Human Barriers to International Trade

Irene Fensore   Stefan Legge   Lukas Schmid*

December 31, 2016

Abstract

This paper investigates the impact of human barriers on international trade, using ancestral distance as a measure of the relatedness of populations. In a new data set covering the universe of global trade, our analysis documents that country pairs with a large ancestral distance are less likely to trade with each other (extensive margin) and if they do trade, we find that a reduced volume and number of goods traded (intensive margin). These results are robust to including a vast array of micro-geographic controls variables. We provide evidence suggesting that the inverse relationship between bilateral trade flows and ancestral distance arises from differences in values, preferences, technology, as well as network effects.

---

* Fensore and Legge: University of St.Gallen, Department of Economics, SIAW-HSG, Bodanstrasse 8, CH-9000 St.Gallen, Switzerland. E-mail: irene.fensore@unisg.ch, stefan.legge@unisg.ch; Schmid: University of Lucerne, Faculty of Economics and Management, P.O. Box 4466, CH-6002 Lucerne, Switzerland. E-mail: lukasdavid.schmid@unilu.ch

1   Introduction

Despite substantial reductions in transportation costs and significant decreases in tariffs and non-tariff trade barriers, geographic distance continues to appear as a major obstacle to trade. A large body of research provides evidence of a negative estimated impact of geographic distance in a gravity-type regression. This estimate has been persistently large since the middle of the 20th century (Disdier and Head, 2008; Yotov, 2012). Thus far, however, the literature has provided an incomplete understanding of factors behind the negative relationship between geographic distance and bilateral trade flows.

This paper offers a new explanation for the unresolved 'distance puzzle'. We argue that human differences which increase with geographic distance reflect a significant barrier to trade. In particular, we show that populations with a more recent ancestral past are more likely to have similar preferences, values, norms, habits, and technologies. Using ancestral distance as a measure of the relatedness of populations, we explicitly take into account migration patterns of the past. Populations with more recent common ancestors are likely to adapt technological innovations from each other (Spolaore and Wacziarg, 2013a). This facilitates both directly and indirectly the establishment of trade relationships. Furthermore, similar norms and values increase mutual trust that serves as a complement for incomplete contracts (Guiso, Sapienza and Zingales, 2009). In addition, countries with similar demand structures are more likely to trade with each other (Linder, 1961; Economides, 1984). Finally, the spread of technological innovations leads to higher growth rates in similar countries (Spolaore and Wacziarg, 2013a) and countries with similar income levels tend to trade more with each other (Anderson and Van Wincoop, 2003).

Evidence provided by Spolaore and Wacziarg (2013b) suggests that economic development is affected by human traits that have been transmitted across generations over the very long run. This paper examines how such traits affect current bilateral trade flows, holding fixed geographic barriers. We develop and empirically test the hypothesis that international trade depends on measures of the long-term historical relatedness between populations. For example, the United Kingdom trades more with Australia than with the Philippines although the geographic distance is roughly the same. The key to understanding this empirical fact is to realize that

the populations of the UK and Australia are historically much closer; their ancestral distance is significantly smaller. Hence, they share many of the same values, preferences, technologies, traits, and habits. Using data on the near-universe of global trade, for the first time we study how the relatedness of populations affects global bilateral trade flows.

We combine the most recent data set on ancestral distance (Spolaore and Wacziarg, 2016) with detailed trade data as well as a large set of bilateral geographic measures for 172 countries. To capture ancestral distance, we use genetic distance, a measure that describes the time elapsed since two populations' last common ancestors. Ancestral distance is positively but not perfectly correlated with geographic distance and thus constitutes an important factor through which geographic distance reduces trade.[1] Using a standard gravity equation framework, we find that a larger ancestral distance between two countries' populations reduces the probability that a trade relationship exists (extensive margin). Furthermore, the results indicate that both the volume and the number of commodities traded are lower among country pairs with a larger ancestral distance (intensive margin). We perform several tests to explore the robustness of our results. First, we include of a vast array of measures for micro-geographical distance including contiguities, access to the same sea, latitude, longitude, terrain ruggedness, shares of fertile soil, desert, and tropical climate, as well as the average distance to the nearest ice-free coast. Second, we control for a set of political variables including corruption, civil and political liberty, as well as free trade agreements. None of these variables alter the finding that bilateral trade flows are negatively associated with ancestral distance. Finally, our findings are also robust to including measures of linguistic and religious distance.

Having established a negative relationship between ancestral distance and trade, this paper also sheds light on proximate determinants of trade which create this relationship. The estimated effect of ancestral distance can reflect differences in technology, preferences, values, as well as an obstacle to communication, social interaction, and learning across different societies. In a first step, we find that countries with a larger ancestral distance differ in the technologies they use. This in turn negatively affects bilateral trade flows. Second, using 4-digit trade data we find that ancestral distance is positively associated with a differences in preferences. Coun-

---

[1]Note that, for example, Chile has about the same geographical distance from the United States as Argentina. The genetic distance of Chile, however, is about twice the genetic distance of Argentina.

tries whose population share more distant common ancestors import and export (i.e. consume) a different basket of goods which partly explains why they trade less with each other. Third, in line with Desmet et al. (2011) as well as Becker, Enke and Falk (2016) we find a positive relationship between ancestral distance and differences in values. Using data on 861 questions from the World Values Survey for the period 1981–2014, our findings suggest that countries trade more if their people give more similar answers to questions on a wide range on values.[2] Finally, we follow previous work on migration networks and trade. Using comprehensive data on migration matrices from Artuc et al. (2015), we find that a larger overlap in populations increases bilateral trade. For all these proximate determinants of trade, we expect to find larger effects among differentiated goods. Hence, we use data from Rauch (1999) and classify commodities into three different groups: homogeneous goods, commodities with reference prices, and differentiated goods. We find that ancestral distance reduces trade for all categories. The largest estimated effects, however, are found in the sample of differentiated products.

Our work is related to several strands of literature. First, we contribute to prior research on the 'distance puzzle' which describes the fact that we know very little about why the estimated effect of geographic distance on trade flows has remained consistently high even after transportation costs as well as tariffs and non-tariff barriers have decreased substantially (Disdier and Head, 2008). Head and Mayer (2013) coined the term 'dark costs' and argue that 72–96% of the rise in trade costs associated with distance is attributable to the dark sources of resistance.[3] With respect to international trade and the impact of cultural traits (Alesina and Giuliano, 2015), our paper is closest to Guiso, Sapienza and Zingales (2009) who find that differences in bilateral trust across country pairs have substantial explanatory power in a standard gravity equation.

Second, we deviate from studying proximate determinants of trade and analyze more fundamental factors which are rooted in long-term historical differences across countries and populations. This contributes to prior work showing that ancestral distance has an impact on

---

[2]This result is in line with previous work by Cyrus (2012) who investigates to what extent cultural proximity influences bilateral trade flows using responses to the World Value Survey over time.

[3]Feyrer (2009) uses the closing of the Suez Canal between 1967 and 1975 and concludes that dark trade costs account for 50%–85% of the effect of distance on trade flow.

outcomes beyond technology diffusion and civil conflict. Guiso, Sapienza and Zingales (2009) find a negative effect of genetic distance on trade in the set of relatively homogeneous European countries but only for the intensive margin of trade. Using genetic distance as a proxy for common linguistic and cultural roots, it has further been shown that European countries with similar populations are more likely to trade with each other.[4] Felbermayr and Toubal (2010) construct a proxy for cultural proximity based on score data from the Eurovision Song Contest. The authors find that their measure of proximity is positively correlated with bilateral trade volumes. It remains unclear, however, whether these findings extend beyond the set of relatively homogeneous European countries.

Finally, our work adds to the literature investigating the consequences of heterogeneous preferences across countries. Following Linder's hypothesis, two countries may trade more with each other if they have more similar demand structures. If ancestral distance increases differences in preferences, we should observe a negative correlation of genetic distance with trade flows. As suggested by the literature on the 'home bias' (Trefler, 1995), the same observation could be made in the presence of very localized tastes, which are historically determined and change only slowly. There is number of papers discussing the impact of geography on preferences. Research has addressed this association in the context of oil versus butter (Head and Mayer, 2013), music (Ferreira and Waldfogel, 2013), websites (Blum and Goldfarb, 2006), and cereals (Bronnenberg, Dubé and Gentzkow, 2012; Atkin, 2013).

The remainder of the paper is organized as follows. Section 2 provides information on the construction of our data set as well as several descriptive statistics on the relationship between genetic and geographic distance. Section 3 describes the econometric approach, shows the main empirical results as well as a series of robustness checks. In Section 4, we discuss channels through which ancestral distance affects trade. Finally, Section 5 concludes.

---

[4]Guiso, Sapienza and Zingales (2009, p. 1128) emphasize that their "results are obtained within the boundaries of the old European Union, which comprises fairly culturally homogeneous nations" and that the impact of genetic distance "might be much larger on world trade".

2   Data

In this section, we describe our data sources and how we combine them into a single data set. Moreover, we provide descriptive statistics on all variables employed in the analysis. Our empirical work is based on a novel data set which contains information on international bilateral trade flows, country characteristics, and numerous measures of genetic, linguistic, religious, and geographic distances. We explain the source and definitions of each part separately.

2.1   Trade Data

Our data on international trade flows is taken from UN COMTRADE, a database that contains all bilateral trade flows for the year 2000. For each recorded trade flow, the data includes both the value and weight, which is available at the 6-digit commodity code level. Notably, every reporting country ('reporter') has a large set of partner countries ('partners'). For the set of countries that do not report imports and exports (i.e., a large set of poorer countries), we follow the method by Feenstra et al. (2005) as well as Helpman, Melitz and Rubinstein (2008) who impute exports and imports of non-reporting countries from the reports of (richer) countries trade flows. For example, Albania might not provide information on their *exports* to the United States. In this case, we use the *import* data from the United States. By using this method, our data set contains virtually all of the world's countries and their trade flows.[5]

A significant shortcoming of the UN COMTRADE data is that it only includes positive trade flows. In other words, the missing (or zero) trade flows are not recorded. To overcome this issue, we save the full list of (reporter and partner) countries. Using this list we create a template that contains all possible country pairs.[6] For every pair, our data set has a separate entry with each 6-, 4-, 3-, or 2-digit commodity code. As a result our template data file covers all *possible* trade flows. This allows us to investigate not only the intensive but also the extensive margin of trade.

---

[5]The only trade flows we miss are those between two countries, both of which do not submit information to the UN COMTRADE data base. These trade flows, however, comprise a negligible fraction of world trade.

[6]In terms of countries, we only remove those nations whose population is smaller than ten thousand. These countries account for only a tiny fraction of international trade. Moreover, crucial information such as GDP is usually not available.

2.2 Country Information

We merge the trade flow data with country-level information. In particular, we add data on GDP and population size for each country. As primary source for this information, we use the Penn World Table (PWT) 9.0, for which we take into account the most recent update by Feenstra, Inklaar and Timmer (2015). If there is no information for a particular country, we use the World Development Indicators as secondary or, if necessary, UNdata as third data source. Note that we use the secondary (or tertiary) data sources to predict the GDP or population value that is missing in the PWT. This makes the GDP (per capita) values comparable even if they stem from different sources.

The literature on political regimes and trade has found empirical support for the hypothesis that democracies are more likely to set up free trade areas and trade more with each other (Mansfield, Milner and Rosendorff, 2000). We follow this insight and use data from the Polity IV Project to test whether regime types affect our estimates. In particular, we use a dummy variable that takes the value one if both countries' democracy score (which ranges from 0 to 10 with higher values indicating more democratic) is above eight.

To account for trade policy, free trade areas (FTA) as well as political unions, we extend the list of variables by dummy variables for each country's membership in the EU, NAFTA, EFTA, AFTA, and Mercosur. Furthermore, we add data by Baier, Bergstrand and Feng (2014) as well as Bergstrand, Larch and Yotov (2015) who provide a database on Economic Integration Agreements (EIA). For each bilateral pair, this indicator ranges from 0 to 6 with higher values reflecting deeper integration. Finally, we use data on political rights and civil liberties from Freedom House as well as information about corruption from Transparency International.

2.3 Geographic Variables

We add a large set of geographic information to our data. The *Centre d'Études Prospectives et d'Informations Internationales* (CEPII) provides a database that comprises both information for each country as well as bilateral variables. The former includes each country's continental location, currency as well as a dummy for being landlocked. The bilateral variables provide information on geodesic distance between largest cities, contingency, common official languages,

6

colonial ties, common currencies, and legal origins.[7]

In addition, we follow insights from Giuliano, Spilimbergo and Tonon (2014) and add several 'microgeographic' variables. Drawing on data provided by Nunn and Puga (2012), these variables include longitude, latitude, a measure of terrain ruggedness, as well as the average distance to the nearest ice-free coast. Note that some of those variables pick up within-country transportation costs. Overall, the addition of variables of microgeographic factors is supposed to capture travel and communication costs between two countries. Both are barriers to international trade and likely be related to ancestral distance.[8] Furthermore, for every country we recorded to which sea it has direct access. As described in detail in the Appendix, we use this information to generate a dummy that takes the value one if two countries have access to the same sea. This leaves us with all but one variable used by Giuliano, Spilimbergo and Tonon (2014). The one missing control for bilateral geographic distance is the number of mountain chains between a country pair that is only available for European countries. However, with the plethora of the above-mentioned controls we feel comfortable that we control to the largest possible extent for geography and geographic barriers to international trade.

Linguistic Distance — Prior research by Spolaore and Wacziarg (2009, 2015) shows that genetic distance is highly correlated with other measures of cultural distance and can be used as a summary statistic for a wide array of cultural traits. Nevertheless, heterogeneity in culture is multidimensional and the use of a single variable might constitute an approximation of the multifaceted cultural distance between countries. For this reason and notwithstanding the focus on genetic distance, we also include other measures of cultural diversity. Following Spolaore and Wacziarg (2015), we take into account linguistic and religious distance between countries. Linguistic distance is determined as an index based on language trees. Depending on their similarities, linguists group languages into families. Every language belongs to a set of families according to its different characteristics and the higher the number of common families, the more

---

[7]Instead of using the geodesic distance between two countries' largest cities, we can also apply a population-weighted distance. However, the two measures are highly correlated and we decided to use the distance between the largest cities in all specifications.

[8]In Table A.3 in the Appendix, we show how these micro-geographic variables are correlated with bilateral genetic distance.

similar two languages are. Vice versa, if two languages do not belong to any common family, they have the greatest language difference. The language-trees approach allows to calculate distances for 157 countries and 12,246 country pairs.

Religious Distance — For religious distance, we use the measure developed by Spolaore and Wacziarg (2015). They group religions into broad categories and further divide them into fine classifications. Using methodology analogous to the language tree cited above, they find measures of religious distance. The authors use two sources of religious trees and build both pluralistic and weighted measures. The reason for considering multiple sources is that is that the most detailed tree is only available for 157 countries (12,246 country pairs) whereas a less refined one is available for 197 countries (19,306 country pairs).

2.4 Genetic Distance

In order to investigate the impact of ancestral distance on international trade flows, we use information on genetic differences to proxy for ancestral distances.[9] Our analysis uses a bilateral data set on genetic distance provided by Spolaore and Wacziarg (2016). This data set combines two sources of information. First, genetic differences between a large number of distinct populations measured by Pemberton, DeGiorgio and Rosenberg (2013). Second, Alesina et al. (2003) provide data on the composition (fractionalization) of more than one hundred countries. Spolaore and Wacziarg combine these two sources into an extensive database on bilateral genetic distances between countries. Notably, this data set updates Spolaore and Wacziarg's earlier data on bilateral genetic distances that was used in a number of research papers.[10]

In our paper, we can use both the older and newer data set on genetic differences. The main advantage of the newer data is that it is based on a significantly larger set of unique populations. Instead of relying on 42 populations defined by Cavalli-Sforza, Menozzi and Piazza (1994), the new data set is based on 267 worldwide populations compiled by Pemberton, DeGiorgio and Rosenberg (2013). Hence, we have much more detailed information on the relationship between

---

[9]Dawkins (2004) provides an extensive, nontechnical discussion of history in the context of ancestral distance.

[10]For example, the study by Spolaore and Wacziarg (2009) documents that genetic distance to the most technologically advanced country is strongly related to income differences across countries.

populations in all countries, especially within Africa and Asia.

While genetic distance can be considered as a summary statistic for intergenerationally transmitted traits across populations like culture, we also draw on data for other measures of cultural differences between countries. For this, we rely on Spolaore and Wacziarg (2015) who provide a comprehensive database that includes several measures of linguistic, religious, and cultural distance. They also show that although measures of cultural distance are poorly correlated to one another, genetic distance is positively correlated with all of them. Thus we use genetic differences as our preferred measure of ancestral distance in the empirical analysis. In what follows, we provide a brief summary of how genetic distance is measured.

Genetic Distance as Molecular Clock — It is crucial for our analysis to understand why genetic differences between populations can be used as a proxy for ancestral distance. Hence, we explain in detail how genetic distance is measured and how it might affect bilateral trade flows. Whereas all people in the world share the same gene variants, their frequencies are different across populations. In order to measure genetic distances between populations, researchers use differences in genetic markers at the molecular level. Pemberton, DeGiorgio and Rosenberg (2013) consider genetic differences occurring as a consequence of microsatellite variations.[11] Microsatellites are tracts of repetitive noncoding DNA in which short DNA motifs are repeated between 5 to 50 times.

The advantage of using microsatellites in population genetics is due to their high mutation rates, their high diversity, and their noncoding nature. The last point implies that microsatellites do affect the chromosome structure but not the ribonucleic acid (RNA), which is the molecule implicated in the expression and codifications of genes.[12] This aspect is very important because it renders the genetic distance measure independent from possible somatic differences and genetic endowments of different populations. As a result, what we are measuring

---

[11]The work by Pemberton, DeGiorgio and Rosenberg (2013) follows extensive research based on the Human Genome Diversity Project (HGDP-CEPH) which is described in detail by Cann et al. (2002).

[12]Though the vast majority of microsatellites are indeed situated between codifying regions and remain biologically silent, some of them might end up inside regulatory of even coding DNA, giving rise to phenotypic changes and/or diseases. However, data on populations genetics is based on the highly mutational microsatellites located in noncoding DNA regions.

is ancestral relatedness between populations independently from their looks and leaving aside any quality judgment with respect to their genetic outfit. In other words, our main explanatory variable (ancestral distance) can affect trade flows because it captures similarities in beliefs, conventions, customs, or habits.

As an important final note, we emphasize that all measures of genetic distance are symmetric in Spolaore and Wacziarg (2016). The same applies to measures of religious or linguistic distance that are provided by Spolaore and Wacziarg (2015).

Quantifying Genetic Distance — Having data on differences in microsatellites among populations, Pemberton, DeGiorgio and Rosenberg (2013) build a summary genetic distance measure F_{ST} based on the probability that two randomly selected alleles at a given locus are different within a population (heterozygosity). Considering the average heterozygosity between two populations, h_m, and the heterozygosity in the sum of the two populations, h, genetic distance is defined as:

$$F_{ST} = 1 - \frac{h_m}{h} \quad \text{with } 0 \le F_{ST} \le 1 \tag{1}$$

In order to better understand this measure, we can consider the example provided by Spolaore and Wacziarg (2009). Suppose we have two populations called a and b as well as a biallelic gene taking values 1 and 2. The gene frequencies are p_a and q_a for each allele in population a and p_b, q_b for population b, respectively[13] so that the average allele frequencies between the two populations are $\bar{p} = \frac{p_a + p_b}{2}$ and $\bar{q} = \frac{q_a + q_b}{2}$. The heterozygosity for population a and b is, respectively:[14]

$$h_a = 1 - (p_a^2 + q_a^2) = 2p_a q_a \tag{2}$$

$$h_b = 1 - (p_b^2 + q_b^2) = 2p_b q_b \tag{3}$$

---

[13]Note that a biallelic gene can only take one of the two forms so that $p_i = 1 - q_i$ for $i \in a, b$. In this case $p_i + q_i = 1$ so that $(p_i + q_i)^2 = p_i^2 + q_i^2 + 2p_i q_i = 1$.

[14]Note that Ashraf and Galor (2013) used this measure of within-country heterozygosity to investigate the impact of genetic differences on productivity.

We can then find the heterozygosity in the sum of the two populations:

$$h = 1 - (\bar{p}^2 + \bar{q}^2) = 2\bar{p}\bar{q} \tag{4}$$

and the average heterozygosity between the two populations:

$$h_m = \frac{h_a + h_b}{2} \tag{5}$$

In this case, genetic distance between the two populations is given by

$$F_{ST} = 1 - \frac{h_m}{h} = \frac{(p_a - p_b)^2}{4\bar{p}(1 - \bar{p})} \tag{6}$$

The genetic distance between populations is only equal to zero when the frequencies of genetic markers is the same across two populations, i.e. $p_a = p_b$. However, if one population only displays alleles which are not present in the other population, F_{ST} take the value of one.

Spolaore and Wacziarg (2016) provide two measures of genetic distance between countries: a weighted and a non-weighted one. The non-weighted measure simply shows the value of F_{ST} for the ethnic group having the largest share in the country's population. The weighted measure, in contrast, is based on a weighted average of all ethnic groups residing in a country:

$$F_{ST}^W = \sum_{i=1}^{I} \sum_{j=1}^{J} (s_{1i} s_{2j} d_{ij}) \tag{7}$$

where s_{1i} is the share of population i in country 1, s_{2j} is the share of population j in country 2, and d_{ij} is the genetic distance F_{ST} between population i and population j. It is worth noting that the correlation between the two measures of genetic distance is very high (0.917). For our empirical analysis, we prefer using the weighted measure as it represents more precisely the expected genetic distance between two randomly selected individuals from two different countries. Note that genetic distance is a continuous measure (unlike, for example, linguistic and religious distance). In the robustness tests, we use both the non-weighted genetic distance based on Pemberton, DeGiorgio and Rosenberg (2013) as well as the older measures of genetic distance based on Cavalli-Sforza, Menozzi and Piazza (1994).

2.5 Descriptive Statistics

We use two different data sets on trade statistics for the year 2000. The first one includes information on trade flows for all commodities, while the second one is based on a 2-digit commodity dimension for country pairs. Table 1 shows descriptive statistics for each variable in our data set.

— Table 1 about here —

Overall, we have 172 countries in our data set. Hence, there are $172 \times 171 = 29{,}412$ bilateral observations. In our sample the GDP per capita ranges from 192 to 101,248 U.S. dollar. The average country has an income level of about 10,674 U.S. dollar. As the numbers of observation indicate, we have country information available in all cases except for the corruption index as well as Freedom House variables on political rights and civil liberty. However, we only use these variable in a few regressions. At the 2-digit level, we have 69 commodities and can use $69 \times 172 \times 171 = 2{,}029{,}428$ observations. In the 4-digit sample that comprises 782 goods, we have 23,000,184 observations.

2.6 Relationship between Genetic and Geographic Distance

A straightforward way of illustrating the relationship between genetic distance and geographic distance is to depict genetic distance for each potential trading partner on a map. We do this in Figure A.1 in the Appendix for both the United States and Uganda. While the U.S. shows a small genetic distance to, for example, European countries, Brazil, or Australia, we see that Uganda is genetically very distant from virtually all major economic markets. Notably, there is a concave relationship between geographic and genetic distance as we show in Figure A.2 in the Appendix.

Using our data, we can document a clear positive relationship between geographic and genetic distance. Figure 1 depicts the relationship between genetic a country's average genetic distance (x-axis) and its average geographic distance (y-axis). Panel (a) provides an unweighted version in which every country's geographic and genetic distance enters with the same weight, while Panel (b) weights both distance variables according to the trading partners' GDP.

Panel (a) in Figure 1 reveals that, on average, a country that is separated from the rest of the world by a larger geographic distance also shows a higher genetic distance to its potential trading partners. However, this relationship is not perfect. In particular, a set of African countries is genetically far more distant from the rest of the world than any other group. When we weight each potential trading partner by its GDP in Panel (b), those (mostly) African countries are even more remote.

3 Ancestral Distance and Trade

We now turn to our econometric analysis of the effect of ancestral distance on trade. In particular, we test whether ancestral distance reflects a barrier to trade for both the extensive and intensive margin.

3.1 Econometric Approach

Our empirical model mimics a gravity equation as in Tinbergen (1962), Anderson and Van Wincoop (2003) as well as Anderson (2011). For the econometric analysis, our baseline model is given by

$$y_{o,d} = \beta G_{o,d} + \alpha D_{o,d} + \mathbf{X}_{o,d}\boldsymbol{\beta} + \varepsilon_{o,d} \tag{8}$$

where $y_{o,d}$ denotes the dependent variable which can either be either the extensive or intensive margin of trade. For the former, we use a dummy variable taking the value one if two countries trade with each other. In contrast, the intensive margin is measured by the total trade volume in U.S. dollar. We denote the origin country by o and the destination country by d. Our main variable of interest, $G_{o,d}$, indicates the genetic distance between the country pair. As explained in Section 2, genetic distance measures the time since the two countries' populations have been the same population ('molecular clock'). Additional variables on the right-hand side include the geodesic distance $D_{o,d}$ between the countries.[15] Furthermore, we

---

[15]The the geographic distance, we use the shortest route between two countries' largest cities on the Earth's surface. This is typically referred to as geodesic distance. Alternatively, we can use the geodesic distance between

add a vector of control variables denoted by $\mathbf{X}_{o,d}$. The latter includes a varying set of variable in order to examine the conditional correlation of ancestral distance with trade. Note that $\mathbf{X}_{o,d}$ can include both country-specific as well as bilateral terms. We follow Cameron, Gelbach and Miller (2011) as well as Egger and Tarlea (2015) and cluster the standard error ($\varepsilon_{o,d,c}$) at the country-pair level.

In the first part of our empirical analysis, we use data on aggregate bilateral trade flows from the year 2000. This choice reflects the fact that ancestral distance between countries was measured based on populations at that time. For the robustness section, we will also use trade data from other periods and assume that the composition of country populations remains very stable in the short and medium run. Additionally, we use an extended data set on bilateral trade flows for a set of 69 different commodities. In this case, the model is given by

$$y_{o,d,c} = \beta G_{o,d} + \alpha D_{o,d} + \mathbf{X}_{o,d}\boldsymbol{\beta} + \gamma_{o,c} + \delta_{d,c} + \varepsilon_{o,d,c} \tag{9}$$

where c denotes the commodity and fixed effects are added to the right-hand side of the equation. These fixed effects reflect multilateral resistance terms as suggested by Anderson and Van Wincoop (2003). Note that we use origin-commodity and destination-commodity fixed effects as suggested by Head and Mayer (2014). As before, the standard error ($\varepsilon_{o,d,c}$) is clustered at the country-pair level.

Whether or not we use data with commodity dimension, we use a Probit estimator for the extensive margin. With respect to the intensive margin of trade, we follow Santos Silva and Tenreyro (2006, 2011) as well as Fally (2015) and apply a Poisson pseudo-maximum-likelihood (PPML) estimation technique to account for heteroskedasticity as well as zero trade flows.

Transportation Costs — If ancestral distance constitutes a barrier to international trade, it appears crucial to control for transportation costs. Anderson and Wincoop (2004) define trade costs as "all costs incurred in getting a good to a final user, other than the marginal cost of producing the good itself". Our paper aims at investigating the importance of ancestral distance

---

two countries weighted by population. Due to the very high correlation between the two distance measures, our results are not affected by which measure we apply.

on trade. Insofar as this distance is a good proxy for familiarity, controlling for transportation costs (which should not depend on ancestral distance) would be ideal. There are, however, a number of issues with transportation costs data. First, information are only available for those country-pairs who actually engage in trade. This means that we would have a missing value every time that there is a zero in the bilateral trade matrix. Second, the most widely available measures are published by the International Monetary Fund's *Direction of Trade Statistics* and are based on the ratio of carriage, insurance, and freight (henceforth c.i.f.) to free on board (f.o.b.) values. Such measures are aggregated over commodities and might not represent the true value due to their dependence on compositional change. Third, the IMF database includes imputed observations. The imputation method, however, is pretty inaccurate (Hummels, 2007).

Due to lack of data, however, we decided not to include a direct measure of transportation costs in the empirical analysis. Instead, we control for all factors that might have a direct impact on bilateral transportation costs. More specifically, we control for geographical distance, ruggedness of the terrain, the presence of a common border, access to the same sea as well as numerous other geographic and political variables shown in Table 1. In addition, we add GDP per capita (of both the exporting and importing country) to each specification as prior research suggests that poor countries face higher transport costs (Foellmi, Hepenstrick and Zweimüller, 2012; Simonovska and Waugh, 2014). The fixed effects for exporter and importer account for additional time-invariant country-specific factors that have an impact on trade costs.

3.2 Descriptive Evidence

Before turning to econometric estimates, we briefly discuss the relationship between ancestral distance, geographic distance, and trade at the descriptive level. Do worldwide bilateral trade flows correlate with ancestral distance between populations? In Figure 3, we show the relationship between ancestral distance and bilateral trade flows. Netting out geographic distance and breaking genetic distance into deciles, we find a significant negative gradient.[16] This suggests that countries with a larger genetic distance trade less with each other.

---

[16]Note that we plot on the horizontal axis the actual genetic distance minus the predicted genetic distance based on geographic distance. For the vertical axis, we use the residuals from regressing trade on geographic distance.

Notably, the negative association can be found for both the intensive and extensive margin of trade. Countries that are genetically further apart are less likely to trade with each other and, if they do trade, they trade a smaller volume. We can also plot these two figures with all country pair observations. In addition, we provide univariate regression results in column (1) of Table 2. The coefficients we obtain for the extensive and intensive margin of trade show the expected negative sign and are highly statistically significant.

The question, however, remains whether geographic and genetic distance measure the exact same thing and thus its effects cannot be disentangled. We argue that while genetic and geographic distance are highly correlated, as documented by Figure A.2 in the Appendix, the correlation is not perfect. In particular, countries like Australia and the United Kingdom are genetically far closer than what one could expect based on their geographic distance. We can show this more formally by first regressing trade flows on geographic distance and storing the residuals. In the second step we regress bilateral genetic distances on geographic distance and again store the residuals. Finally, we plot the two residuals against each other.

Panel (a) of Figure 3 in the Appendix plots on the vertical axis the residuals of regressing log Trade Value on geographic distance. On the horizontal axis, we show the difference between genetic distance and predicted genetic distance (based on geographic distance). Hence observations on the right-hand side reflect country pairs with a genetic distance that is lower than one would expect based on their geographic distance (e.g., Australia and UK). In panel (b) of Figure 3, we replace the vertical axis by the extensive margin of trade. The key observation of both plots in Figure 3 is that country pairs with a large genetic distance —larger than expected by their geographic distance— are less like to trade and trade smaller volumes if they do trade. This finding suggests that genetic distance is a barrier to international trade on top of geographic distance. In other words, while geographic distance picks up a substantial fraction of the effect of genetic distance on trade, it does not fully account for the effect. Thus if we intend to understand trade flows between countries and measure trade costs, we have to include genetic distance as a significant barrier to trade.

3.3 Main Regression Results

We now turn to the main results of our econometric estimation. Using data from the year 2000, we fit the empirical model describe in equation (8). First, we examine the extensive margin of trade and estimate whether genetic distance affects the propensity of positive trade flows. In a second step, we investigate whether positive trade flows are reduced by a higher ancestral distance. For both margins of trade, different specifications (i.e., sets of control variables) are used to examine the robustness of the correlation between ancestral distance and trade. In all specifications, the coefficient of interest, α in equation (8), shows the correlation of trade between two countries with the respective their ancestral distance.

For both the extensive and intensive margin, we begin with a univatirate specification to see how much of the variation in trade can be explained by ancestral distance. Subsequently, we add various control variables to examine the robustness of the correlation of ancestral distance with trade. Furthermore, by adding variables on the right-hand side of in equation (8), we can potentially investigate why trade correlates with ancestral distance. Note that we first control only for measures of geographic isolation. When adding political variables or other measures of cultural distance, we expect that the negative correlation between ancestral distance and trade might be weakened. This would reflect the fact that ancestral distance is a summary statistic for the relatedness of two populations. Common beliefs, conventions, or habits that we proxy for by using genetic distance, also explain why two countries have a free trade agreement, colonial ties, similar languages, and so on. As a result, part of the effect of ancestral distance on trade will be picked up by other variables in the regression.

Extensive Margin of Trade — Prior research by Baldwin and Harrigan (2011) emphasizes the importance of distinguishing between the number of goods shipped and the value of trade flows. Given the large number of country-pair-commodity observations with zero trade volume, we first investigate the extensive margin of trade. In the top part of Table 2, we show the results of estimating equation (8) using our data set with a 2-digit commodity dimension.

— Table 2 about here —

In column (1), we replicate Figure 3 and show that genetic distance is negatively associated

17

with the probability that two countries have positive trade flows. It is important to note that ancestral distance alone accounts for 5.3% of the variation in trade flows at the extensive margin. In columns (2) and (3), we add geodesic distance to the right-hand side of the regression and find that it has a statistically significant negative coefficient. Since both genetic and geodesic distance are entered in logarithmic form, we can compare the magnitudes. While column (3) suggests genetic distance to be more important than geodesic distance, this differs in the following columns.

We add genetic distance to a standard gravity equation with both countries' GDP and geodesic distance in column (4). This yields a much smaller coefficient on genetic distance than in any of the other specification. We interpret this finding as being in line with research by Spolaore and Wacziarg (2009) who document that income differences are linked to genetic distance. Next, we follow research by Giuliano, Spilimbergo and Tonon (2014) and add numerous micro-geographic measures to the equation. While we find that the coefficients on these variable show the expected signs, the estimated effect of genetic distance on trade remains similar to column (3) where we only controlled for the geodesic distance.

In columns (6), we apply controls for political variables. In particular, we find that colonial ties and deeper economic integration (i.e., free trade areas or common currencies) are positively correlated with trade. The correlation between genetic distance and trade, however, remains unaffected by the addition of political control variables. We obtain a similar result when adding linguistic and religious distance to the equation. Column (7) reveals that countries trade more if they share a common official language. Conversely, they are less likely to trade the larger their linguistic and religious distance.[17] This is in line with our expectation and previous research suggesting that language and religion are only two of the many characteristics that increasingly differ across populations the longer the time since they were one population.[18] In other words, ancestral distance is correlated with linguistic and religious difference but this correlation does

---

[17]Note that we do not use the natural logarithm of linguistic and religious distance because this would imply losing all country pairs with a distance of zero. This is a particularly severe problem with linguistic distance due to the high number of zeros in the data.

[18]Heterogeneity in culture is multidimensional and the use of a single variable might constitute an approximation of the multifaceted cultural distance between countries. Recent work by Desmet, Ortuño-Ortín and Wacziarg (2015) argues that within-ethnic-group variation is larger than between-ethnic-group variation. Hence, ethnic and cultural diversity should be considered separately.

not capture why ancestrally distant countries are less likely to trade with each other.

Finally, in column (8) we address the concern that ancestral distance might only reflect cross-continental barriers to trade. As shown in Figure A.1, populations living on the same continent usually have a similar ancestral past. In order to test whether our estimated coefficient on ancestral distance in fact picks up such continental fixed effects, we follow a suggestion by Spolaore and Wacziarg (2009). In particular, we add two sets of dummy variables to the right-hand side of equation (8): for each continent we have a dummy variable taking the value one if both countries are located on this continent. Additionally, we have dummy variables for each continent that take the value one if one country is on it but the other is not. The inclusion of these continent dummy variables does not alter the estimated coefficient on genetic distance. In contrast, we obtain a standardized beta of 22.6% which is very similar to what we obtained in column (3) with geodesic distance as control variable.

Intensive Margin of Trade — Given the evidence that countries with a large ancestral distance are less likely to establish trade relations, we now turn to the intensive margin of trade. Hence, we estimate equation (8) using the volume of trade flows on the right-hand side. Note that applying a PPML estimator, we use the large number of zeros in the trade data for our estimation. The results are shown in the lower part of Table 2. First, we observe that ancestral distance is negatively correlated with the total trade volume in a univariate regression. While this resembles the finding of Figure 3, it is important to note that ancestral distance alone accounts for only 1% of the variation in trade flows at the intensive margin. This is much smaller than in the case of the extensive margin. When adding geodesic distance as control variable, the coefficient on genetic distance remains statistically significant at the 1% level but is reduced to half its magnitude. The standardized beta shows that increasing the genetic distance by one standard deviation reduces the bilateral trade volume by 7.1% of a standard deviation.[19]

In column (3), we estimate a standard gravity equation with GDP, geodesic distance, and genetic distance. As was the case with the extensive margin, we observe that the coefficient

---

[19]The interpretation of the coefficients from the Poisson model is straightforward. Despite having the dependent variable specified as trade flows in levels (rather than in logarithms), the coefficients of any log-transformed independent variable can be interpreted as simple elasticity.

on genetic distance is strongly reduced by adding GDP to the regressors. However, for the intensive margin we find that the coefficient actually becomes insignificant. The same occurs in columns (6) and (8) where we add political controls and continental dummy variables, respectively. Notably, by adding measures for micro-geographic distance (column 5) or linguistic and religious distance (column 7) we still obtain a highly statistically significant coefficient on genetic distance.

Overall, we can conclude that genetic distance is negatively correlated with the total bilateral trade volume. This correlation remains statistically significant even if we add a large set of micro-geographic control variables to the regression. However, the results of Table 2 suggest that ancestral distance picks up some aspects (like income differences) which drive the intensive margin of trade.

3.4   Robustness Tests

Our main empirical findings based on the model specified in equation (8) suggest that the ancestral distance between two countries significantly affects trade flows, especially at the extensive margin. In order to explore the robustness of this finding, we conduct several tests which we describe in the following subsections.

3.4.1   Control Variables

In the analysis shown in Table 2, we piece-wise add and remove various sets of control variables. This allows us to study the robustness of the correlation between ancestral distance and trade with respect to numerous potential omitted variables. However, we can also explore whether the negative effect of ancestral distance on trade is also present if we add all control variable to the right-hand side. Furthermore, we add fixed effects for both origin and destination country. This yields the model described in equation (9) which we apply to both the extensive and intensive margin of trade.

— Table 3 about here —

The results we obtain are shown in column (1) of Table 3. We find evidence that ancestral distance negatively affects both the extensive and intensive margin of trade. Furthermore, we

20

observe an expected substantial reduction in the magnitude of the coefficient. This reflects the fact that ancestral distance also affects institutional and other economic differences (Spolaore and Wacziarg, 2009, 2015).

3.4.2 Set of Countries

A possible concern with our analysis could be that the estimated effect of ancestral distance on trade depends on the selection of countries in the estimation sample. In particular, we address the question whether the effect we find in the main specification is driven by specific countries who trade very little because of their status of development and also happen to be genetically distant from one-another. The fact that a set of African countries constitutes an outlier group in Figure 1 raises the question whether the estimated effects are solely driven by this group. In order to address these concerns, we replicate the analysis and exclude the subsample of mostly Sub-Saharan African countries with a very high average genetic distance (above the 90th percentile). The results shown in column (2) of Table 3 indicate that the negative impact of genetic distance trade is not sensitive to this reduction in the sample of countries.

3.4.3 Alternative Genetic Distance Measures

Throughout the empirical analysis so far, we used the weighted genetic distance provided by Spolaore and Wacziarg (2015) as a proxy for ancestral distance. As explained in Section 2, this variable is favorable to alternatives for two main reasons. First, it is based on a large set of 267 worldwide populations. Second, it takes into account the diversity within populations of the same country. However, we can use alternative proxy variable for ancestral distance to explore the robustness of our main finding.

In column (3) of Table 3 we replace the weighted genetic distance by the genetic distance between the dominant populations of origin and destination country. This variable is also provided by Spolaore and Wacziarg (2016). Notably, we obtain an almost identical coefficient as in column (1), suggesting that ancestral distance between the dominant groups is driving the effect on trade. A second alternative measure is the weighted genetic distance based on a smaller set of populations defined by Cavalli-Sforza, Menozzi and Piazza (1994). This variable

21

is provided by Spolaore and Wacziarg (2015) and has been used in a number of research papers. Using this variable in our specification, we obtain negative coefficient that is significant at the 5% level for the extensive margin and insignificant with respect to the intensive margin.

3.4.4 Different Time Period

In the main analysis we use the year 2000 as benchmark as it roughly corresponds to the year in which the data on genetic distances (and the composition of countries) were measured (Alesina et al., 2003). Under the assumption that the population of each country does not change drastically within a relatively short period, we can replicate our analysis using trade data from the other years prior or after 2000. Results shown in column (5) of Table 3 reveal that the estimates for the effect of genetic distance on trade are very similar to our main results for the extensive margin. For both the intensive margin of trade, however, we obtain an insignificant coefficient.

3.4.5 Commodity-Level Trade Data

In our empirical work so far, we relied on aggregate bilateral trade data. We can extend this analysis by using detailed trade data at the 2-digit SITC commodity code level. This comprises a total of 69 separate groups of goods. With this data, we re-run the model specified in equation (8). Results shown in column (6) of Table 3 indicate that the estimated effect of genetic distance on trade is largely identical to the results we obtain with aggregate trade data. For both the extensive and intensive margin of trade, we find a statistically significant negative relationship with ancestral distance.

3.4.6 Number of Goods Traded

Following research by Hummels and Klenow (2005) as well as Huberman, Meissner and Oosterlinck (2015), we investigate whether ancestral distance also affects the number and range of goods countries trade with each other. The key finding of our empirical analysis is that ancestral distance is negatively associated with both the intensive and extensive margin of trade. Another approach for testing this effect is to replace the volume of bilateral trade by the number

of goods that a country pair trades with each other. In column (7) of Table 3, we show that genetic distance also reduces this number. Using the 2–Digit commodity data, the estimated coefficient indicates that countries which are ancestrally further apart trade fewer goods with each other.

4   Proximate Determinants of Trade

The empirical analysis thus far has provided evidence that countries with a larger ancestral distance are less likely to trade and, if they engage in trade, they ship fewer goods and lower quantities. As discussed in Section 2, ancestral distance serves as a summary statistic of relatedness between populations. In this section, we explore the role of different dimensions of ancestral distance to further investigate the role of several so called "proximate determinants of trade". We first document how genetic distance is related to these proximate determinants and then show how these determinants relate to trade.

4.1   Ancestral Distance and Proximate Determinants of Trade

Differences in Technology — A large literature has emphasized that ancestral distance affects the spread of technological inventions (Diamond, 1997; Spolaore and Wacziarg, 2009). Individuals are more likely to learn from closed relatives and friends than from strangers because the former tend to speak the same language and share similar beliefs and norms. Similarly, closely related countries are more likely to adopt innovations developed by the other country. To explore the relationship between ancestral distance and differences in technology, we use a recent dataset that includes measures of the technological level in five sectors, namely agriculture, transportation, communication, military and industry (Comin, Easterly and Gong, 2010). For each country, we take the average of the technological level across sectors. We then construct a country-pair dataset that calculates the difference in the average technology level and standardizes this variable to a mean of zero and a standard deviation of 1. As an example, the resulting technology difference amounts to 0.66 for the United States and Ghana, while it is only 0.17 for the United States and France. More interestingly, ancestral distance is

strongly related to the average technological level as suggested by Panel (a) in Figure 4. A one standard deviation increase in ancestral distance increases technological differences by 32.5% of a standard deviation when we account for country-specific fixed effects as documented in Table A.4 in the Appendix.

— Figure 4 about here —

Differences in Preferences — A second important proximate determinant through which ancestral distance affects trade are preferences. Previous research has documented that local customs, habits, and culture may determine regional food preferences (Dubois, Griffith and Nevo, 2014; Atkin, 2016). This high geographical correlation of preferences extends to non-food products. Several scholars have put forward the explanation that past experiences are an important driver of present consumption (Bronnenberg, Dubé and Gentzkow, 2012). As parents shape individual preferences in childhood, ancestral distance is likely to be positively correlated with differences in preferences.

To investigate the relationship between ancestral distance and preferences, we construct a measure of differences in preferences. We first define a dummy variable that takes the value one if a country either exports or imports a certain commodity, and zero otherwise. Assuming that any good that is produced and consumed domestically is —- at least to a small amount — also exported to some partner country, we argue that this measure is a proxy for the extensive margin of consumption.[20] For a given country pair, we then calculate the share of goods that is consumed in both countries. This measure reflects the probability that two countries A and B both consume a commodity. Panel (b) in Figure 4 documents that this measure is strongly correlated with ancestral distance. A one standard deviation increase in ancestral distance increases differences in preferences by 32.0% of a standard deviation when we account for country-pair-specific fixed effects (see Table A.4 in the Appendix).

---

[20]The population of a country can only consume a good if it is produced domestically or imported. While imports show up directly in our data set, we assume that any local production would show up as exports to at least one country.

Differences in Values — Countries with common ancestry not only share similar technology and preferences, but also tend to have similar values regarding the role of the family, individual perceptions of life, and moral and religious questions. Desmet et al. (2011) document that genetic distance is strongly correlated with answers to the World Values Surveys (WVS). Following Desmet et al. (2011), we link ancestral distance to differences in values using all seven waves of the WVS. We use the longitudinal multiple-wave edition of the World Value Survey dataset, which includes a total of 341,271 surveyed individuals from 100 countries[21] for the period 1981–2014 with $(89 \times 90)/2$ unique pair observations. We compute the Manhattan distance based on the answers of 861 questions for which WVS information is available. The average Manhattan distance between countries j and k is calculated as

$$w_{jk} = \frac{1}{861} \sum_{i=1}^{861} \sum_{s=1}^{q} \left| x_{i,j}^s - x_{i,k}^s \right|$$

where i indexes questions and s indexes answer categories. $x_{i,j}^s$ is the share of respondents in country j who chose answer category s when answering question i. Panel (c) in Figure 4 documents that this measure is strongly correlated with ancestral distance. A one standard deviation increase in ancestral distance increases differences in values by 16.6% of a standard deviation.

Networks — Finally, ancestral distance is closely related to past and current migration patterns. An emerging literature has documented that historical migration has shaped the genetic composition within and between populations (Ashraf and Galor, 2013). In particular, the migratory distance to Africa is negatively related to genetic diversity *within* a population because — according to the so called serial founder effect — subgroups of settlers who expanded across the planet carried with them only a subset of the overall genetic diversity of their parental colonies. As a consequence of these ancient migration patterns, the common gene pool *between* populations has decreased. More recent migration patterns are likely to affect ancestral distance in a similar way. This recent migration creates networks that are likely to directly affect inter-

---

[21]We exclude the following countries because we lack data on genetic distance: Puerto Rico, Tanzania, Yemen, Arabic Republic, countries in former Yugoslavia.

national trade (Rauch, 1999; Rauch and Trindade, 2002). We use data on migration stocks to further investigate the effects of ancestral distance on networks based on a collection of census and register data compiled by Artuc et al. (2015). Panel (d) in Figure 4 documents that recent migration stocks are negatively correlated with ancestral distance. A one standard deviation increase in ancestral distance decreases the by 8.0% of a standard deviation in the estimation without fixed effects. However, when we take into account the heterogeneity by including fixed effects for the origin and destination country, the coefficient decreases to close to zero and is no longer statistically significant.

4.2   The Importance of Proximate Determinants of Trade

The previous section has shown that ancestral distance affects several proximate determinants of trade. In the following, we assess whether one of these proximate determinants of trade is able explain the overall impact of ancestral distance. If ancestral distance primarily operates through one specific proximate determinant, the relationship between genetic distance and trade should weaken once we include this specific determinant. Table 5 reports the results of this test for both the extensive and intensive margin of trade. The entries in column (3) and (5), for which we have the same number of observations as in the baseline specification in column (1), suggest that the effect of ancestral distance is slightly weaker but still significant. Similarly, the effect of ancestral distance is not reduced in columns (2) and (4). Overall, these findings suggest that none of the proximate determinants is able to fully explain the effect of ancestral distance on trade.

— Table 5 about here —

4.3   Ancestral Distance, Culture and Incomplete Contracts

Ancestral distance is strongly linked to culture, often defined as "the customary beliefs and values that ethnic, religious, and social groups transmit fairly unchanged from generation to generation" (Sapienza, Zingales and Guiso, 2006). Recent research has provided evidence that ancestral distance influences trade primarily via differences in culture (Guiso, Sapienza and Zingales, 2009; Desmet et al., 2011). Most prominently, scholars have pointed out that higher

26

levels of trust between two populations lead to more trade. To formally examine the consequences of trust on trade, the seminal paper by Guiso, Sapienza and Zingales (2009) proposes a novel test.

The test is based on a dataset by Rauch (1999) who distinguishes three types of goods, namely goods traded in an organized exchange, goods with a reference price, and differentiated goods. Rauch argues that goods can be traded in an organized exchange only if they are very homogeneous in quality. In the same vein, they can have a reference price only if they are similar in terms of their intrinsic quality. Guiso, Sapienza and Zingales (2009) point out that Rauch's classification can also be interpreted as a classification of the degree of trust intensiveness of the different goods. While iron ore and cotton are both homogeneous product and thus can be traded on organized exchanges, cut flowers and foliage are more differentiated in quality. According to the test, this higher variation in quality makes it more difficult to write contracts and hence gaps in the contract are more likely for differentiated goods. These incomplete contracts then lead to deals that are often made just by shaking hands, which requires a higher level of trust between the contracting parties. As a consequence, we expect a higher effect of ancestral distance for differentiated goods.

— Table 6 about here —

We merged Rauch's commodity classification to our 4-digit trade data and ran our preferred regression specification using the number of goods traded per country-pair as a dependent variable (in logs). Table 6 reports the results. We find that the effect of ancestral distance is substantially larger for differentiated than for homogeneous and reference-priced products. This result is consistent with the notion that ancestral distance operates through cultural variables, most prominently trust.

5 Conclusion

This paper investigates the impact of ancestral distance on both the extensive and intensive margin of trade. We use a new data set on bilateral trade flows for 172 countries combined with detailed information about their ancestral distance which we proxy using genetic differences.

Our results show that ancestral distance has a statistically and economically significant negative effect on both the probability of establishing trade relations (extensive margin) and on the amount of goods traded (intensive margin). These findings are robust to the inclusion of a large set of control variables for geography and other cultural distance measures.

We show that ancestral distance captures differences in preferences, values, norms, habits, and technology across populations. Hence, we interpret our empirical results as evidence for deep-rooted human barriers to international trade. Insofar as international trade constitutes an important channel through which technological innovations spread across countries, ancestral distance thus represents a significant barrier to the diffusion of development. From a policy perspective, one might argue that genetic distances between countries cannot be changed, at least in the short run. Therefore, the question arises what we learn from establishing a negative association between trade flows and genetic distance. We argue that —while it is not possible to alter genetic distances— it is possible to change the effect of genetic distance as a barrier to trade. For example, removing tariff barriers between ancestrally distant countries seems to be particularly efficiency-enhancing because these countries are unlikely to trade with each other even in the presence of pronounced comparative advantages. For example, Frazer and Biesebroeck (2010) find that tariff reductions between African countries and the U.S. in the African Growth and Opportunity Act (AGOA) had a positive effect of exports from Africa into the United States.

Based on our findings, we conclude that there is a strong relationship between ancestral distance and trade. There is a large body of literature documenting that a key benefit from international trade is the expansion of varieties for consumers (Krugman, 1979; Broda and Weinstein, 2004). Thus investigating the welfare implications of ancestral distance appears to be a fruitful task for future research.

References

Alesina, Alberto, and Paola Giuliano. 2015. "Culture and Institutions." *Journal of Economic Literature*, 53(4): 898–944.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth*, 8(2): 155–194.

Anderson, James E. 2011. "The Gravity Model." *Annual Review of Economics*, 3(1): 133–160.

Anderson, James E., and Eric Van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review*, 93(1): 170–192.

Anderson, James E., and Eric Van Wincoop. 2004. "Trade Costs." *Journal of Economic Literature*, 42(3): 691–751.

Artuc, Erhan, Frédéric Docquier, Çaglar Özden, and Christopher Parsons. 2015. "A Global Assessment of Human Capital Mobility: The Role of Non-OECD Destinations." *World Development*, 65: 6–26.

Ashraf, Quamrul, and Oded Galor. 2013. "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review*, 103(1): 1–46.

Atkin, David. 2013. "Trade, Tastes, and Nutrition in India." *The American Economic Review*, 103(5): 1629–1663.

Atkin, David. 2016. "The Caloric Costs of Culture: Evidence from Indian Migrants." *American Economic Review*, 106(October): 47.

Baier, Scott L., Jeffrey H. Bergstrand, and Michael Feng. 2014. "Economic integration agreements and the margins of international trade." *Journal of International Economics*, 93(2): 339–350.

Baldwin, Richard, and James Harrigan. 2011. "Zeros, Quality, and Space: Trade Theory and Trade Evidence." *American Economic Journal: Microeconomics*, 3(2): 60–88.

Becker, Anke, Benjamin Enke, and Armin Falk. 2016. "Ancient Origins of the Global Variation in Economic Preferences."

Bergstrand, Jeffrey H., Mario Larch, and Yoto V. Yotov. 2015. "Economic integration agreements, border effects, and distance elasticities in the gravity equation." *European Economic Review*, 78: 307–327.

Blum, Bernardo S., and Avi Goldfarb. 2006. "Does the internet defy the law of gravity?" *Journal of International Economics*, 70(2): 384–405.

Broda, Christian, and David E. Weinstein. 2004. "Variety Growth and World Welfare." *American Economic Review (Papers and Proceedings)*, 94(2): 139–144.

Bronnenberg, Bart J., Jean-Pierre H. Dubé, and Matthew Gentzkow. 2012. "The Evolution of Brand Preferences: Evidence from Consumer Migration." *The American Economic Review*, 102(6): 2472–2508.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. "Robust Inference With Multiway Clustering." *Journal of Business & Economic Statistics*, 29(2): 238–249.

Cann, Howard M., Claudia de Toma, Lucien Cazes, and Others. 2002. "A Human Genome Diversity Cell Line Panel." *Science*, 296(5566): 261–262.

Cavalli-Sforza, L. Luca, Paola Menozzi, and Alberto Piazza. 1994. "The History and Geography of Human Genes." *Princeton University Press*.

Comin, Diego, and Bart Hobijn. 2008. "An Exploration of Technology Diffusion." *The American Economic Review*, 100(1997): 1–49.

Comin, Diego, William Easterly, and Erick Gong. 2010. "Was the Wealth of Nations Determined in 1000 BC?" *American Economic Journal: Macroeconomics*, 2(3): 65–97.

Cyrus, Teresa L. 2012. "Cultural Distance and Bilateral Trade." *Global Economy Journal*, 12(4): 1–23.

Dawkins, Richard. 2004. *Ancestor's Tale: A Pilgrimage to the Dawn of Evolution.* Boston, MA:Houghton Mifflin.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Romain Wacziarg. 2015. "Culture, Ethnicity and Diversity." *mimeo*.

Desmet, Klaus, Michel Le Breton, Ignacio Ortuño-Ortín, and Shlomo Weber. 2011. "The stability and breakup of nations: A quantitative analysis." *Journal of Economic Growth*, 16(3): 183–213.

Diamond, Jared. 1997. *Guns, Germs, and Steel: The Fates of Human Societies.* New York and London:Norton.

Disdier, Anne-Célia, and Keith Head. 2008. "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *Review of Economics and Statistics*, 90(1): 37–48.

Dubois, Pierre, Rachel Griffith, and Aviv Nevo. 2014. "Do prices and attributes explain international differences in food purchases." *American Economic Review*, 104(3): 832–867.

Eakins, B.W., and G.F. Sharman. 2010. "Volumes of the World's Oceans from ETOPO1." NOAA National Geophysical Data Center, Boulder, CO.

Economides, Nicholas. 1984. "Do increases in preference similarity (across countries) induce increases in trade? An affirmative example." *Journal of International Economics*, 17(3-4): 375–381.

Egger, Peter H., and Filip Tarlea. 2015. "Multi-way clustering estimation of standard errors in gravity models." *Economics Letters*, 134: 144–147.

Fally, Thibault. 2015. "Structural gravity and fixed effects." *Journal of International Economics*, 97(1): 76–85.

Feenstra, Robert C., Robert E. Lipsey, Haiyan Deng, Alyson C. Ma, and Hengyong Mo. 2005. "World Trade Flows: 1962-2000." *NBER Working Paper #11040*.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review*, 105(10): 3150–3182.

Felbermayr, Gabriel J., and Farid Toubal. 2010. "Cultural proximity and trade." *European Economic Review*, 54(2): 279–293.

Ferreira, Fernando, and Joel Waldfogel. 2013. "Pop internationalism: Has half a century of world music trade displaced local culture?" *Economic Journal*, 123(569): 634–664.

Feyrer, James. 2009. "Distance, Trade, and Income - The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment." *NBER Working Paper #15557*.

Foellmi, Reto, Christian Hepenstrick, and Josef Zweimüller. 2012. "The Threat of Parallel Imports and the Extensive Margin of International Trade." *CEPR Discussion Paper No. 7939*.

Frazer, Garth, and Johannes Van Biesebroeck. 2010. "Trade Growth under the African Growth and Opportunity Act." *The Review of Economics and Statistics*, 92(1): 128–144.

Freedom House. 2016. "Freedom House." *Freedom in the World 2016: Anxious Dictators, Wavering Democracies: Global Freedom under Pressure.*

Giuliano, Paola, Antonio Spilimbergo, and Giovanni Tonon. 2014. "Genetic distance, transportation costs, and trade." *Journal of Economic Geography*, 14(1): 179–198.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2009. "Cultural Biases in Economic Exchange?" *The Quarterly Journal of Economics*, 124(3): 1095–1131.

Head, Keith, and Thierry Mayer. 2013. "What separates us? Sources of resistance to globalization." *Canadian Journal of Economics*, 46(4): 1196–1231.

Head, Keith, and Thierry Mayer. 2014. *Gravity Equations: Workhorse,Toolkit, and Cookbook.* Vol. 4, Amsterdam:Elsevier B.V.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein. 2008. "Estimating Trade Flows: Trading Partner and Trading Volumes." *The Quarterly Journal of Economics*, 123(2): 441–487.

Huberman, Michael, Christopher M. Meissner, and Kim Oosterlinck. 2015. "Technology and Geography in the Second Industrial Revolution: New Evidence from the Margins of Trade." *mimeo.*

Hummels, David. 2007. "Transportation Costs and International Trade in the Second Era of Globalization." *Journal of Economic Perspectives*, 21(3): 131–154.

Hummels, David, and Peter J. Klenow. 2005. "The Variety and Quality of a Nation's Exports." *The American Economic Review*, 95(3): 704–723.

Krugman, Paul. 1979. "Increasing Returns, Monopolistic Competition, and International Trade." *Journal of International Economics*, 9(4): 469–479.

Linder, Staffan Burenstam. 1961. *An essay on trade and transformation.* Uppsala:Almqvist and Wiksell.

Mansfield, Edward D., Helen V. Milner, and B. Peter Rosendorff. 2000. "Free to Trade: Democracies, Autocracies, and International Trade." *The American Political Science Review*, 94(2): 305–321.

Marshall, Monty G., Ted Robert Gurr, and Keith Jaggers. 2014. "POLITY IV PROJECT Dataset Users' Manual."

Mayer, Thierry, and Soledad Zignago. 2011. "Notes on CEPII's Distance Measures: the GeoDist Database." *mimeo, cepii*, 2011(25): pp. 1–12.

Nunn, Nathan, and Diego Puga. 2012. "Ruggedness: The Blessing of Bad Geography in Africa." *The Review of Economics and Statistics*, 94(1): 20–36.

Pemberton, Trevor J., Michael DeGiorgio, and Noah A. Rosenberg. 2013. "Population structure in a comprehensive genomic data set on human microsatellite variation." *G3 (Bethesda)*, 3(5): 903–919.

Rauch, James E. 1999. "Networks versus markets in international trade." *Journal of International Economics*, 48(1): 7–35.

Rauch, James E., and Vitor Trindade. 2002. "Ethnic Chinese Networks in International Trade." *Review of Economics and Statistics*, 84(1): 116–130.

Santos Silva, J. M. C., and Silvana Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics*, 88(4): 641–658.

Santos Silva, J. M. C., and Silvana Tenreyro. 2011. "poisson: Some convergence issues." *Stata Journal*, 10(3): 207–212.

Sapienza, Paola, Luigi Zingales, and Luigi Guiso. 2006. "Does Culture Affect Economic Outcomes?" *The Journal of Economic Perspectives*, 20(2): 23–48.

Simonovska, Ina, and Michael E. Waugh. 2014. "The elasticity of trade: Estimates and evidence." *Journal of International Economics*, 92(1): 34–50.

Spolaore, Enrico, and Romain Wacziarg. 2009. "The Diffusion of Development." *Quarterly Journal of Economics*, 124(2): 469–529.

Spolaore, Enrico, and Romain Wacziarg. 2013*a*. "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature*, 51(2): 325–369.

Spolaore, Enrico, and Romain Wacziarg. 2013*b*. "Long-Term Barriers to Economic Development." *Handbook of Economic Growth, Vol. 3*, forthcoming.

Spolaore, Enrico, and Romain Wacziarg. 2015. "Ancestry, Language and Culture." In *The Palgrave Handbook of Economics and Language.* , ed. Victor Ginsburgh and Shlomo Weber, Chapter 7. London:Palgrave Macmillan.

Spolaore, Enrico, and Romain Wacziarg. 2016. "Ancestry and Development: New Evidence." *mimeo.*

Strahler, Alan H., and Arthur N. Strahler. 1992. *Modern Physical Geography.* Hoboken NJ:John Wiley & Sons.

Tinbergen, Jan. 1962. *The World Economy. Suggestions for an International Economic Policy.* New York:Twentieth Century Fund.

Trefler, Daniel. 1995. "The Case of the Missing Trade and Other Mysteries." *American Economic Review*, 85(5): 1029–1046.

Yotov, Yoto V. 2012. "A simple solution to the distance puzzle in international trade." *Economics Letters*, 117(3): 794–798.

Figures and Tables

Figure 1: Geographic and Genetic Distance
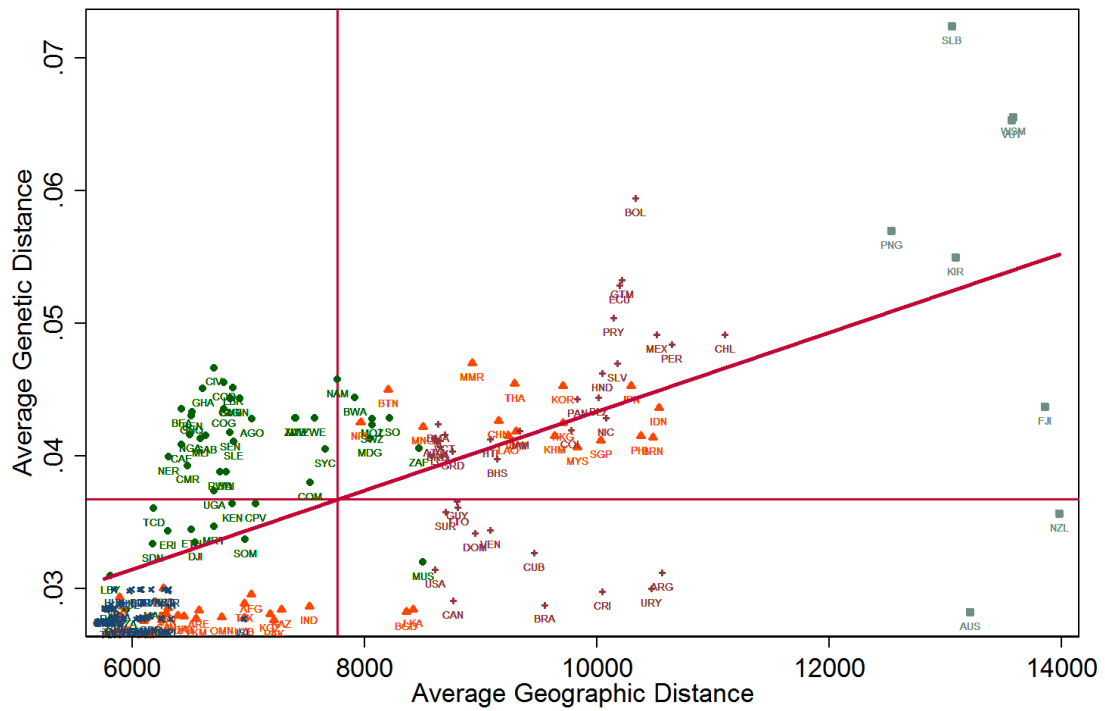
Note: The figure shows each country's average genetic distance to its (potential) partners on the vertical axis plotted against each country's average geographic distance to all its (potential) trading partners on the horizontal axis. Countries are colored according to their continent. The horizontal and vertical lines show the average for the genetic and geographic distance, respectively. We obtain a t-value of 7.25 in a linear regression with clustered standard errors.

Figure 2: Illustration of Channels
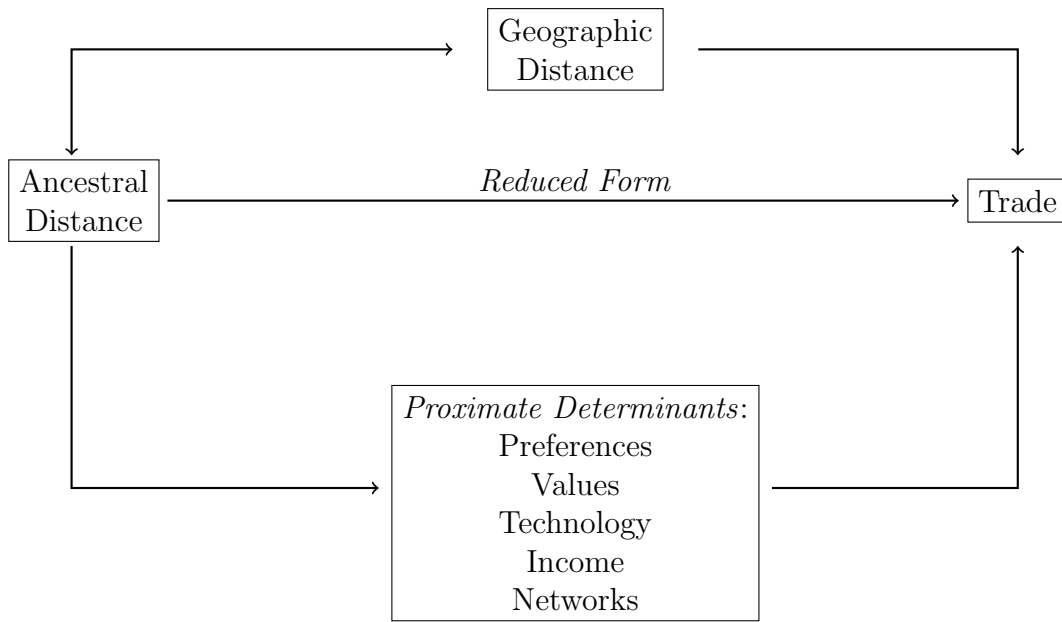
Note: The figure illustrates the relationship between ancestral distance and bilateral trade flows. In Section 3, we first estimate the reduced form. Subsequently, in Section 4, we examine the relationship between ancestral distance and proximate determinants of international trade.

Figure 3: Genetic Distance and Trade Residuals
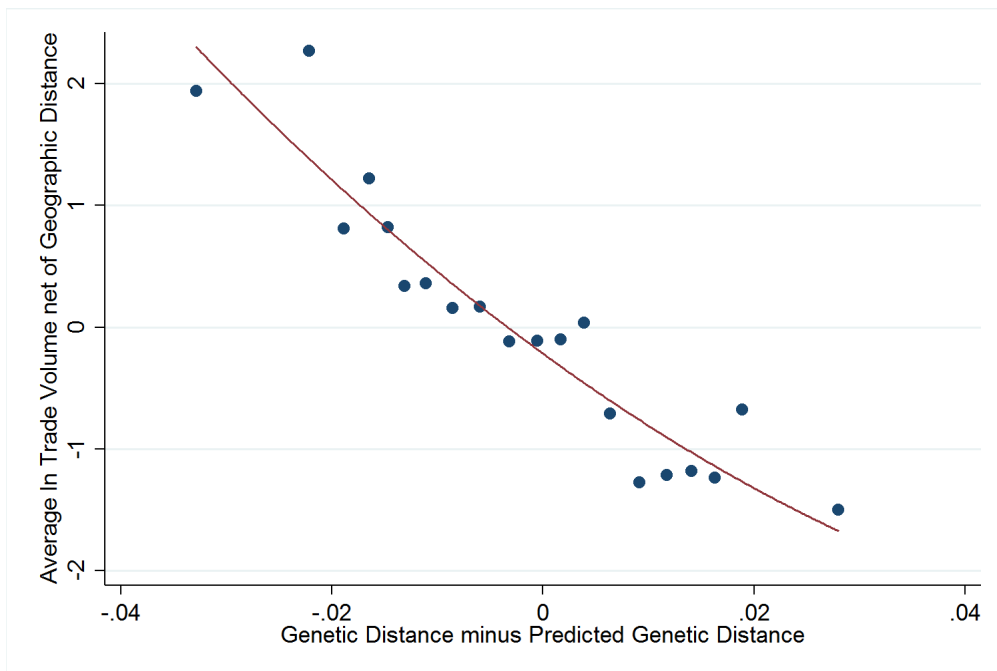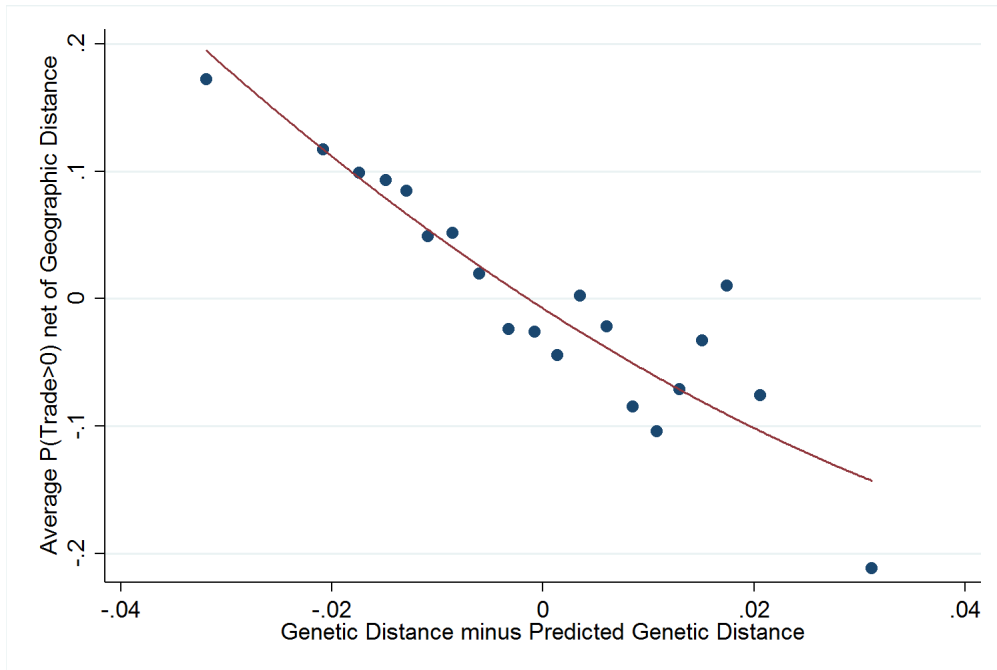
Note: The figures plot on the x-axes the difference between genetic distance and predicted genetic distance based on geographic distance. On the y-axes we see the residuals of trade regressed on geographic distance, with trade defined as log of Trade Value in panel a) and trade index in panel b).

34

Figure 4: Genetic Distance and Proximate Determinants of Trade

(a) Technology

(b) Preferences

(c) Values

(d) Networks
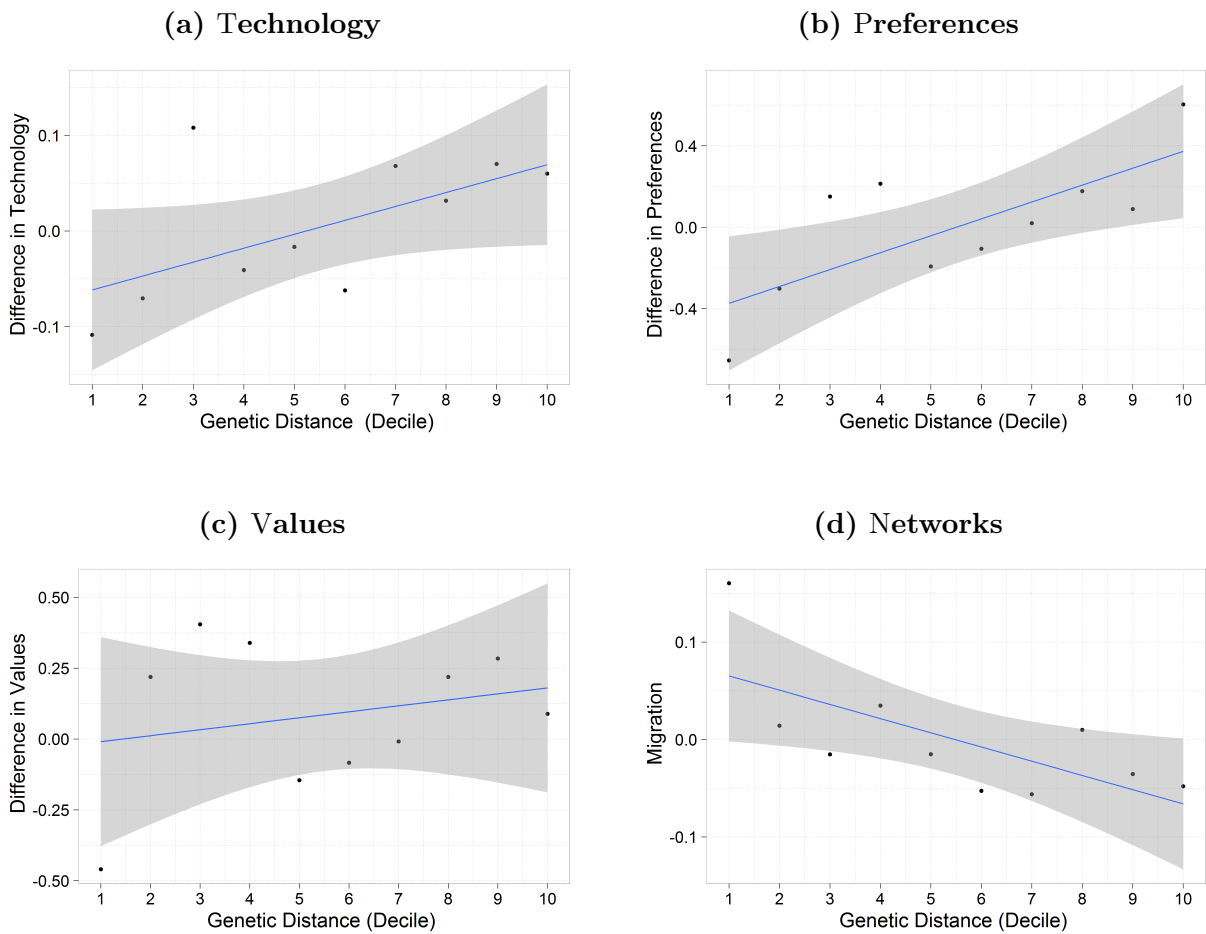
Note: The figure shows the relationship between genetic distance (horizontal axis) and four important channels (vertical axis). Each dot reflects the mean of the dependent variable for a decile of genetic distance. Plot (a) shows the relationship between genetic distance and the average difference in technology levels for a country pair based on Comin and Hobijn (2008). Plot (b) shows the relationship between genetic distance and a measure that captures differences in preferences on the vertical axis. Plot (c) shows the relationship between genetic distance axis and bilateral differences in answers to questions in the World Values Survey (WVS) on the vertical axis. Plot (d) shows the relationship between genetic distance axis and the average difference in migration stocks. All variables are expressed in standard scores. The solid lines indicate a linear fit with a corresponding 95% confidence interval.

35

Table 1: Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Part I: Country-Level Variables:					
GDP (mil. 2005 USD)	268,538.35	964,605.68	93.75	11,158,100	172
GDP per capita	10,674.73	14,071.72	191.62	101,248.76	172
Population (mil.)	34.77	128.82	0.05	1,262.65	172
Landlocked	0.22	0.41	0	1	172
Island	0.21	0.41	0	1	172
Ruggedness of Terrain	1.33	1.25	0	6.74	172
Latitude	19.02	24.3	-41.81	64.99	172
Longitude	16.95	62.64	-172.17	171.48	172
Avg. Distance to Nearest Coast (km)	324.86	417.47	0.05	2206.17	172
Within 100km of Ice-Free Coast (%)	44.84	40.12	0	100	172
Corruption Index (in 2015)	43.1	20.13	8	91	155
FH Political Rights	3.49	2.25	1	7	169
FH Civil Liberty	3.61	1.81	1	7	169
Part II: Bilateral Variables:					
Geodesic Distance (biggest city, km)	7769.03	4404.82	10.48	19904.45	29,412
Difference in Latitude	27.74	20.27	0	106.8	29,412
Difference in Longitude	69.68	54.7	0.02	343.65	29,412
Absolute Climate Difference	1.67	0.5	0	2	22,350
Difference in Share of Tropic	0.42	0.41	0	1	22,350
Difference in Average Elevation (m)	564.22	575.36	0	3176.75	22,350
Common Border	0.02	0.13	0	1	29,412
Access to Same Sea	0.16	0.37	0	1	29,412
Same Continent	0.23	0.42	0	1	29,412
Different Continent	1.99	1.54	0	5	29,412
Landlocked (none, one, both)	0.43	0.58	0	2	29,412
Island (none, one, both)	0.42	0.57	0	2	29,412
Common Off. Language	0.16	0.36	0	1	29,412
Common Currency	0.02	0.15	0	1	29,412
Colonial Ties	0.01	0.11	0	1	29,412
Same Country in Past	0.01	0.09	0	1	29,412
Economic Integration	0.36	0.94	0	6	29,412
Both in EU	0.01	0.08	0	1	29,412
Both in NAFTA	0	0.01	0	1	29,412
Both in EFTA	0	0.01	0	1	29,412
Both in AFTA	0	0.06	0	1	29,412
Both in Mercusor	0	0.03	0	1	29,412
Both Democratic	0	0.07	0	1	29,412
Genetic Distance (weighted, old)	0.11	0.07	0	0.35	29,412
Genetic Distance (dominant, old)	0.12	0.08	0	0.34	29,412
Genetic Distance (weighted, new)	0.04	0.02	0	0.09	29,412
Genetic Distance (dominant, new)	0.04	0.02	0	0.11	29,412
Linguistic Distance (weighted)	0.97	0.1	0	1	21,852
Religious Distance (weighted)	0.85	0.15	0.09	1	22,144
Part III: Trade Variables:					
Trade Value (0-Digit)	200.53	2,758.71	0	241,590.89	29,412
Positive Trade Flow (0-Digit)	0.69	0.46	0	1	29,412
Trade Value (2-Digit)	13.86	219.36	0	56570.2	2,029,428
Positive Trade Flow (2-Digit)	0.21	0.41	0	1	2,029,428
Trade Value (4-Digit)	2.90	65.67	0	34,712.82	23,000,184
Positive Trade Flow (4-Digit)	0.09	0.29	0	1	23,000,184

Note: The table shows descriptive statistics for each variable used in the empirical analysis. The data is from the year 2000. Note that we have data on 172 countries. Hence, we have 172 * 171 = 29,412 country pairs and 29,412 * 69 commodities = 2,029,428 observations for the 2-digit commodity code data set.

Table 2: Main Regression Results with Aggregate Data

	Pr(Trade>0)			Trade Value		
Mean of dep. variable	0.686			200.527		
	(1)	(2)	(3)	(4)	(5)	(6)
log Genetic Distance	-0.201***		-0.065***	-0.608***		-0.064***
	(0.008)		(0.008)	(0.036)		(0.024)
log Geodesic Distance		-0.176***	-0.123***		-0.456***	-0.493***
		(0.012)	(0.014)		(0.062)	(0.064)
Difference log GDP p.c.		-0.036**	-0.033**		0.247***	0.249***
		(0.014)	(0.014)		(0.062)	(0.061)
Common Border		-0.047	-0.015		0.546***	0.472***
		(0.037)	(0.038)		(0.084)	(0.083)
Model	Genetic Dist.	HMR	HMR + Gen. Dist.	Genetic Dist.	HMR	HMR + Gen. Dist.
Origin FE	yes	yes	yes	yes	yes	yes
Destination FE	yes	yes	yes	yes	yes	yes
Observations	25,506	18,931	18,887	29,358	22,350	22,296
R-squared	0.477	0.504	0.507	0.609	0.949	0.951

Note: The first three columns of the table shows the result of Probit regressions using the extensive margin of trade as dependent variable. All coefficients are marginal effects at means. In columns (4) to (6), the table shows the result of PPML regressions using the intensive margin of trade (trade volume divided by 1 million) as dependent variable. The data comprises aggregate bilateral trade flows from the year 2000. Additional control variables follow Helpman, Melitz and Rubinstein (2008) (HMR) and include differences in GDP per capita, being an island, being landlocked, access to the same sea, difference in latitude, difference in longitude, difference in climate, difference in tropics, difference in elevation, common currency, colonial relationship, same country in the past, economic integration index, common memeber ship in the EU, EFTA, AFTA, and MERCUSOR, as well as both democratic. Standard errors are shown in parentheses and clustered at the country-pair level. Significance at the 10% level is indicated by *, at the 5% level by **, and at the 1% level by ***.

Table 3: Robustness Tests

I. Extensive Margin

	Pr(Trade>0)	
Mean of dep. variable	0.686	0.788	0.733	0.686	0.686
	(1)	(2)	(3)	(4)	(5)
log Genetic Distance	-0.065***	-0.018**	-0.059***		
	(0.008)	(0.007)	(0.008)		
log Geodesic Distance	-0.123***	-0.022***	-0.114***	-0.147***	-0.153***
	(0.014)	(0.008)	(0.013)	(0.016)	(0.013)
log Genetic Distance (dominant)				-0.060***	
				(0.007)	
log Genetic Distance (Cavalli-Sforza)					-0.018***
					(0.004)
Difference log GDP p.c.	-0.033**	0.006	-0.060***	-0.036**	-0.032**
	(0.014)	(0.008)	(0.013)	(0.016)	(0.014)
Common Border	-0.015	-0.027	-0.066**	0.001	-0.025
	(0.038)	(0.020)	(0.031)	(0.059)	(0.042)
Sample	full	Africa	2005	full	full
	2000	out	2000	2000	2000
Origin FE	yes	yes	yes	yes	yes
Destination FE	yes	yes	yes	yes	yes
Observations	18,887	7,814	16,419	17,915	18,873
R-squared	0.507	0.531	0.534	0.500	0.505

II. Intensive Margin

	\multicolumn{5}{c}{Trade Value}		# Goods
Mean of dep. variable	0.200	0.190	0.200	0.200	0.329	10.391
	(1)	(2)	(3)	(4)	(5)	(6)
log Genetic Distance	0.033***	0.024***	0.032***			-18.771***
	(0.005)	(0.004)	(0.005)			(1.505)
log Geodesic Distance	-0.174***	-0.064***	-0.135***	-0.166***	-0.170***	-32.121***
	(0.011)	(0.010)	(0.010)	(0.013)	(0.011)	(2.717)
Genetic Distance (dominant, new)				-0.005		
				(0.008)		
Genetic Distance (weighted, old)					0.022***	
					(0.005)	
Difference log GDP p.c.	-0.017***	-0.003	-0.010**	-0.004	-0.019***	
	(0.006)	(0.005)	(0.004)	(0.007)	(0.006)	
Common Border	0.005	0.010	-0.014	0.034	0.003	
	(0.023)	(0.019)	(0.021)	(0.032)	(0.022)	
Sample	full	Africa	2005	full	full	full
	2000	out	2000	2000	2000	4-Digit
						2000
Origin FE	yes	yes	yes	yes	yes	
Destination FE	yes	yes	yes	yes	yes	
Observations	22,296	10,884	21,408	20,918	22,290	15,841,756
R-squared	0.362	0.333	0.363	0.364	0.362	0.778

Note: The top part of the table shows the result of six Probit regressions using the extensive margin of trade as dependent variable. All coefficients are marginal effects at means. In the lower part, the table shows the result of six PPML regressions using the intensive margin of trade (trade volume divided by 1 billion) as dependent variable. Additional control variables include differences in GDP per capita, being an island, being landlocked, access to the same sea, difference in latitude, difference in longitude, difference in climate, difference in tropics, difference in elevation, common currency, colonial relationship, same country in the past, economic integration index, common memeber ship in the EU, EFTA, AFTA, and MERCUSOR, as well as both democratic. Standard errors are shown in parentheses and clustered at the country-pair level. Significance at the 10% level is indicated by *, at the 5% level by **, and at the 1% level by ***.

Table 4: Number of Commodities Traded

	Number of Commodities Traded			
Mean of dep. variable	58.449	3.365	46.441	14.660
Std. of dep. variable	131.989	8.749	90.193	34.321
	(1)	(2)	(3)	(4)
log Genetic Distance	-33.660***	-1.646***	-21.680***	-7.417***
	(1.899)	(0.133)	(1.154)	(0.483)
log Geodesic Distance	-55.622***	-3.017***	-35.433***	-12.696***
	(3.822)	(0.262)	(2.340)	(0.958)
Difference log GDP p.c.	10.522***	0.261***	7.135***	2.243***
	(0.389)	(0.022)	(0.250)	(0.091)
Common Border	45.326***	7.134***	17.809***	16.889***
	(11.224)	(0.856)	(6.699)	(2.887)
Commodities	all	Homo-genous	Differ-entiated	Ref.-Price
Origin FE	yes	yes	yes	yes
Destination FE	yes	yes	yes	yes
Commodity FE	yes	yes	yes	yes
Observations	16,741,056	1,648,416	8,841,504	4,538,496
R-squared	0.366	0.390	0.352	0.368

Note: The table shows the result of four OLS regressions using the number of bilaterally traded commodities as dependent variable. The data comprises bilateral trade flows at the 4-digit level from the year 2000. The sample differs based on commodities as defined by Rauch (1999): homogeneous, differentiated, and reference-priced. Additional control variables include differences in GDP per capita, being an island, being landlocked, access to the same sea, difference in latitude, difference in longitude, difference in climate, difference in tropics, difference in elevation, common currency, colonial relationship, same country in the past, economic integration index, common memeber ship in the EU, EFTA, AFTA, and MERCUSOR, as well as both democratic. Standard errors are shown in parentheses and clustered at the country-pair level. Significance at the 10% level is indicated by *, at the 5% level by **, and at the 1% level by ***.

Table 5: Proximate Determinants of Trade

	Pr(Trade>0)					Trade Value				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
log Genetic Distance	-0.051***	-0.074***	-0.047***	-0.088***	-0.047***	-0.068	-0.069	-0.069	-0.071	-0.061
	(0.006)	(0.010)	(0.006)	(0.005)	(0.005)	(0.046)	(0.044)	(0.046)	(0.061)	(0.043)
log Geodesic Distance	-0.132***	-0.092***	-0.130***	-0.071***	-0.121***	-0.906***	-0.925***	-0.907***	-0.925***	-0.901***
	(0.005)	(0.008)	(0.005)	(0.011)	(0.005)	(0.035)	(0.034)	(0.035)	(0.042)	(0.035)
Diff. in Technology		-0.159***					2.447***			
		(0.059)					(0.161)			
Diff. in Preferences			-0.102***					0.781		
			(0.015)					(0.822)		
Diff. in Values				-0.014**					0.099*	
				(0.007)					(0.053)	
Networks					0.041***					0.595***
					(0.005)					(0.141)
Origin FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Destination FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	25,506	13,142	25,506	4,162	25,506	29,358	16,732	29,358	7,806	29,358
R-squared	0.513	0.494	0.514	0.431	0.515	0.890	0.894	0.890	0.905	0.891

Note: Columns (1) to (5) of the table shows the result of Probit regressions using the extensive margin of trade as dependent variable. All coefficients are marginal effects at means. In columns (6) to (10), the table shows the result of PPML regressions using the intensive margin of trade (trade volume divided by 1 million) as dependent variable. Columns (2) and (6) use a measure of differences in technology, columns (3) and (7) include a measure of differences in preferences, columns (4) and (8) use a measure of differences in values, columns (5) and (10) include a measure of migration networks. The data comprises aggregate bilateral trade flows from the year 2000. Standard errors are shown in parentheses and clustered at the country-pair level. Significance at the 10% level is indicated by *, at the 5% level by **, and at the 1% level by ***.

## Table 6: Effect Heterogeneity across Types of Commodities

	log Number of Goods Traded			
Mean of dep. variable	3.199	1.394	2.997	2.225
	(1)	(2)	(3)	(4)
log Genetic Distance	-0.398***	-0.192***	-0.370***	-0.288***
	(0.019)	(0.015)	(0.018)	(0.017)
log Geodesic Distance	-0.825***	-0.346***	-0.746***	-0.572***
	(0.042)	(0.032)	(0.041)	(0.038)
Difference log GDP p.c.	0.183***	0.058***	0.191***	0.152***
	(0.007)	(0.005)	(0.007)	(0.006)
Common Border	0.342***	0.537***	0.242***	0.363***
	(0.095)	(0.063)	(0.091)	(0.082)
Commodities	all	Homo-geneous	Differ-entiated	Ref.-Price
Observations	12,307,898	748,440	6,124,790	2,619,684
R-squared	0.297	0.284	0.284	0.293

Note: The table shows the result of four separate OLS regressions using the number of traded 4-digit commodities as dependent variable. The data comprises bilateral trade flows at the 4-digit level from the year 2000. The samples differ based on commodities as defined by Rauch (1999): homogeneous, differentiated, as well as reference-priced. Additional control variables include differences in GDP per capita, being an island, being landlocked, access to the same sea, difference in latitude, difference in longitude, difference in climate, difference in tropics, difference in elevation, common currency, colonial relationship, same country in the past, economic integration index, common memebership in the EU, EFTA, AFTA, and MERCUSOR, as well as both democratic. Standard errors are shown in parentheses and clustered at the country-pair level. Significance at the 10% level is indicated by *, at the 5% level by **, and at the 1% level by ***.

Appendix

Appendix A: Additional Figures and Tables

Figure A.1: Genetic Distance in a World Map

United States
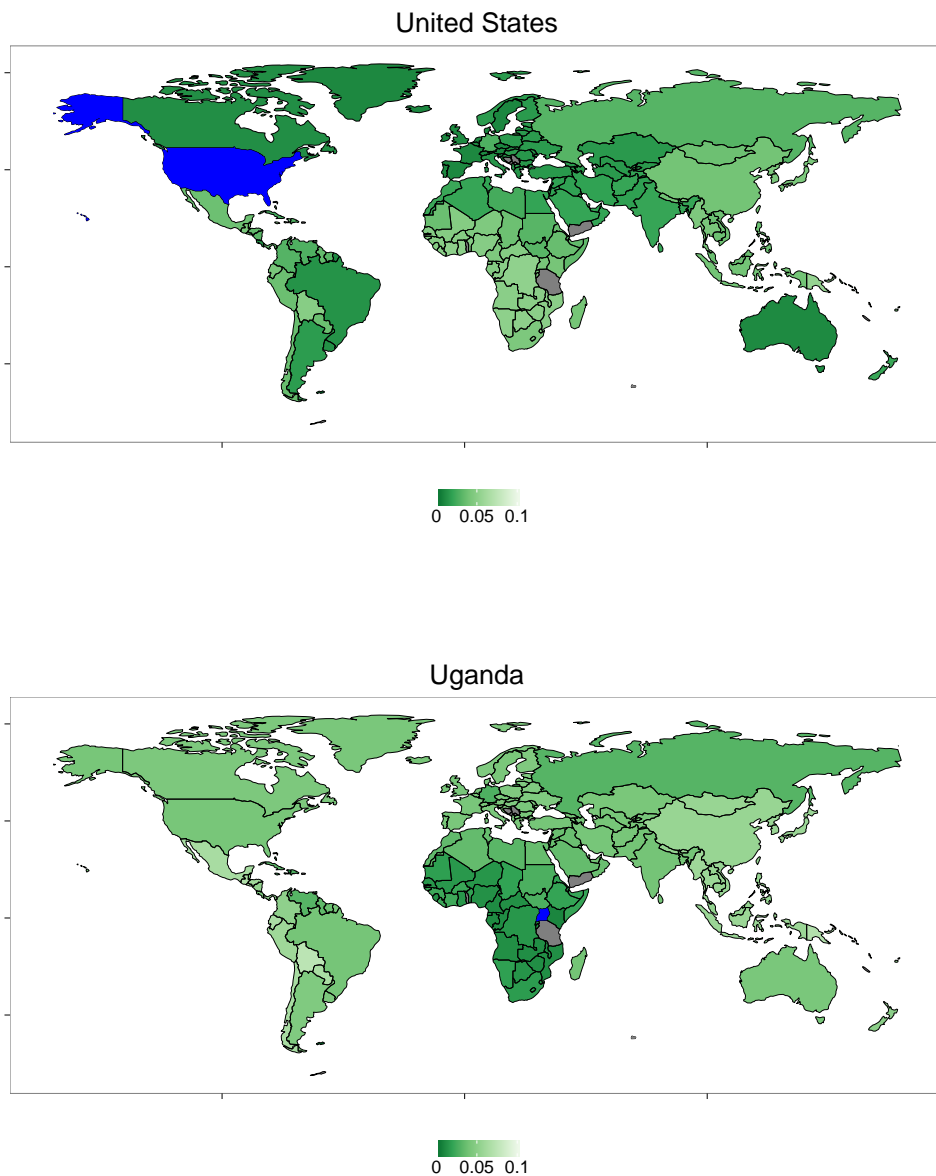
Uganda

Note: The figures illustrate the bilateral genetic distances of the United States (top figure) as well as Uganda (lower figure) to all other countries. We use the weighted genetic distance. There is no data for countries shaded gray.

Figure A.2: Geographic and Genetic Distance (Quintiles)
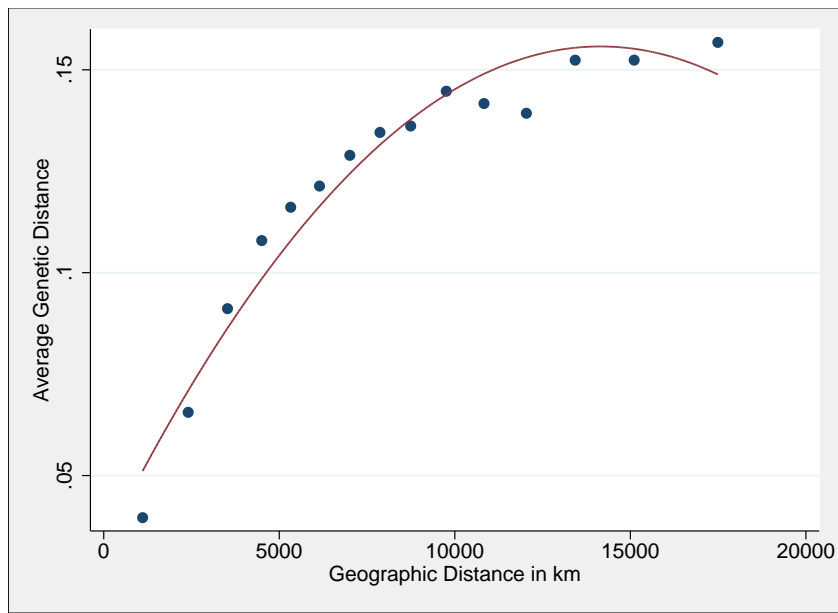
Note: The figure shows quintiles of each country's average genetic distance to its (potential) partners on the y-axis plotted against quintiles of each country's average geographic distance to all its (potential) trading partners on the y-axis.

Table A.1: Countries in the Sample

Afghanistan	Cape Verde	Gabon	Korea Rep.	Nigeria	Sri Lanka
Albania	Cen. African Rep.	Gambia	Kuwait	Norway	St Kitts Nevis
Algeria	Chad	Georgia	Kyrgyzstan	Oman	St Vincent Grenadine
Angola	Chile	Germany	Laos	Pakistan	Sudan
Antigua & Barbuda	China	Ghana	Latvia	Panama	Suriname
Argentina	Colombia	Greece	Lebanon	Papua N. Guinea	Swaziland
Armenia	Comoros	Grenada	Lesotho	Paraguay	Sweden
Australia	Congo	Guatemala	Liberia	Peru	Switzerland
Austria	Congo, Dem. Rep.	Guinea	Libya	Philippines	Syria
Azerbaijan	Costa Rica	Guinea-Bissau	Lithuania	Poland	Tajikistan
Bahamas	Cote dIvoire	Guyana	Macedonia	Portugal	Thailand
Bahrain	Croatia	Haiti	Madagascar	Qatar	Trinidad and Tobago
Bangladesh	Cuba	Honduras	Malawi	Rep. of Moldova	Tunisia
Barbados	Cyprus	Hong Kong	Malaysia	Romania	Turkey
Belarus	Czech Rep.	Hungary	Mali	Russia	Turkmenistan
Belgium-Lux.	Denmark	Iceland	Malta	Rwanda	Uganda
Belize	Djibouti	India	Mauritania	Saint Lucia	Ukraine
Benin	Dominica	Indonesia	Mauritius	Samoa	United Arab Emirates
Bhutan	Dominican Rep.	Iran	Mexico	Saudi Arabia	United Kingdom
Bolivia	Ecuador	Iraq	Mongolia	Senegal	United States
Botswana	Egypt	Ireland	Morocco	Seychelles	Uruguay
Brazil	El Salvador	Israel	Mozambique	Sierra Leone	Uzbekistan
Brunei	Equ. Guinea	Italy	Myanmar	Singapore	Vanuatu
Bulgaria	Eritrea	Jamaica	Namibia	Slovakia	Venezuela
Burkina Faso	Estonia	Japan	Nepal	Slovenia	Vietnam
Burundi	Ethiopia	Jordan	Netherlands	Solomon Isds	Zambia
Cambodia	Fiji	Kazakhstan	New Zealand	Somalia	Zimbabwe
Cameroon	Finland	Kenya	Nicaragua	South Africa	
Canada	France	Kiribati	Niger	Spain	

Note: The table shows the set of 172 countries for which we have data on bilateral trade flows, weighted genetic distance as well as country-specific information about GDP, population size, and a large set of geographic variables.

Table A.2: Countries with Missing Cultural Distance Data

(A) No Genetic Distance in Spolaore and Wacziarg (2016): 0.701% of Trade Volume

Andorra	China, Macao SAR	New Caledonia	Togo
Aruba	Cook Isds	Palau	Tonga
Bermuda	French Polynesia	Sao Tome and Principe	Turks and Caicos Isds
Bosnia Herzegovina	Greenland	Serbia and Montenegro	Yemen
Br. Virgin Isds	Maldives	Tanzania	
Cayman Isds	Marshall Isds	Timor-Leste	

(B) No Linguistic Distance in Spolaore and Wacziarg (2015): 7.032% of Trade Volume

Andorra	China, Macao SAR	Malta	St Kitts Nevis
Antigua and Barbuda	Comoros	Marshall Isds	St Vincent Grenadine
Aruba	Cook Isds	New Caledonia	Suriname
Bahamas	Dominica	Palau	Tanzania
Barbados	Equatorial Guinea	Papua New Guinea	Timor-Leste
Belize	French Polynesia	Qatar	Togo
Bermuda	Greenland	Saint Lucia	Tonga
Bosnia Herzegovina	Grenada	Samoa	Turks and Caicos Isds
Br. Virgin Isds	Hong Kong	Sao Tome and Principe	Vanuatu
Brunei	Iceland	Serbia and Montenegro	Yemen
Cape Verde	Kiribati	Seychelles	
Cayman Isds	Maldives	Solomon Isds	

(C) No Religious Distance in Spolaore and Wacziarg (2015): 6.997% of Trade Volume

Andorra	China, Macao SAR	Malta	St Vincent Grenadine
Antigua and Barbuda	Comoros	Marshall Isds	Suriname
Aruba	Cook Isds	New Caledonia	Tanzania
Bahamas	Dominica	Palau	Timor-Leste
Barbados	Equatorial Guinea	Qatar	Togo
Belize	French Polynesia	Saint Lucia	Tonga
Bermuda	Greenland	Samoa	Turks and Caicos Isds
Bosnia Herzegovina	Grenada	Sao Tome and Principe	Vanuatu
Br. Virgin Isds	Hong Kong	Serbia and Montenegro	Yemen
Brunei	Iceland	Seychelles	
Cape Verde	Kiribati	Solomon Isds	
Cayman Isds	Maldives	St Kitts Nevis	

Note: The table lists all countries with missing genetic, linguistic, or religious data separated in three different categories. Panel (A) reports all countries that have no weighted genetic distance data in Spolaore and Wacziarg (2016), Panel (B) lists all countries with missing linguistic distance data, and Panel (C) those with missing religious distance data as provided by Spolaore and Wacziarg (2015).

Table A.3: Determinants of Genetic Distance

Mean of dep. variable	log Genetic Distance			
	-3.499			
	(1)	(2)	(3)	(4)
log Geodesic Distance	0.583***	0.566***	0.522***	0.524***
	(0.028)	(0.026)	(0.027)	(0.122)
Landlocked Exporter		-0.039	-0.003	-0.134
		(0.037)	(0.041)	(0.182)
Landlocked Importer		-0.039**	-0.003	-0.140
		(0.015)	(0.012)	(0.141)
GDP p.c. Exporter		-0.000***	-0.000***	-0.000
		(0.000)	(0.000)	(0.000)
GDP p.c. Importer		-0.000***	-0.000***	-0.000
		(0.000)	(0.000)	(0.000)
Population (mil.) Exporter		-0.000	-0.000	0.014**
		(0.000)	(0.000)	(0.006)
Population (mil.) Importer		-0.000***	-0.000**	0.014***
		(0.000)	(0.000)	(0.001)
Colonial Ties		0.034	0.071	0.267
		(0.060)	(0.064)	(0.241)
Same Country in Past		0.250**	0.105	0.440*
		(0.103)	(0.099)	(0.255)
Common Currency		-0.330***	-0.324***	-0.065
		(0.100)	(0.096)	(0.126)
Latitude			-0.004***	0.013*
			(0.001)	(0.007)
Longitude			-0.000	0.021***
			(0.000)	(0.007)
Latitude			-0.004***	0.013
			(0.001)	(0.008)
Longitude			-0.000	0.021***
			(0.000)	(0.005)
Distance Nearest Coast Importer			-0.000	-0.000
			(0.000)	(0.000)
Distance Nearest Coast Exporter			-0.000	-0.000
			(0.000)	(0.000)
Access to Same Sea			0.051	-1.012***
			(0.033)	(0.224)
Island (none, one, both)			0.009	-0.168*
			(0.023)	(0.086)
Countries	all	all	all	Europe
Observations	29,358	29,358	29,358	1,228
R-squared	0.329	0.370	0.390	0.282

Note: The table shows the result of four regressions using the (weighted) genetic distanced between each country pair as dependent variable. In column (4), the sample is reduced to countries in Europe. Standard errors are shown in parentheses and clustered at the origin country. Significance at the 10% level is indicated by *, at the 5% level by **, and at the 1% level by ***.

Table A.4: The Effect of Genetic Distance on Proximate Determinants

	Differences in Technology		Differences in Preferences		Differences in Values		Networks	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log Genetic Distance (standardized)	0.048	0.325***	0.047***	0.320***	0.170***	0.166***	-0.080***	-0.001
	(0.039)	(0.025)	(0.001)	(0.006)	(0.011)	(0.011)	(0.003)	(0.003)
log Geodesic Distance	-0.032	-0.210***	0.002	0.142***	-0.049***	0.194***	-0.137***	-0.199***
	(0.048)	(0.022)	(0.002)	(0.008)	(0.016)	(0.015)	(0.004)	(0.004)
Observations	8,366	8,366	29,358	29,358	7,806	7,806	29,358	29,358
R-squared	0.002	0.569	0.974	0.073	0.029	0.559	0.115	0.440
Fixed Effects	No	Origin	No	Origin & Destination	No	Origin & Destination		Origin & Destination

Note: The table shows the effect of genetic distance on four proximate causes, namely differences in technology (columns (1) and (2)), differences in preferences (columns (3) and (4)), differences in values (columns (5) and (6)), and networks (columns (7) and (8)). The data comprises aggregate bilateral trade flows from the year 2000. Note that the estimation in columns (1) and (2) uses only one exporter-importer observation because genetic distance does not vary at the country-pair-level and the difference in technology is symmetric. Standard errors are shown in parentheses and clustered at the importer level in columns (1) and (2)and at the country-pair level for columns (3) to (8). Significance at the 10% level is indicated by *, at the 5% level by **, and at the 1% level by ***.

Table A.5: Proximate Causes of Trade without Genetic Distance

	Pr(Trade>0)				Trade Value			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Differences in Technology	-0.145**				-2.532***			
	(0.058)				(0.125)			
Differences in Preferences		0.122***				-0.202		
		(0.014)				(0.731)		
Differences in Values			-0.019***				0.079	
			(0.007)				(0.051)	
Networks				0.047***				0.576***
				(0.005)				(0.143)
Origin FE	yes	yes	yes	yes	yes	yes	yes	yes
Destination FE	yes	yes	yes	yes	yes	yes	yes	yes
Observations	13,176	25,552	4,180	25,552	16,770	29,412	7,832	29,412
R-squared	0.487	0.511	0.421	0.512	0.893	0.886	0.902	0.886

Note: The first four columns of the table shows the result of Probit regressions using the extensive margin of trade as dependent variable. All coefficients are marginal effects at means. In columns (5) to (8), the table shows the result of PPML regressions using the intensive margin of trade (trade volume divided by 1 million) as dependent variable. Columns (1) and (5) use a measure of differences in technology, columns (2) and (6) include a measure of differences in preferences, columns (3) and (7) use a measure of differences in values, columns (4) and (8) include a measure of migration networks. The data comprises aggregate bilateral trade flows from the year 2000. Standard errors are shown in parentheses and clustered at the country-pair level. Significance at the 10% level is indicated by *, at the 5% level by **, and at the 1% level by ***.

Appendix B: Description of Ocean Data Set

For our analysis, we created a variable indicating whether two countries have access to the same sea. Here, we explain how we have built this measure. First, we selected a list of ten major seas according to the classification by Eakins and Sharman (2010). The authors modified the list of world-seas published by the International Hydrographic Organization (IHO) to include only major oceans and marginal seas. This results in a sample of ten major seas, which we show as a list in the table below. Figure B.1, which is taken from Eakins and Sharman (2010), also shows the borders that apply once minor seas are merged with bigger ones.

There are two major differences with respect to the official IHO classification. First, we incorporate the Arabian Sea and the Bay of Bengali into the Indian Ocean. Second, we include the Southern Ocean south of 60°S. A revision of the IHO code in the year 2000 includes the Southern Ocean among the official sea-list. However, the revision has not yet been approved.

Table B.1: Seas as Defined in our Data

Arctic Ocean	Atlantic (North)	Atlantic (South)	Baltic Sea	Indian Ocean
Mediterranean Sea	Pacific (North)	Pacific (South)	South China Sea	Southern Ocean

Figure B.1: Illustration of Sea Borders
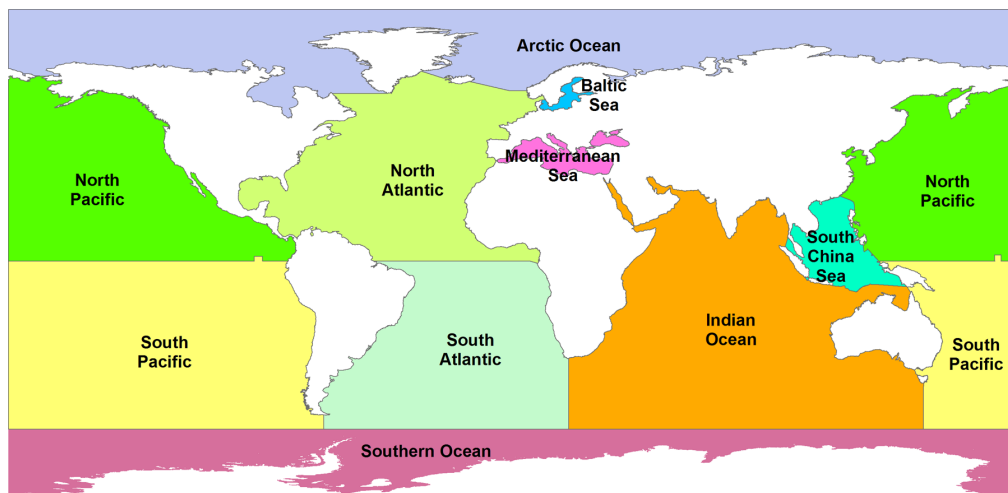
Note: The figure highlights the borders of all oceans as we use them in our data. The illustration is taken from Eakins and Sharman (2010).

For every country in the data set, we generated a dummy variable for each sea indicating whether the country has access to it. Moreover, we computed a dummy variable ('access to the same sea') that takes the value one if two countries have access to the same sea.

Appendix C: List of Variables

In this section, we describe in detail how each variable is defined and measured. Furthermore, we outline the sources of all variables.

Geographic Variables

Geodesic Distance the distance in km between origin o and destination country d (in logs). A geodesic is the shortest route between two points on the Earth's surface, a segment of a great circle.

Common Border a binary variable that equals one if the origin o and destination country d are neighboring countries sharing a common physical boundary, and zero otherwise.

Island a binary variable taking value 1 if either the origin o or the destination d country is an island and 0 otherwise.

Landlocked a binary variable taking value 1 if either the origin o or the destination d country is landlocked and 0 otherwise.

Access to the Same Sea a binary variable equal to 1 if both origin and destination countries have access to the same sea (see illustration of sea-borders above) and equal to 0 otherwise.

Difference in Latitude a numerical variable reporting the absolute value of the difference in degrees of latitude between the origin o and destination d country.

Difference in Longitude a numerical variable reporting the absolute value of the difference in degrees of longitude between the origin o and destination d country.

Difference in Climate a numerical variable reporting the sum of the absolute values of the difference between the percentage of the area of origin o and destination d country in each climate zone. Climate zones are defined according to the Koeppen-Geiger climate zones map as reported by Strahler and Strahler (1992).

Difference in Tropics a numerical variable reporting the absolute difference between the share of the origin and destination country's territory in a tropical climate. The tropical climate zone is defined according to the Koeppen-Geiger climate zones map as reported by Strahler and Strahler (1992).

Difference in Elevation A numerical variable reporting the absolute difference between the origin and the destination country's average elevation, where elevation is measured in meters above sea level.

Economic and Political Variables

Difference in GDP per capita a numerical variable reporting the log of the difference in per capita GDP between the origin o and the destination d country. Per capita GDP is measured in millions of US dollars and population is measured in millions of individuals.

Common Currency a binary variable taking value 1 if the origin o and destination d countries use the same currency and 0 otherwise.

Economic Integration Agreements

Common Membership in FTA a set of binary variables which equal one if both countries are a member of the EU, NAFTA, EFTA, AFTA, or MERCUSOR, and zero otherwise.

Democracy a binary variable that equals one if both countries score eight or higher (on a scale from -10 to 10) on the democracy index provided by Polity IV.

Common Official Language a binary variable that equals one if both countries have at least one official language in common.

Colonial Ties a binary variable that equals one if either of the two countries was the colony of the other one in the past.

Same Country in the Past a binary variable that equals one if the origin and destination countries used to be one country at some point in the past.