the same historic linguistic family, which is correlated to genetic distance (Cavalli-Sforza, 1994), is not necessarily a good measure of mutual

¹⁰ As we previously said, differences between populations arise largely through random genetic drift when they are separated by distance, geographical barriers or culture. Europe has been considered an excellent area to study the importance of the different factors because its archeology, linguistics and genetics are fairly well known.

¹¹ In terms of languages: Swedes and Norwegians belong to the Germanic subfamily of the Indo-Europeans, Latvian and Lithuanians belong to the Baltic subfamily of the Indo-Europeans and Finns, Saami and Estonians belong to the Uralic family. Cultural differences are defined by the authors as livelihood differences: the Saami were hunters and nomadic, whereas the other populations were mainly farmers.

¹² The dataset is the biggest existing available on Y-chromosome diversity.

¹³ These findings are compatible with the dual inheritance theory developed by Cavalli-Sforza and Feldman (1980) according to which humans are products of interactions between natural selection and cultural evolution if the cultural traits that were important for cultural evolution are not important for contemporary economic outcomes.

comprehension or of cultural commonality. For instance, historical linguistics shows that Indians from New Delhi are linguistically closer to Icelanders than to Indians from Chennai but this does not suggest any strong cultural commonality between Icelanders and Northern Indians. Moreover, to an Italian-speaker Hungarian, Hindi, or Armenian are equally incomprehensible despite the fact that the Italian is much closer historically to Armenian and Hindi than Hungarian. On the other hand, for an English-speaker French could be more intelligible than German despite the fact that German and English are both Germanic languages. In other words, belonging to the same historical linguistic group, which is what genetic distance is correlated with, helps communication only in some specific cases and could not have direct economic effect.

In summary, the relationship between genetic distance and geography is uncontroversial; at the same time, there is evidence that genetic distance is correlated to some cultural traits, including linguistic historical relationship, while the relation of genetic distance to other cultural traits of economic interest is still being explored. For the main point of this paper, the simple correlation between genetic and cultural distances is not relevant. The relevant point is that the correlation between genetic distance and economic outcomes seems to be in large part explained by geography. We explore the relationship between genetic distance and geography in the next section.

III. GENETIC DISTANCE, GEOGRAPHY, AND TRADE

A. Data

The **index of genetic distance** F_{ST} is taken from Cavalli-Sforza (1994). We use only the values in which there is a correspondence between genetic population and country. We do not use data on sub-national population with distinct genetic information, including the Sardinians, Basques, and the Lapps.

The **bilateral export data** are obtained from the United Nations COMTRADE database revised by Feenstra et al. (2005). Our GDP data are obtained from the World Development Indicator of the World Bank; distance between capitals, common official language

and contiguity dummies are obtained from a new dataset compiled at CEPII.¹⁴ Both exports and GDP data refers to 2002. We also repeat the analysis using a panel approach for the period 1975-2000 and our results do not change.¹⁵

We use a newly constructed **measure of transportation costs**.¹⁶ This measure (tc_{ij}) is taken from shipping company quotes collected by Import Export Wizard (IEW), a shipping company providing estimates of transportation costs around the world. IEW calculates the surface freight data based on a survey of inter-modal 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 data refers to 2006. The advantage of this measure is that it represents the actual average transport costs and not indirect measures or proxies, which are often plagued by measurement errors.

We construct a set of **measures of geographical barriers** using information on sea, mountain chains, and the average elevation of countries. We define a variable (mountains) identifying the number of major mountain chains between countries. According to 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 UK), the Balkan Mountains, the Massif Central, the Meseta, the Pyrenees, the Urals, the Carpathian Mountains and the Caucasus. We define a dummy “common sea” equal to one if a pair of countries shares the same sea, which can be the Mediterranean, the Atlantic Ocean, or the Northern/Baltic Sea. Finally we construct a variable measuring the average elevation of countries which are between two trading partners. For instance, for the pair Germany-Italy this variable is equal to the average elevation of Germany, Austria, and Italy¹⁷. This variable measures the difficulty of transportation/migration between countries.

¹⁴ The data on bilateral exports can be found at <http://cid.econ.ucdavis.edu/>. The data on distance, language and contiguity is available at <http://www.cepii.fr/anglaisgraph/bdd/distances.htm>.

¹⁵ Results are reported in the appendix. We have decided to perform a cross-section analysis since our measure of transportation costs is available only for last year.

¹⁶ For a review of the literature on transport costs see Anderson and van Wincoop (2004).

¹⁷ An alternative would be the variance of the elevation of the countries. We prefer our measure because the cost of transportation is usually proportional to the average elevation.

The list of countries in our sample is described in Table A1. We cover 21 pairs of countries for Europe. Table A2 shows the correlation between genetic distance, transportation costs and several measure of geography. It is apparent from the table that genetic distance shows a very high correlation not only with the standard measure of geographic distance (the log distance between two countries) but also with the average elevation between countries and number of mountain chains between countries; the correlation is lower for the presence of a common sea. Table A3 shows descriptive statistics, while Table A4 describes the distribution of mountains and seas across countries.

Figure 4 shows the correlation between our measure of transportation costs and genetic distance. The correlation between the actual measure of transportation costs is striking, confirming our intuition that genetic distance in a gravity equation is probably picking up geographic impediments that are relevant in the determination of transportation costs.

B. Genetic Distance and Geography

This section analyzes the relationship between genetic distance and geography. As we saw in the previous session, geography (including not only the distance between countries, but also the presence of major mountains chains and common 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: i) the same factors that determined genetic distance in the past also have a strong influence in current transportation costs; ii) once we properly control for transportation costs, the impact of genetic distance on trade disappears, indicating that the correlation between trade and genetic distance may be spurious.

Our first goal is to show how geography has shaped genetic differences within Europe. Our starting point is Figure 3 (Sokal et al., 1990), which shows the main genetic ‘discrete’ changes within Europe. Sokal et al. (1990) have identified 33 boundaries of sharp changes in gene frequencies across Europe and have shown that the zones of abrupt genetic change in European populations correspond mostly to geographical boundaries. Specifically, the authors have counted 22 physical, 4 mountainous, and 18 marine boundaries. “In the 22 cases in which there are both physical barriers and genetic boundaries, it is reasonable to postulate that the causal arrow is likely to go more from physical barriers to both genetic and linguistic

differentiation, rather than in other directions” (Cavalli-Sforza, 1996, pag. 271).¹⁸ The importance of geography is also confirmed by classical genetic studies in humans and other organisms, also showing a strong association between geographic boundaries and genetic distance. Finally, note the ambiguous effect of sea. Ancient migrations often followed the sea coasts; therefore, sharing the same sea is a unifying factor. At the same time, crossing large seas was relatively complicated so islands are usually genetically isolated.

In order to investigate more systematically how geographical factors shape genetic distance, we run a regression with genetic distance between populations as dependent variable and several geographical variables as control variables. As discussed before, the measure of genetic distance is derived from Cavalli-Sforza et al. (with F_{ST} derived from the analysis of the allele frequencies of 88 genes). The choice of geographical variables, following Sokal et al. (1990), includes distance, number of mountains between countries, the presence of a common sea, and average terrain elevation between two countries (as defined below). In addition, all regressions have country fixed effects to control for country specific characteristics, including insularity, remoteness and presence of major mountain chains. The results presented in the first three columns of Table 1 use different combinations of geographical variables and confirm that geographical measures and genetic distance between European countries are indeed correlated.¹⁹

The regressions reported in Table 1 and the literature reviewed above show that geography (including the distance between countries, the presence of major mountains chains, and common 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. Given the strong correlation between geography and genetic distance, we hypothesize that geography affects both genetic distance and, via transport costs, trade flows and that the correlation between trade and genetic distance is largely spurious. In the next section, we

¹⁸ Note also that some genetic boundaries cut countries in the middle (for instance in Germany, Italy, Finland, Island, Spain, and Greece.) This observation reinforces the point that sharp genetic differences exist within national population, i.e. national borders are not always genetic borders. Moreover some abrupt changes follow linguistic and not geographical lines, for instance between Germany and the Netherlands. In other cases, there are linguistic barriers but no big genetic discontinuity (see for instance Germany and France, or Germany and Poland).

¹⁹ Note that the effect of insularity is captured by country fixed effects; as a consequence, sharing the same sea is always a strongly unifying factor. In a regression which includes island dummies among the regressors, the sign on island dummies is positive and significant consistent with Sokal’s findings.

show that: i) the same geographic factors that contribute to genetic distance are also important determinants of modern transportation costs; ii) in a standard gravity equation the impact of genetic distance on trade disappears once we introduce transport costs.

C. Determinants of Transportation Costs and Genetic Distance

In this section we analyze whether the measures of geographic distance we have constructed have an impact on transportation costs and genetic 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 transport costs (see Hummels, 1998, Limao and Venables, 2001, and Eaton and Kortum, 2002).²⁰ A substantial literature in engineering has long observed that topographical characteristics, such as terrain variability, affect transportation costs (see Tsunokawa, 1983, and World Bank, 1987). For example, for the same horizontal distance, moving goods across variable terrain requires more energy and time. Similarly, Limao and Venables (2001) and the World Bank (1998) find that non primary export high performers are island countries and none is landlocked.

To understand the geographic determinants of transport costs, we run the following regression:

$$\log(tc_{ij}) = \alpha_0 + \alpha_1 \log(D_{ij}) + \alpha_2 \text{geography}_{ij} + \alpha_3 \text{country}_i + \alpha_4 \text{country}_j + \varepsilon_{ijt}$$

where tc_{ij} are the transport costs, D_{ij} is the distance between country i and country j , geography is a vector with our proxies for geographical barriers (number of mountains, common sea, and average elevation of intermediate countries), country_i and country_j control for importer and trading countries fixed effects. The results of our regressions are reported in the last three columns of Table 1 for several combinations of geographical barriers. All variables have the expected sign. (Log)-distance, number of mountain chains, and average elevation between countries increase transportation costs, whereas the presence of a common sea reduces

²⁰ See Clark, Dollar, and Micco (2004) for a more nuanced measure of transport costs.

them.²¹ In conclusion, the same geographical determinants explain genetic distance and transport costs between countries.

D. Genes versus Geography in Explaining Trade

In this section, we show that, when we instrument transportation costs with those measures of geographical characteristics that influenced also migration patterns in the past (and hence largely genetic distance nowadays), the impact of genetic distance on economic outcome (in this case trade) disappears. In order to do so, we estimate the following gravity equation:

$$\ln(X_{ij}) = \beta_0 + \beta_1 gendist_{ij} + \beta_2 \ln(Y_i) + \beta_3 \ln(Y_j) - \beta_4 \ln(D_{ij}) + \beta_5 C_{ij} + \beta_6 L_{ij} + \beta_7 E_{ij} + \varepsilon_{ij}$$

where X_{ij} is the value of annual exports from country i to country j for the year 2002, $gendist_{ij}$ is the genetic distance between country i and j as defined by Cavalli-Sforza et al. (1994), Y_i is the real GDP of country i , D_{ij} is the distance between 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 i and j , E_{ij} is a dummy equal to one if country i and country j both use the Euro at time t , and ε_{ij} is the error term. We control for country of origin and country of destination fixed effects.

We begin by estimating a standard gravity equation. The first specification (Table 2, column 1) does not include genetic distance. Contiguity, distance, language, the presence of the Euro and GDP for both the importing and the exporting country have all the expected sign. As a next step, we introduce genetic distance among the regressors (specification 2). Genetic distance has a negative sign and is significant at the one percent level. From this regression, one would be inclined to infer that cultural distance has a negative impact on trade. Specification (2)

²¹ We used all the 1,332 observations for which data were available. If we restrict the same to the 470 observations for which genetic distance is available we obtain similar results. The only difference is column 6: with the smaller sample the three measures of geography appear to be significant once they are all included in the regression for the determinants of transportation costs.

could be misleading, since genetic distance could capture the effect of omitted variables (transport costs). We argue that genetic distance is related to trade because it is a proxy for geographic impediments, which increase transportation costs between countries.

To address this issue we introduce transportation costs. However, OLS regressions suffer from endogeneity bias because transport costs depend on the volume of trade (Hummels, 1999). For this reason, we instrument transport costs with a combination of our measures of geography. The IV regression (Columns 3) confirms that, when we introduce transportation costs, genetic distance is no longer significant, whereas the effect of transportation costs is negative and significant.²² Therefore, transportation costs appear to be crucial in the determination of trade, while genetic distance is once again not significant.

For all our specifications we run the appropriate tests to check that our instruments are weakly correlated with transportation costs (see Stock and Yogo, 2002, and Moreira, 2004). The F statistics for the joint hypothesis that the instruments' coefficients are zero in the first stage regression, has always a significance level lower than 5% for all specifications. In addition all the tables report the Hansen J statistics of over-identifying restrictions.

In conclusion, we find that genetic distance plays no role in explaining total bilateral trade flows once we control for transport costs.

IV. ALTERNATIVE EXPLANATIONS?

The findings of the section above show that *total* bilateral trade is not explained by genetic distance once we control for transport costs. However, this result could be driven by the fact that only a small part of bilateral trade flows is explained by cultural factor as proxied by genetic distance. To address this issue we classify the data using Rauch's classification to identify goods that should be more sensitive to cultural differences. Our findings could also be potentially plagued by problems of multicollinearity if the data of geography, cultural traits,

²² Table 2 reports a specification that uses average elevation and the presence of a common sea as instruments. We find very similar results using different combinations of instruments.

trade, and genetic distance are highly correlated (see Table A2 for a correlation table).²³ To address this issue, we propose a new classification of goods according to the easiness of transportability to identify goods for which transport costs may be more relevant.

A. Homogeneous vs. non-homogeneous goods

If genetic distance is a proxy for cultural differences, countries with less genetically similar populations should trade relatively less in goods for which the role of cultural similarities are more relevant. In order to test this hypothesis, we divide the goods into two categories: differentiated products and homogeneous goods following Rauch's classification (Rauch, 1999).²⁴ Table 3 reports the results for regressions using the sample restricted to these two types of goods. As before, we run a regression with genetic distance and a regression where transportation costs are instrumented with our measures of geography. Genetic distance is significant in the regressions without transport costs for both homogeneous and differentiated goods, however its significance disappears when we introduce transportation costs. In addition, consistent with our story, transportation costs are not relevant for differentiated goods, while they are relevant for the homogeneous ones, in line with the idea that homogenous goods are, on average, heavier and more costly to move than other goods (Rauch, 1999).

B. Bulky vs. non-Bulky goods

While the Rauch classification is based on the existence of organized trade and referenced prices, we also propose a new classification based on the easiness to transport.²⁵ Goods which are easy to transport because the values per weight is very high (for instance, diamonds, gold, and electronics) should not be sensitive to transport costs. On the opposite, bulky goods which are difficult to move should be more sensitive to transport costs. On the

²³ The correlation between genetic distance and cultural traits, as proxied by answers to the World Value Survey, has been shown by Desmet et al., 2006.

²⁴ Homogeneous goods are goods traded on organized exchanges, reference priced goods are goods not traded on organized exchanges but nevertheless possessing a reference price, differentiated goods are other commodities (Rauch, 1999). Over 60 percent of the value is in differentiated goods; about 20 percent belongs to reference priced goods, and the remaining to the organized trade.

²⁵ We thank David Hummels for useful discussion on this particular index and for providing us the data to construct it.

other hand, there is no reason as to why genetic distance should make easier trade in goods which are easy to move or bulky. We construct the index of bulkiness by looking at the freight to value ratio for the US imports from Mexico and Canada at 4 SITC digits.²⁶ We consider bulky all goods whose freight to value ratio is higher than the median, the other half of the goods is classified as easy to transport. Table 4 reports the results of our basic specification for easy-to-transport and bulky goods; as before, transport costs are instrumented for using our geographical variables. As expected, the coefficient on transport costs is large and highly significant for bulky goods while is not significant for easy-to-transport goods. Genetic distance is always insignificant when we control for transportation costs.

In conclusion, classifying trade between goods with organized markets or according to the easiness to transport allows us to perform a better test on the importance of geography versus cultural similarities in the determination of trade. Using both Rauch and our measure of bulkiness, we reach the same conclusion: transport costs are significant for bulky goods and/or homogeneous goods while genetic distance is never significant where we expected it to be significant, i.e. in goods in which cultural differences may be relevant. We take this as a strong indication that, even though genetic distance, geographical barriers, and cultural differences are correlated, trade is explained mostly by transport costs.

V. POSSIBLE USES OF GENETIC DISTANCE IN ECONOMICS

In this section we discuss possible uses of genetic distance in economics. From the previous discussion it appears that the first obvious use of genetic distance is as instrument for transportation costs. Being exogenously determined by geographical barriers, genetic distance could be simply used as an indirect measure for geographical impediments. In Table 5, we run a gravity equation using genetic distance as an instrument for transportation costs. As expected the coefficient on transportation costs is negative and significant.²⁷ The first stage of this

²⁶ We use US data because they give detailed information on freight rate and values at 4 digit level. We choose import from Mexico and Canada because, being contiguous countries, all modes of transportation are used, including sea, land, and air. We make the assumption that the ranking of this ratio is the same in North America as in Europe.

²⁷ As before, Column 1 controls for country of origin and country of destination fixed effects, whereas Column 2 adds country specific linear trends.

regression is reported in Table A6. As before, in the IV specification, not only the effect of transportation costs remains significant, but its magnitude increases substantially.

Besides being a proxy for geographical impediments, being a heritage of the past, a population's genetic make-up can also provide valuable insights about the history of a country. Inasmuch ancient history is believed to be an important determinant of current economic outcomes, genetics can add useful information. For instance, the genetic composition of a country, which was formerly a colony, can inform us on how much intermingling there was between colonizers and natives. Little intermingling will suggest that the colonizers' culture was simply imposed while more intermingling will indicate that population transfers also occurred.²⁸ Even more specifically, analysis of mitochondrial DNA and chromosome Y analysis can unveil the type of intermingling – if just male colonizers contributed or if whole nuclear families moved.

The genetic make-up of a population could also provide information on past environmental conditions, including the endemic diffusion of diseases such as malaria or the dietary habits, which shaped and acted upon specific phenotypes. Note that in this case, the measure of genetic distance proposed by Cavalli-Sforza, which is intentionally based on genes that likely do not affect phenotypes subject to natural selection, is not appropriate. Some of these applications are left for future research.

VI. CONCLUSIONS

Our paper is motivated by the surprising finding that genetic distance and trade flows are correlated. We have presented original evidence of this, in addition to what was uncovered in other papers. Our main contribution is to move a step forward in understanding the reason behind this correlation.

²⁸ We credit Simon Johnson for this specific idea. Note that the idea that genetic analysis is a good instrument to understand the pattern of cultural spreading was the original motivation of Cavalli-Sforza et al. (1994). During the Neolithic revolutions in Europe new artifacts or techniques (or cultures) appeared to spread westwards. However, it is unclear if this spreading of culture was brought about by a spread of population or not. Cavalli—Sforza has argued that genetic analysis is compatible with the hypothesis of migration. Similarly, the transfer of techniques from colonizers to colonies may have happened with or without population intermingling and the subsequent economic outcome and institutions may depend on the way techniques have spread.

Genetic distance is strongly correlated with geographical variables. Our starting point is the observation that the same geographical barriers which shaped genetic patterns play a role in contemporary transport costs. We confirm this intuition with simple regressions. We have taken the case of European trade for which data on genetic distance are uncontroversial and there is considerable (but not total) overlapping between genetic population and countries. If geographical barriers play a role for both genetic distance and modern day transport costs, the variable genetic distance in a trade equation simply captures an omitted variable. We show the point by presenting gravity equations with genetic distance and with or without transport costs. As expected, the coefficient on genetic distance becomes insignificant once transport costs are introduced.

We check the robustness of our results by repeating our regression restricting the sample to goods with an organized market (using Rauch classification) and for easiness to transport (using a new classification). The results are confirmed: bulky and homogeneous goods are particularly sensitive to transport costs while genetic distance does not play any role even in differentiated goods.

Our main conclusion is that genetic distance loses any significance in explaining trade once one controls properly for geography, therefore genetic distance can be used as a proxy for cultural distance once geography is appropriately controlled for. On the other hand, genetic distance could serve as an exogenous measure of transportation costs.

Looking forward, our paper has built a solid ground on which future research on genetic distance and economics can proceed. Future research should recognize that genetic distances between populations largely reflect geographical impediments. However, once geography is appropriately taken into account, genetic distance may reflect other factors of possible interest to economists, including cultural differences, population intermingling, and ancient migrations.

References

- Acemoglu, Daron, Simon Johnson and James A. Robinson, 2001, “The Colonial Origins of Comparative Development: and Empirical Investigation,” *American Economic Review*, Vol. 91 No. 5 (December).
- Alesina, A., W. Easterly and J. Matuszeski (2006), “Artificial States”, NBER WP 12328
- Akey, J. M., K. Zhang, L. Jin, and M.D. Shriver, 2002, “Interrogating a High-Density SNP Map for Signatures of Natural Selection,” *Genome Research*, Vol. 12, pp. 1805-14.
- Anderson, J. and E. van Wincoop, 2003, “Gravity with Gravitas: a Solution to the Border Puzzle,” *American Economic Review*, Vol. 93 No. 1, pp. 170-92.
- , 2004, “Trade Costs,” *Journal of Economic Literature*, Vol. 42 No. 3, pp. 691-751.
- Carlson, C.S., Thomas, D.J., Eberle, M.A., Swanson, J.E., Livingston, R.J., Rieder, M.J., Nickerson D.A. 2005. Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res.* **15**: 1553–1565
- Cavalli-Sforza, L. L., and M. W. Feldman, 1981, *Cultural Transmission and Evolution* Princeton: Princeton University Press.
- Cavalli-Sforza, L. L., and M. W. Feldman, 2003, “The application of Molecular Genetic Approaches to the Study of Human Evolution,” *Nature Genetic Supplement*, Vol. 33 pp. 266-75.
- Cavalli-Sforza, L., P.Menozzi, and A. Piazza, 1994, *The History and Geography of Human Genes*, (Princeton, New Jersey: Princeton University Press).
- Desmet, K., M. Le Breton, I. Ortuño-Ortín, and S. Weber, 2006 “Nation Formation and Genetic Distance” CEPR Discussion Paper No. 5918.
- Eaton, J. and S. Kortum, 2002, “Technology, Geography and Trade,” *Econometrica*, Vol. 70 No. 5, pp. 1741-79.
- Clark, X., D. Dollar, and A. Micco, 2004, “Port Efficiency, Maritime Transport Costs, and Bilateral Trade,” *Journal of Development Economics*, pp. 417-50.
- Fearon, James D., 2006, “Is Genetic Distance a Plausible Measure of Cultural Distance.” Mimeo. Stanford University.
- Feenstra, R., R. Lipsey, H. Deng, A. Ma, H. Mo, 2005, “World Trade Flows: 1960-2000,” NBER Working Paper 11040, (Cambridge, Massachusetts: National Bureau of Economic Research).

- Gallup, John Luke, and Jeffrey Sachs D., with Andrew D. Mellinger, 1998, “Geography and Economic Development,” in *Annual World Bank Conference on Development Economics*, (Washington: The World Bank).
- Goldstein, D. B., and L. Chikhi, 2002, Human Migrations and Population Structure: What We Know and Why it Matters,” *Annual Reviews Genomics Hum Genet*, Vol. 3 pp. 129-52.
- Guiso, L., Sapienza, P. and L. Zingales, 2005, “Cultural Biases in Economic Exchange,” NBER Working Paper No. 11005, (Cambridge, Massachusetts: National Bureau of Economic Research).
- , 2006, “Does Culture Affect Economic Outcomes?” *Journal of Economic Perspectives*, Vol. 20, No. 2, pp. 23-48.
- Hofstede, G., 1980, *Culture’s Consequences: International Differences in Work-Related Values*, (Beverly Hills: Sage Publication).
- Hofstede, G., and R. McCrae, 2004, “Personality and Culture Revised: Linking Traits and Dimension of Culture,” *Sage Publications-Cross Cultural Research*
- Hummels, David, 1998, “Towards a Geography of Transport Costs,” Mimeo, (Chicago: University of Chicago, Department of Economics).
- Hummels, David, and V. Lugovskyy, 2006, “Usable Data? Matched Partner Trade Statistics as a Measure of Transportation Cost,” Forthcoming, *Review of International Economics*.
- The International HapMap Consortium. 2005. “A haplotype map of the human genome” *Nature* 437: 1299–1320.
- Joanna L. Kelley, Jennifer Madeoy, John C. Calhoun, Willie Swanson and Joshua M. Akey, 2006. “Genomic signatures of positive selection in humans and the limits of outlier approaches” *Genome Res.* 16: 980-989.
- Kimura, M., 1968, “Evolutionary Rate at the Molecular Level,” *Nature*, No. 217, pp. 624-26.
- Lamason, R. L., Mohideen, M. A., Mest, J. R., *et al.*, 2005, “A Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans Science,” *SLC24A5* No. 310, pp. 1782-86.
- Lewontin, R.C., Rose L., and Kamin L., 1984, *Not in Our Genes: Biology, Ideology and Human Nature*, (New York: Pantheon Books).
- Limao, N. and Venables, A., 2001, “Infrastructure, Geographical Disadvantage, Transport Costs and Trade,” *World Bank Economic Review*, Vol. 15, pp. 451-79.

- McCrae, R. and J. Allik, 2002, *The Five-Factor Model of Personality Across Cultures*, (New York: Kluwer Academic/Plenum Publisher).
- Putnam, R., 1993, *Making Democracy Work*, (Princeton, New Jersey: Princeton University Press).
- Rauch, James, 1999, "Networks Versus Markets in International Trade," *Journal of International Economics*, Vol. 48 (June 1999). Pp. 7-35.
- Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K., and M. W. Feldman, 2005, "Clines, Clusters, and the Effect of Study Design on the Inference of Human Population Structure," *PLoS Genet* No. 1, p. 70.
- Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi, 2004, "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development", *Journal of Economic Growth*, Vol. 9, no.2.
- Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd, L. A. Zhivotovsky, and M. W. Feldman, 2002, "Genetic Structure of Human Populations," *Science* No. 298, pp. 2381-85.
- Sachs, Jeffrey, 2003, "Institutions Don't Rule: Direct Effects of Geography on Per Capita Income," [NBER Working Paper 9490](#).
- Soejima, M., H. Tachida, T. Ishida, A. Sano, and Y. Koda, 2006, "Evidence for Recent Positive Selection at the Human AIM1 Locus in a European Population," *Mol Biol Evol*, No. 23, pp. 179-88.
- Spolaore, E. and R. Wacziarg, 2005, "The Diffusion of Development," mimeo
- Terracciano, A., et al., 2005, "National Character Does Not reflect Mean Personality Trait Levels in 49 Countries," *Science*, Vol. 310, (October), pp. 96-100.
- Voight, B.F., Kudaravalli S., Wen X., Pritchard J.K. 2006. "A map of recent positive selection in the human genome". *PLoS Biol*. 4: e72.
- Wang E.T., Kodama G., Baldi P., Moyzis R.K. 2006. Global landscape of recent inferred Darwinian selection for *Homo sapiens*. *Proc. Natl. Acad. Sci.* **103**: 135–140.
- Weber, M., 1958, *The Protestant Ethic and the Spirit of Capitalism*, Translated by T. Parsons, (New York: Scribner's Library).
- Zerjal, T., et al., 2001, "Geographical, Linguistic and Cultural Influences on Genetic Diversity: Y-Chromosomal Distribution in Northern European Populations," *Molecular Biology and Evolution*, Vol. 18 No. 6, pp. 1077-87

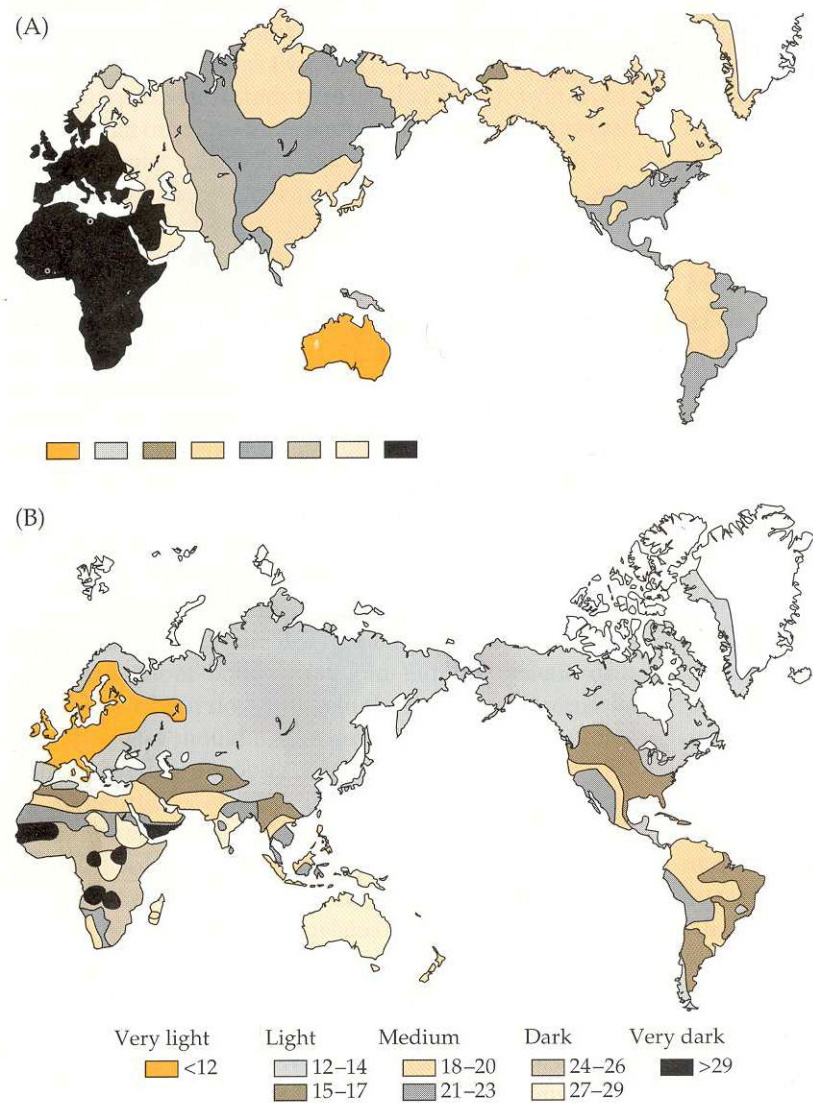
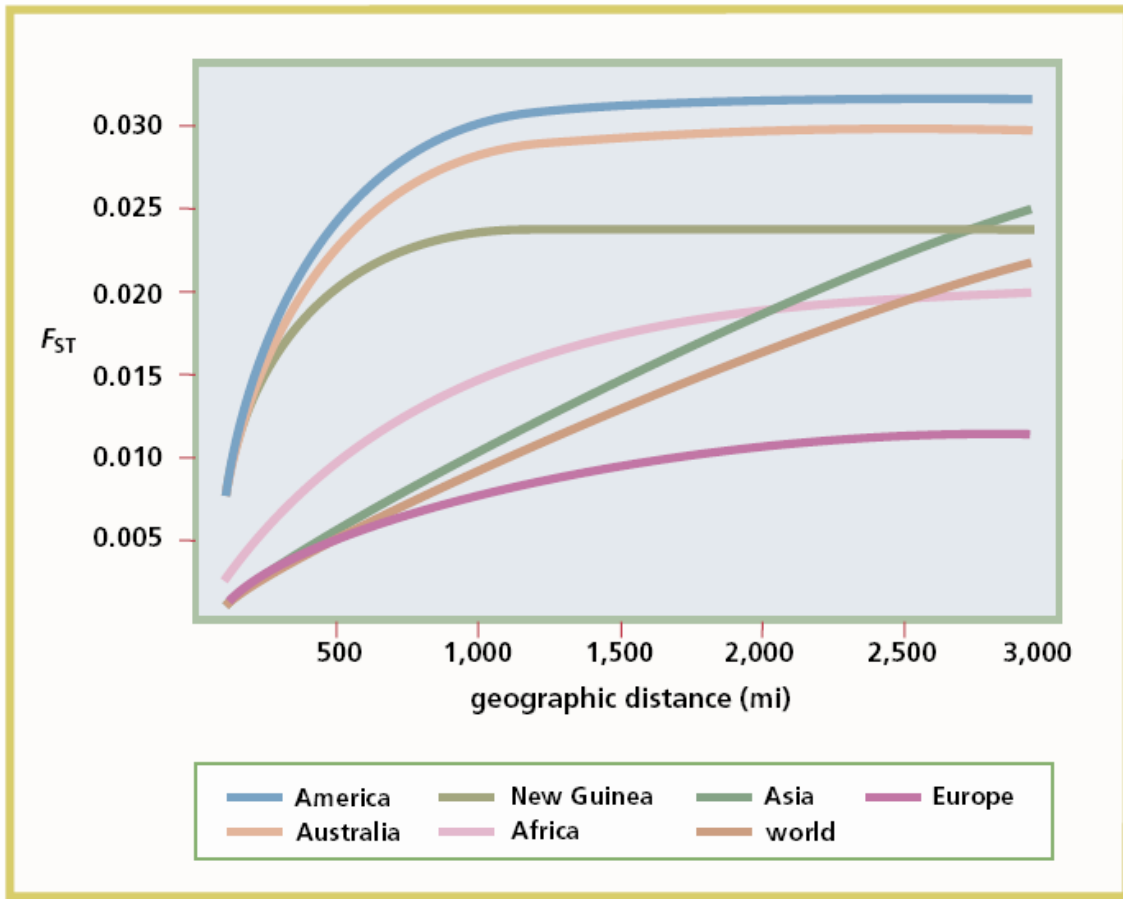


Figure 1. A. Division of the world's human population into eight classes of genetic similarity, based on overall difference and similarity at numerous enzyme and blood-group loci. The eight classes represented are arrayed in order of increasing difference. B. Geographic distribution of skin color, classified in eight grades of pigmentation intensity (adapted from Cavalli Sforza et al., 1994)



BOB CRIMI

Figure 2

Relationship between Genetic and Geographic Distance (Cavalli-Sforza and Feldman, 2003)

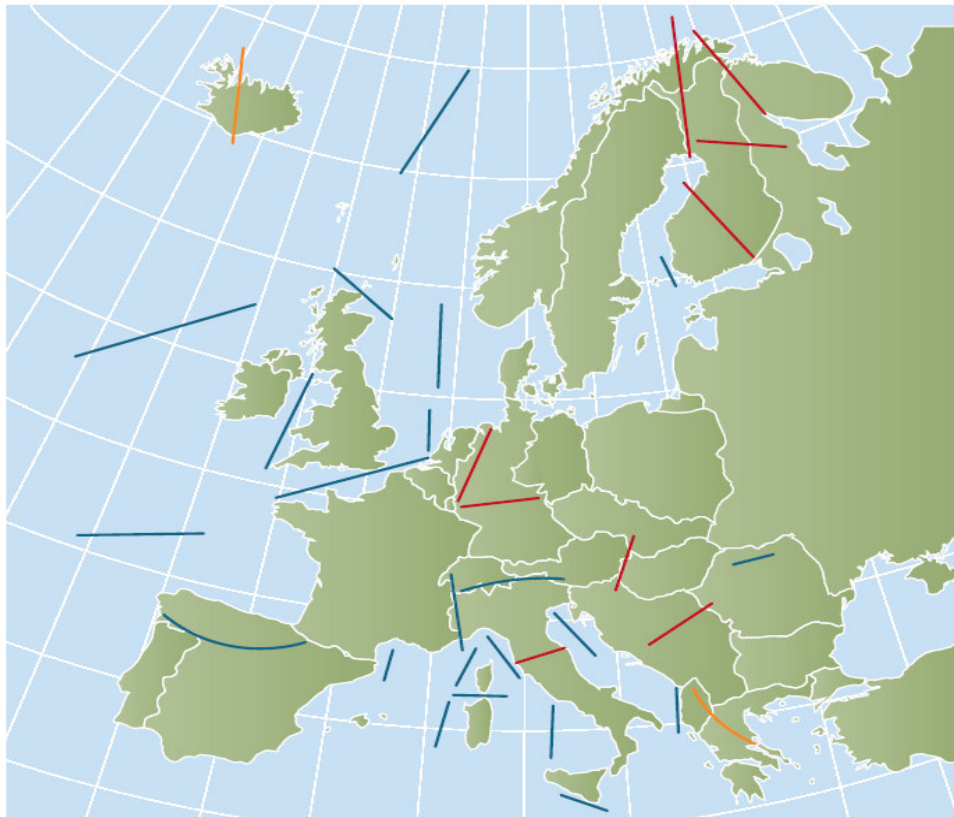


Figure 3

Zones of sharp genetic changes in Europe (Sokal, 1990)

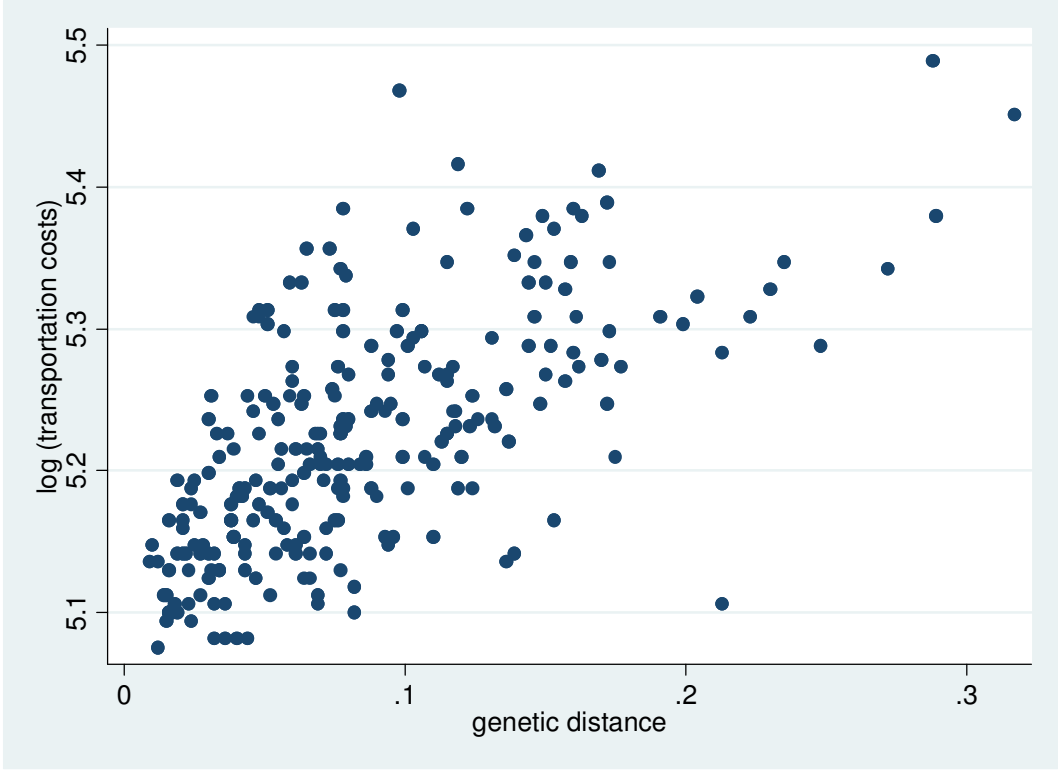


Figure 4.
Genetic Distance and Transportation Costs

Table 1.
Geographic Determinants of Genetic Distance and Transportation Costs

	Genetic Distance	Genetic Distance	Genetic Distance	ln(tc_{ij}) Transp. cost	ln(tc_{ij}) Transp. cost	ln(tc_{ij}) Transp. cost
	(1)	(2)	(3)	(4)	(5)	(6)
Log (distance)	0.022*** (.0031)	.0173*** (.0037)	.0173*** (.0037)	.0871*** (.0036)	.0880*** (.0067)	.0856*** (.0039)
Number of mountain Chains	.0010 (.0024)		-.0008 (.0025)	.0079*** (.0013)		.0069*** (.0014)
Average elevation between countries		.0414** (.0214)	.0441** (.0225)		.0381*** (.0133)	.0216 (.0139)
Common sea	-.0280*** (6.723)	-.0238*** (.0075)	-.0237*** (.0075)	-.0338*** (.0013)	-.0336*** (.0042)	-.0322*** (.0041)
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country of dest. fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	470	470	470	1332	1332	1332
R-squared	0.95	0.95	0.95	0.93	0.93	0.93

* significant at 10%; ** significant at 5%, *** significant at 1%. Robust standard errors in parentheses

Table 2
Genetic Distance, Transportation Costs and Trade
Dependent Variable: Log Total Bilateral Exports

	(1)	(2)	(3)
	<i>OLS</i>	<i>OLS</i>	<i>IV</i>
Transportation costs			-10.862 (1.79)*
Genetic distance		-5.485 (4.21)***	-1.073 (0.37)
Common language	0.434 (1.92)*	0.378 (1.87)*	0.603 (3.14)***
Euro	0.188 (1.68)*	0.158 (1.44)	0.156 (1.49)
Log (distance)	-0.940 (10.45)***	-0.764 (8.16)***	0.259 (0.46)
Contiguity	0.320 (2.64)***	0.368 (3.13)***	0.394 (3.29)***
GDP importer	0.970 (20.26)***	0.878 (16.97)***	0.923 (17.88)***
GDP exporter	1.085 (23.48)***	0.991 (20.68)***	1.038 (19.49)***
Observations	470	470	470
R-squared	0.94	0.94	.99
Hansen J statistic			.32

Robust standard errors in parentheses. significant at 10, %, ** significant at 5%, *** significant at 1%.
Instruments include average elevation and common sea.

Table 3. Genetic Distance, Transportation Costs and Trade
Dependent Variable: Log Bilateral Exports for Differentiated and Homogeneous Goods

	Differentiated goods		Homogeneous goods	
	<i>OLS</i>	<i>IV</i>	<i>OLS</i>	<i>IV</i>
Transportation costs		-3.816 (0.66)		-35.238 (2.45)**
Genetic distance	-4.090 (2.91)***	-2.540 (0.94)	-8.096 (3.20)***	6.000 (0.86)
Language	0.320 (1.58)	0.399 (1.99)**	0.654 (2.26)**	1.384 (3.05)***
Euro	0.248 (2.07)**	0.247 (2.23)**	0.117 (0.57)	0.101 (0.43)
Log (distance)	-0.671 (6.81)***	-0.311 (0.58)	-1.387 (8.83)***	1.932 (1.42)
Contiguity	0.441 (3.55)***	0.450 (3.85)***	0.340 (1.54)	0.417 (1.46)
Total GDP importer	0.904 (17.10)***	0.919 (17.66)***	0.740 (8.91)***	0.886 (8.28)***
Total GDP exporter	1.124 (18.73)***	1.141 (18.73)***	1.019 (10.13)***	1.170 (8.99)***
Observations	470	470	466	466
R-squared	0.94		0.86	
Hansen J Statistics		.71		0.05

Robust standard errors in parentheses. significant at 10, %, ** significant at 5%, *** significant at 1%. Instruments include average elevation and common sea.

Table 4
Genetic Distance, Transportation Costs and Trade for “Easy to Transport” and “Bulky” Goods
Dependent Variable: Log Bilateral Exports

	Easy to transport goods		Bulky goods	
	(1)	(2)	(1)	(2)
Transportation costs		-4.321 (0.57)		-14.481 (2.24)**
Genetic distance	-3.549 (2.00)**	-1.823 (0.53)	-6.160 (4.56)***	-0.278 (0.09)
Language	0.289 (1.24)	0.378 (1.53)	0.465 (2.28)**	0.765 (3.77)***
Euro	0.234 (1.44)	0.231 (1.53)	0.136 (1.19)	0.133 (1.17)
Log distance	-0.531 (4.58)***	-0.123 (0.18)	-0.893 (9.01)***	0.471 (0.79)
Contiguity	0.386 (2.46)**	0.395 (2.69)***	0.364 (2.94)***	0.399 (2.95)***
Total GDP importer	1.285 (13.44)***	1.302 (14.00)***	0.904 (19.30)***	0.963 (17.62)***
Total GDP exporter	0.988 (16.15)***	1.007 (15.92)***	0.835 (14.87)***	0.897 (15.84)***
Observations	469	469	470	470
R-squared	0.90	.99	0.93	.99
Hansen J statistics		.96		.13

Robust standard errors in parentheses. significant at 10, %, ** significant at 5%, *** significant at 1%.
Instruments include average elevation and common sea.

Table 5
Genetic Distance, Transportation Costs and Trade
Dependent Variable: Log Total Exports
Instrumenting Transportation Costs with Genetic Distance

	IV
Transportation costs	-13.503 (4.20)***
Language	0.658 (3.89)***
Euro	0.155 (1.44)
Log (distance)	0.508 (1.50)
Contiguity	0.400 (3.16)***
Total GDP importer	0.933 (23.31)***
Total GDP exporter	1.049 (25.15)***
Observations	470

Robust standard errors in parenthesis.

* significant at 10%; ** significant at 5%, *** significant at 1%.

Table A1
Countries included in our sample

Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Macedonia, Netherlands, Norway, Poland, Russia, Spain, Sweden, Switzerland, Turkey, UK.

Table A2
Correlations between Genetic Distance, Several Measures of Geography and Transportation Costs

	Genetic distance	Log(transp. costs)	Log(geogr. distance)	Mountains	Common sea	Average elevation
Genetic distance	1					
Log (trans. costs)	0.6505**	1				
Log (geogr. distance)	0.5839**	0.9117**	1			
Mountains	0.3405**	0.3143**	0.3371**	1		
Common sea	-0.0937*	0.0895*	0.0808*	-0.2914**	1	
Average elevation	0.3078**	0.4351**	0.4923**	0.5612**	-0.3852**	1

** , significant at the 1%, * significant at the 5%

Table A3
Summary Statistics

	Mean	Std. Dev.	Observations
Log Distance	7.134	.6365	1480
Contiguity	.0959	.2946	1480
Common Language	.0135	.1154	1480
Genetic Distance	99.22	115.24	470
Islands	.1277	.3338	1480
Landlocked	.1540	.3611	1480
Euro	.0235	.1517	26460
Number of mountains	.9202	.9420	1480
Average elevation b/w countries	159.08	122.04	1474
Common sea	.3594	.4800	1480
Log(GDP) Exporter	25.22	1.85	22280
Log(GDP) Importer	25.22	1.85	22280
Log(Total Exports)	11.29	2.74	25872
Log Transp. Costs (direct transport costs)	5.27	.1603	1332

Table A4
Distribution of shared mountains and seas

	Sea	Freq.	Percent
	0	262	56.71
	1	200	43.29
Mountains			
	0	226	48.92
	1	156	33.77
	2	52	11.26
	3	24	5.19
	4	4	0.87

Table A5
IV Regression, First Stage
Geographical Barriers as a Measure for Transportation Costs

	Log (Transp. Costs)
Average Elev. b/w Countries	.0564*** (.0179)
Common sea	-.0036 (.0050)
Common language	.0189** (.0090)
Euro	.0006 (.0045)
Genetic Distance	.3656*** (.0455)
Log(distance)	.0869*** (.0041)
Contiguity	.0040 (.0052)
GDP exporter	.0048 (.0015)
GDP importer	.0046 (.0015)
Observations	470
R-squared	.92

Robust standard errors in parentheses.

* significant at 10%; ** significant at 5%, *** significant at 1%.

Table A6
IV Regression, First Stage
Genetic Distance as an Instrument for Transportation Costs

Genetic Distance	.4532*** (.0478)
Common Language	.0207** (.0106)
Euro	-.0002 (.0046)
Log(Distance)	.0941*** (.0046)
Contiguity	.0023 (.0061)
GDP exporter	.0042*** (.0014)
GDP importer	.0041*** (.0013)
Observations	470
R-squared	.92

Robust standard errors in parentheses

* significant at 10%; ** significant at 5%, *** significant at 1%.

Table A7
Genetic Distance, Transportation Costs and Trade
Dependent Variable: Log Total Bilateral Exports

	<i>OLS</i>	<i>OLS</i>	<i>IV</i>	<i>IV</i>
	(1)	(2)	(3)	(4)
Transportation costs			-28.62** (11.47)	-27.92** (11.22)
Genetic Distance		-3.135** (1.537)	10.02* (5.416)	10.08* (5.321)
Common Language	.7031*** (.1776)	.6708*** (.1652)	1.169*** (.3118)	1.152*** (.3038)
Euro	.0921 (.0666)	.0800 (.0654)	.0880 (.0710)	.0254 (.0752)
Log (distance)	-.9623*** (.0881)	-.8531*** (.0941)	1.684* (1.005)	1.626* (.9789)
Contiguity	.1745 (.1153)	.2026* (.1157)	.3380* (.1845)	.3516** (.1785)
GDP exporter	.6959*** (.2307)	.7037*** (.2289)	.8363*** (.2427)	.4468* (.2357)
GDP importer	-1.090** (.5181)	-1.080** (.5155)	-.9225* (.5240)	.4109 (.3612)
Fixed year effects	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes
Country spec. linear trend	No	No	No	Yes
Observations	9408	9408	9408	9408
Hansen J statistic			.6774	.4847
R-Squared	0.85	.85	.99	.99

Significant at 10, %, ** significant at 5%, *** significant at 1%. Errors are clustered at the bilateral pair level. Instruments include average elevation and common sea.

Table A8
Genetic Distance, Transportation Costs and Trade
Dependent Variable: Log Total Exports

	Differentiated goods		Homogeneous goods	
	<i>IV</i> (1)	<i>IV</i> (2)	<i>IV</i> (1)	<i>IV</i> (2)
Transp. costs	-10.52 (6.841)	-9.562 (6.641)	-38.876** (16.31)	-39.70** (16.56)
Genetic Distance	1.816 (3.360)	1.804 (3.288)	14.379** (7.449)	14.86** (7.569)
Common Language	.8359*** (.1795)	.8111*** (.1747)	1.3916*** (.4288)	1.400*** (.4348)
Euro	.0771 (.0702)	.0445 (.0622)	.1335 (.0940)	-.0543 (.1122)
Log (distance)	.2314 (.5907)	.1525 (.5713)	2.150 (1.405)	2.236 (1.423)
Contiguity	.3730*** (.1290)	.3869*** (.1239)	.4139* (.2487)	.4278* (.2509)
GDP exporter	.8268*** (.2617)	.3443 (.2405)	.5823* (.3315)	.4354 (.3612)
GDP importer	-.9017* (.5026)	.7244* (.3904)	.5749 (.3738)	-.5683 (.4195)
Fixed year effects	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes
Country specific time effects	No	Yes	No	No
Observations	9363	9363	8981	8981
Hansen J Statistics	.3879	.1767	.1150	.1022
R-Squared	.99	.99	.99	.99

* significant at 10%; ** significant at 5%, *** significant at 1%. Standard errors are clustered at the bilateral pair level.

Table A9
Genetic Distance, Transportation Costs and Trade for “Easy to Transport” and “Bulky” Goods
Dependent Variable: Log Total Exports

	Easy to transport goods		Homogeneous goods	
	<i>IV</i> (2)	<i>IV</i> (3)	<i>IV</i> (2)	<i>IV</i> (3)
Transportation costs	-7.133 (7.798)	-6.758 (7.654)	-34.45*** (13.42)	-34.02*** (13.26)
Genetic Distance	.9508 (3.716)	.9975 (3.662)	12.11* (6.198)	12.32** (6.151)
Common Language	.6554*** (.2010)	.6470*** (.1973)	1.346*** (.3820)	1.33*** (.3764)
Euro	.1151 (.0777)	.0788 (.0786)	.0949 (.0700)	.0002 (.0834)
Log (distance)	-.1243 (.6852)	-.1460 (.6723)	2.189* (1.186)	2.154* (1.167)
Contiguity	.1996 (.1347)	.2087 (.1300)	.3988* (.2130)	.4156** (.2084)
GDP exporter	.1651 (.3315)	.6957* (.3733)	-.6622 (.4582)	.2412 (.3103)
GDP importer	.9580*** (.3107)	.0947 (.2904)	.9115*** (.2460)	.4959 (.2642)*
Fixed year effects	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes
Country spec. linear trend	No	Yes	No	Yes
Observations	9290	9290	9498	9498
Hansen J stastistic	.59	.43	.61	.47
R-Squared	.99	.99	.99	.99

significant at 10, %, ** significant at 5%, *** significant at 1%. Errors are clustered at the bilateral pair level. Instruments include average elevation and common sea.

Table A10
Genetic Distance, Transportation Costs and Trade
Dependent Variable: Log Total Exports
Instrumenting Transportation Costs with Genetic Distance

	<i>IV</i> (1)	<i>IV</i> (2)
Transportation costs (direct transport cost)	-7.595** (3.265)	-6.737** (3.150)
Common Language	.7833*** (.1362)	.7622*** (.1380)
Euro	.0361 (.0664)	.0223 (.0659)
Log (distance)	-.1730 (.3263)	-.2397 (.1096)
Contiguity	.2437** (.1147)	.2593** (.1096)
GDP exporter	.8076*** (.2132)	.6507*** (.1989)
GDP importer	-.7772* (.4565)	.3753 (.3288)
Fixed year effects	Yes	Yes
Country dummies	Yes	Yes
Country spec. linear trend	No	Yes
Observations	10348	10348
R-Squared	.99	.99

* significant at 10%; ** significant at 5%, *** significant at 1%.

Standard errors are clustered at the bilateral pair level.