Universität St.Gallen

Human Barriers to International Trade

Irene Fensore, Stefan Legge, Lukas Schmid

September 2017 Discussion Paper no. 2017-12

School of Economics and Political Science,
Department of Economics

University of St. Gallen

Editor: Martina Flockerzi
University of St.Gallen
School of Economics and Political Science
Department of Economics
Müller-Friedberg-Strasse 6/8
CH-9000 St. Gallen
Phone +41 71 224 23 25
Email seps@unisg.ch
Publisher: School of Economics and Political Science
Department of Economics
University of St.Gallen
Müller-Friedberg-Strasse 6/8
CH-9000 St. Gallen
Phone +41 71 224 23 25
Electronic Publication: http://www.seps.unisg.ch

Human Barriers to International Trade[1]

Irene Fensore, Stefan Legge, Lukas Schmid

Irene Fensore:

Irene Fensore, M.A.
SIAW-HSG
University of St.Gallen
Bodanstrasse 8
CH-9000 St.Gallen
Phone +41 71 224 2345
Email irene.fensore@unisg.ch
Website www.siaw.unisg.ch

Stefan Legge:

Stefan Legge, Ph.D.
SIAW-HSG
University of St.Gallen
Bodanstrasse 8
CH-9000 St.Gallen
Phone +41 71 224 2338
Email stefan.legge@unisg.ch
Website www.siaw.unisg.ch

Lukas Schmid:

Dr. Lukas Schmid
University of Lucerne
Frohburgstrasse 3
CH-6005 Luzern
Phone +41 41 229 5850
Email lukasdavid.schmid@unilu.ch
Website www.lukasschmid.net

---

[1] For helpful comments and discussions, we are grateful to James Anderson, Gernot Doppelhofer, David Dorn, Gabriel Felbermayr, Reto Föllmi, Lars Peter Hansen, Beata Javorcik, Dennis Novy, Ulrich Schetter, Enrico Spolaore, Florian Unger, Romain Wacziarg, Josef Zweimüller as well as seminar participants at Stanford University, the University of Zurich, the University of Colorado at Boulder, the Norwegian School of Economics, the University of St.Gallen, the 2017 Annual Meeting of the American Economic Association, the 2017 Lindau Nobel Laureate Meeting, the 2017 EGIT Meeting at the ifo Institute in Munich, the 2016 SNF Sinergia Workshop in Stoos, the 2016 European Trade Study Group, and the 2016 Spanish Economic Association.

Abstract

This paper investigates whether the relatedness of populations across the world shapes international trade flows. Using data on common ancestry for 172 countries covering more than 99% of global trade, we document that country pairs with a larger ancestral distance are less likely to trade with each other (extensive margin) and, if they do trade, they trade fewer goods and smaller volumes (intensive margin). The results are robust to including a vast array of control variables capturing other sources of heterogeneity, including micro-geographic, political, linguistic, and religious differences. We discuss the role of several determinants of trade that lead to this negative relationship, namely differences in trust, values, consumption structures, political institutions, technology, as well as recent migration networks. Exploring the robustness of our findings, we use detailed census information on ancestry and show that U.S. states trade significantly more with ancestrally close countries.

Keywords

Ancestral Distance, Trade Barriers, Trade Flows

JEL Classification

F14, F15, O33

1 Introduction

Despite substantial reductions in transportation costs and significant decreases in tariffs and non-tariff trade barriers since World War II, well-established estimates of bilateral trade costs are still very large (Anderson and van Wincoop, 2004; Balistreri and Hillberry, 2006). Understanding why these trade costs continue to be sizable is a key to several puzzles in the trade literature, including the home bias puzzle (McCallum, 1995), the consumption correlations puzzle (Backus, Kehoe and Kydland, 1992), and the geographic distance puzzle (Disdier and Head, 2008). Thus far, however, the literature has pointed out that a substantial part of bilateral trade costs remains unobserved which has led Head and Mayer (2013) to describe the unobserved part of trade costs as dark costs.

This paper sheds new light on an important part of trade costs and thus a major barrier to international trade. We provide evidence that populations with common ancestors are more likely to establish trade relationships and — if they trade — they exchange larger volumes. We argue that ancestry affects trade because it has an impact on several so-called proximate determinants of trade. Individuals who share common ancestry are more likely to trust each other and mutual trust serves as a complement for incomplete contracts in international markets (Guiso, Sapienza and Zingales, 2009). In addition, countries with more closely related populations are likely to adapt technological innovations from each other (Spolaore and Wacziarg, 2013) and create similar political institutions (Alesina and Giuliano, 2015). Both leads to comparable paths of economic growth and countries with similar income levels tend to trade more with each other (Anderson and van Wincoop, 2003). In addition, ancestry shapes consumption patterns and countries with similar consumption structures are more likely to trade with each other (Linder, 1961; Economides, 1984). Furthermore, country pairs that share common ancestors are more likely to have networks that allow an efficient matching of buyers and sellers (Rauch, 1999). Finally, we argue that common ancestry leads to similar values, norms, and habits which makes countries more likely to adapt similar social, economic, and behavioral changes (Spolaore and Wacziarg, 2016*b*).

We combine the most recent data on genetic distance (Spolaore and Wacziarg, 2016*a*) as a proxy for ancestral relatedness with detailed trade data and complement this with a large set of bilateral geographic measures for 172 countries which encompass more than 99% of global trade. 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). Increasing the ancestral distance between a country pair by 10 percent, which corresponds to the difference in ancestral distance between Canada and Brazil with the United States, decreases the probability that countries establish trade relations by 0.7 to 1.35 percentage points. For the intensive margin, we find that the volume of trade is reduced by 0.7 to 6.0 percent if ancestral distance is 10 percent larger.[2] Given the large standard deviation of genetic distance, the effects are five times larger when considering an increase by one standard deviation. These findings contribute to prior research pointing out the importance of the extensive margin of trade (Hummels and Klenow, 2005; Santos Silva, Tenreyro and Wei, 2014). Notably, our estimated effect of geographic distance on the intensive margin of trade is very similar to the estimated coefficient by Yotov (2012) and decreases by 19 percent (2.5 percent) once we include genetic distance in the estimation of the on the extensive (intensive) margin of trade.

We perform several tests to explore the robustness of our results. First, we include a vast array of measures for micro-geographical distance including contiguities, access to the same sea, as well as differences in latitude and longitude. All of these variables capture differences in transportation costs and none of them alter the finding that bilateral trade flows are negatively associated with ancestral distance. The same holds if we add controls for common currencies, colonial relationships, free trade agreements, economic integration indices, as well as political

---

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

[2] As an example, the genetic distance between the United States and Slovakia is about 10 percent larger than the distance between the United States and Sweden.

systems. Second, our empirical findings are also robust to including measures of linguistic and religious distance. Finally, we account for the potential endogeneity of genetic distance by performing two tests. We use genetic distance based on population data in 1500 AD as instrumental variable and show that the impact of ancestral distance is even larger when adressing endogeneity. Furthermore, we explicitly take into account migration flows over the last 500 years, exploiting data from Putterman and Weil (2010). The estimates show that our results hold in a sample of countries with roughly the same ethnic composition as 500 years ago.

In order to explore the sensitivity of our findings to the measure for ancestral distance, we use data from the 2000 U.S. Census to derive an alternative measure of ancestry. The Census asks each survey participant to name the ancestry group with which they most closely identify. We aggregate these answers at the state level to construct the share of population with an ancestral relationship to other countries in the world. We then combine the ancestry information with state-level export and import data, generating a data set that includes all 50 U.S. states and 27 partner countries. This allows us to explore the relationship between the alternative ancestry measure and trade. Our findings provide evidence of a strong relationship between trade and ancestral relationship and gives us confidence that the relationship between ancestry and international trade is not driven by the measure of ancestry.

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 on trade can reflect differences in bilateral trust, values, consumption structures, political institutions, technology, as well as migration networks. In a first step, show that bilateral trust is larger among countries with a smaller ancestral distance. Furthermore, 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.[3] Third, using 4-digit trade data we find that ancestral distance is positively associated with differences in consumption patterns. Countries whose population share more

---

[3] 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.

distant common ancestors import and export (i.e. consume) a different basket of goods which partly explains why they trade less with each other. As the fourth potential channel, we find that differences political institutions, measured with Polity IV scores, increase with ancestral distance. Furthermore, countries with a larger ancestral distance are shown to differ more in the technologies they use. This in turn negatively affects bilateral trade flows. 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, there are strong correlations with ancestral distance. We expect all channels to be more important for trade of differentiated goods. To test this, 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 add to the literature on the consequences of deeply rooted long-term historical differences across countries and populations on economic outcomes. The previous literature has pointed out that these historical differences can affect the diffusion of technological innovations (Spolaore and Wacziarg, 2009), the dynamics of the fertility transition (Spolaore and Wacziarg, 2016*b*), the likelihood of war (Spolaore and Wacziarg, 2016*c*), foreign direct investment (Burchardi, Chaney and Hassan, 2017), as well as credit lending and default (Fisman, Paravisini and Vig, 2017). Guiso, Sapienza and Zingales (2009) have explored the effect of genetic distance on trade in a seminal paper using data on trade between European countries and find that similar populations are more likely to trade with each other because they trust each other more. Our study builds upon this existing evidence and advances our understanding of the effect of ancestral distance in three ways. First, we extend the data from the set of relatively homogeneous European to the near universe of international trade.[4] Second, our paper is able to make a distinction between the effect of ancestry on the extensive as well as on the intensive margin. Third, we extend the possible mechanisms

---

[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".

through which ancestry affects trade from trust to six important proximate determinants, including political institutions, technology differences, consumption structures, values, trust, and networks. 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 remained unclear from this line of research, however, whether these findings extend beyond the set of relatively homogeneous European countries. Our findings thus complement this literature, highlighting the importance of historical human relatedness for current trade patterns using data on more than 99% of world trade. In addition, our paper also relates to recent evidence by Campante and Yanagizawa-Drott (2016) who document that in-person contact is still important for establishing business links even in a globalized world.

Second, we contribute to prior research on the 'distance puzzle' which describes the fact that the estimated effect of geographic distance on trade flows has remained consistently high even after the sharp decrease of transportation costs, tariffs, and non-tariff barriers (Disdier and Head, 2008; Yotov, 2012). 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.[5] Some existing papers have highlighted channels that may explain part of the distance puzzle. Hornok and Koren (2015), for instance, show that administrative barriers reduce trade volumes and gains from trade. Kropf and Sauré (2014) document that fixed shipment costs that include filling in customs forms, organizing trade credit and monitoring the shipment are large, increase with geographic distance, and decrease with common language.

Finally, our work adds to the literature investigating the consequences of heterogeneous preferences across countries. The prominent Linder hypothesis states that two countries trade more with each other the more similar their demand structures are. This is consistent with our observation that ancestral distance increases differences in consumption patterns and thus reduces trade flows. This finding thus improves our understanding of the consequences of the 'home bias' (Trefler, 1995), the empirical regularity that tastes are historically determined and change only slowly over time. Previous research has found convincing evidence for the home

---

[5] 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.

bias for many goods and services, including the choice between oil and 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.

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 definition of each part separately.

2.1 Trade Data

Our data on international trade 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.[6]

---

[6] 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.

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.[7] For every pair, our data set has a separate entry with each 6-, 4-, 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.

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 finds 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 WTO, 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

---

[7] 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.

pair, this indicator ranges from 0 to 6 with higher values reflecting deeper integration.

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, colonial ties, common currencies, and legal origins.[8]

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 are likely to be related to ancestral distance. 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

---

[8] To measure 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 two countries weighted by population. The two distance measures are very highly correlated and our results do not depend on the selected measure.

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 usually determined as an index based on language trees. Every language belongs to a set of families according to its different characteristics and the higher the number of common families, the more similar two languages are. This approach rests on a strong cardinality assumption over branches of the trees and offers only low variation in distances. For this reason, we use the Levenshtein distance as proposed by Isphording and Otten (2013). The idea is to consider the difference in pronunciation of words having the same meaning in two different languages. The average similarity in pronunciation across a specific set of words is then considered as linguistic distance between two languages. Although this method only considers differences in phonetics, the resulting average distance also correlates with grammatical similarities between languages and allows to calculate bilateral measures that do not depend on a single anchor language or ancestor. The Levenshtein distance also yields the greatest data set, with 28,730 bilateral observations.

Religious Distance — For religious distance, we use data from the World Religion Dataset by the Association of Religion Data Archive (ARDA), which provides information about the number of adherents of each religion and the total population for each country. We then follow the approach of Helpman, Melitz and Rubinstein (2008) and build an index of religious similarity across country pairs based on three main confessions: Protestantism, Catholicism and Islam. For every country pair, we define religious similarity as the probability that two random individuals, one from country i and one from country j, share the same confession, out of these three: (% Protestants in country i × % Protestants in country j) + (% Catholics in country i × % Catholics in country j) + (% Muslims in country i × % Muslims in country j). With this method, we obtain a measure of bilateral religious similarity for 24,180 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*a*), who combine two sources of information. First, genetic differences between a large number of distinct populations measured by Pemberton, DeGiorgio and Rosenberg (2013). Second, data on the composition (fractionalization) of more than one hundred countries provided by Alesina et al. (2003). Spolaore and Wacziarg use these two sources to build 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] The main advantage of the newer data is that it is based on a significantly larger set of unique populations. Whereas the old measure relies on 42 populations defined by Cavalli-Sforza, Menozzi and Piazza (1994), the new data set is based on a much broader set of 267 worldwide populations compiled by Pemberton, DeGiorgio and Rosenberg (2013). Hence, the most recent measure provides much more detailed information on the relationship between populations in all countries, especially within Africa and Asia. For this reason, we use the measure provided by Spolaore and Wacziarg (2016*a*) in the main analysis and make use of the older measure in the robustness section.

Genetic Distance as Molecular Clock — It is crucial for our analysis to understand why we can use genetic differences between populations 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]

---

[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.

[11] The work by Pemberton, DeGiorgio and Rosenberg (2013) follows extensive research based on the Human

Microsatellites are tracts of repetitive non-coding DNA in which short DNA motifs are repeated between 5 to 50 times.

The advantage of using microsatellites in population genetics rests on their high mutation rates, their high diversity, and their non-coding 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 is ancestral relatedness between populations independently from their looks and leaving aside any quality judgement with respect to their genetic outfit. In other words, our main explanatory variable (ancestral distance) can affect trade flows insofar as: (i) it captures relatedness between populations of different countries, and (ii) historical relatedness is manifested as similarities in beliefs, conventions, customs, norms, or habits.

As an important final note, we emphasize that all measures of genetic distance are symmetric in Spolaore and Wacziarg (2016*a*). 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:

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

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

---

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 non-coding DNA regions.

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] Thus, 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]

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

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

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*a*) 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:

---

[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 $(p_i + q_i)^2 = p_i^2 + q_i^2 + 2 p_i q_i = 1$.

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

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

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. 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 three different data sets on trade statistics for the year 2000. The first one includes information on trade flows for all commodities, the second one is based on a 2-digit and the third one on a 4-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 × 171 = 29,412 bilateral observations. In our sample, the GDP per capita ranges from 263 to 130,403 U.S. dollar. The average country has an income level of about 13,985 U.S. dollar. As the numbers of observation indicate, we have country information available in all cases. For some of the bilateral variables we do not have the full set of observations, but we only use these variables in the robustness section and not in the main analysis. 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.

Using our data, we can document a clear positive relationship between geographic and genetic distance. Figure 1 depicts the relationship between a country's average genetic distance (x-axis) and its average geographic distance (y-axis) to the rest of the world.

— Figure 1 about here —

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, African countries are genetically far more distant from the rest of the world than any other group.

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

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

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 country pairs. As explained in Section 2, genetic distance indicates 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. Furthermore, we add a vector of control variables denoted by $\mathbf{X}_{o,d}$. The latter includes a varying set of variables 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 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 — Our paper aims at investigating the importance of ancestral distance on trade. Insofar as this distance is a good proxy for familiarity, we do not wish to control for trade costs. The reason is that trade costs are currently measured as the difference between the price of a good at the shipment port and its price at the port of arrival. This calculation implies that available measures of trade costs reflect "all costs incurred in getting a good to a final user, other than the marginal cost of producing the good itself" (Anderson and van Wincoop, 2004). By its very own nature, this measure includes all human barriers to international trade and controlling for it would yield tautological results.

However, if ancestral distance constitutes a barrier to international trade, it appears crucial to control for transportation costs, which do not depend on genetic relatedness between populations. Yet transportation costs data present a number of issues. First, information is 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 (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, has be shown to be inaccurate (Hummels, 2007).

Given these issues, we do not 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, the presence of a common border, access to the same sea as well as numerous other geographic and bilateral variables shown in Table 1. Moreover, 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 (Simonovska and Waugh, 2014; Föllmi, Hepenstrick and Zweimüller, 2017). The fixed effects for exporter and importer account for additional time-invariant country-specific factors that have an impact on transportation 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 2, we show the relationship between ancestral distance and bilateral trade flows. Both in the raw data as well as when netting out geographic distance and breaking genetic distance into deciles, we find a significant negative gradient.[15] This suggests that countries with a larger genetic distance trade less with each other.

— Figure 2 about here —

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 1, 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. Panels (c) and (d) of Figure 2 plot on the vertical axis the residuals of regressing Trade 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

---

[15] 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.

genetic distance that is lower than one would expect based on their geographic distance (e.g., Australia and UK). The key observation of both plots in Figure 2 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 ancestral distance as a significant determinant 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, β in equation (8), shows the correlation of trade between two countries with the respective their ancestral distance.

We build upon Helpman, Melitz and Rubinstein (2008) to define our main specification in a way that allows us to control for a plethora of both geographic and historical aspects. Specifically, geographical variables include: common border, access to same sea, longitudinal and latitudinal distance, island and landlocked status. Historical and economic variables include: per capita GDP, common currency, colonial ties, same country in the past, WTO membership.

For both the extensive and intensive margin, we begin with a univariate specification to see how much of the variation in trade can be explained by ancestral distance. We then compare the results with a similar univariate regression for geodesic distance. Subsequently, we add various control variables and observe whether genetic distance has any value added over the use of simple geodesic distance. When adding political variables or other measures of cultural distance, we expect that the negative correlation between ancestral distance and trade is 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 sign a free trade agreement, have 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 aggregate data set.

— Table 2 about here —

In column (1), we replicate Figure 2 and show that genetic distance is negatively associated with the probability that two countries have positive trade flows. It is important to note that ancestral distance alone accounts for 5.4 % of the variation in trade flows at the extensive margin. In column (2) we repeat the same exercise for geodesic distance and in column (3) we estimate our main specification omitting genetic distance. We find that geodesic distance has a statistically significant negative coefficient, even after controlling for other geographic, historical and economic variables. However, geodesic distance alone accounts for 3.9% of the variation in trade flows at the extensive margin, thus confirming that geographic and ancestral distance are not perfectly correlated.

We then add genetic distance to our preferred standard gravity equation specification in column (4). This yields a much smaller coefficient on genetic distance than in the univariate specification. This is in line with expectations as ancestral distance also correlates with historical, political and economic ties, thus capturing the effect of those variables when they are omitted. Since both genetic and geodesic distance are entered in logarithmic form, we can compare the magnitudes. While column (4) suggests genetic distance to be more important than geodesic distance, this result needs further investigation. Note that we do not include origin and destination fixed effects in our preferred specification: Some countries, like the USA, export and/or import at least one good from all other countries, which implies that they are excluded once we add country-specific fixed effects. However, the exclusion is non-random and we lose a significant portion of our sample. We report results of fixed-effects regressions in Table A.3 in the Appendix. Once we include fixed effects, we see that the coefficient of genetic distance is lower and that geodesic distance becomes more important. This also holds true when we use two-digit data as in the first four columns of Table A.4, although the magnitude of both coefficients is much lower in this case. Overall our estimates suggest that a 10% increase in ancestral distance decreases the probability that two countries establish a trade relationship by 0.7 to 1.35 percentage points.

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 and we can include both origin and destination fixed effects without loss of observations.

We present the results in the last four columns of Table 2. First, we observe that ancestral distance is negatively correlated with the total trade volume in a univariate regression with fixed effects, as column (1) shows. We repeat the same exercise for geodesic distance and show in column (2) that its coefficient is much larger than the one of genetic distance, though it becomes smaller once we add other control variables in column (3). Consistently with the existing literature (see Anderson and van Wincoop (2003)) we find the border effect to have a positive and significant effect on trade at the intensive margin.

We include genetic distance to the gravity equation and present the results in column (4). Genetic distance has a statistically significant negative impact on the intensive margin of trade, though its the magnitude of the effect is lower than the one of both geodesic distance. The results are confirmed at the two-digit level data: The last four columns of Table A.4 in the Appendix show that the coefficients are indeed all significant and similar in magnitude. Increasing the genetic distance by 10% leads to a reduction in the bilateral trade volume of 0.7 to

6 percentage points.[16] 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 control variables to the regression. However, the results of Table 2 suggest that geographic variables remain significant for the intensive margin even after the inclusion of ancestral distance.

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 main analysis in Table 2, we use a standard gravity equation specification in which we closely follow Helpman, Melitz and Rubinstein (2008) to explore the effect of genetic distance. Yet one potential concern might be that trade is simply higher among closely related countries because they share a common colonial history, a common currency or because countries with common ancestors are generally economically and politically integrated because they are geographically closer. In this case, our estimate for ancestral distance would be biased upward. To explore the sensitivity of our estimate with regard to this concern, we add different bilateral control variables to the right-hand side.

— Table 3 about here —

The entries in column (1) of Table 3 report the result of our baseline specification in Table 2. Column (2) shows the regression results when adding bilateral political control variables. We find evidence that ancestral distance negatively affects both the extensive and intensive margin

---

[16] 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.

of trade and that this coefficient remains unchanged when including economic and political controls. In addition, the previous literature has pointed out that linguistic (Egger and Lassmann, 2015) and religious closeness (Lo Turco and Maggioni, 2016) are important determinants of trade. As language and religion are part of the traits that are passed on from ancestors, we expect that the effect of ancestral distance on trade decreases once we control for these measures. Columns (3) and (4) show regression results after controlling for bilateral linguistic and religious, respectively. The coefficient of ancestral distance remains fairly constant for the extensive margin of trade, but it decreases when we consider the intensive margin. These findings indicate that traits other than language and religion are important for the establishment of a trade relationship (intensive margin). The interpretation for the intensive margin is a bit more difficult. It appears that language and religion are less important once a trade relationship has been established as the coefficients are not statistically different from zero. Overall, the coefficient of ancestral distance on trade is a a bit less robust to the inclusion of additional controls.

3.4.2 Different Time Period

In the main analysis, we use trade data from the year 2000 as the data on the population composition of countries that is used for the aggregation of the ancestral population distances by Pemberton, DeGiorgio and Rosenberg (2013) is measured in this year. To explore the robustness of our estimates to using the year 2000, we can our analysis using trade data from the other years prior and after 2000. We report in column (5) of Table 3 the results of the estimation using data from 2005. The analysis reveals that the estimates for the effect of ancestral distance on trade are almost identical to the main results at both the intensive and extensive margin of trade.[17]

3.4.3 Alternative Genetic Distance Measures

Throughout the empirical analysis so far, we used the weighted genetic distance provided by Spolaore and Wacziarg (2016*a*) as a proxy for ancestral distance based on the genetic population

---

[17] Note that the results are also robust to using data from 1995.

decoding by Pemberton, DeGiorgio and Rosenberg (2013). As explained in Section 2, this variable is our preferred measure because it is based on a large set of 267 worldwide populations. The literature, however, has traditionally used a measure for genetic distance based on Cavalli-Sforza, Menozzi and Piazza (1994) which uses fewer populations.[18] Using this variable in our specification, we obtain a negative coefficient that is significant at the 5% level for both the intensive and extensive margin of trade. The magnitude of the effect is lower for the extensive margin when compared to our preferred measure, while it is slightly higher for the intensive margin. Overall, our finding that ancestral distance reduces trade is robust to the choice of the ancestral distance measure.

3.4.4 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 each 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 negative impact of genetic distance trade is not sensitive to this reduction in the sample of countries.

In order to test whether our main estimate depends on having a specific country included, we estimated the model 172 times, each time omitting one country. For the extensive margin, the point estimate (t-value) ranges from 0.134 to 0.143 (21.1 to 23.1). For the intensive margin, we obtain point estimates from 0.052 to 0.083 which are all significant. Hence, we can rule out the possibility that a single country drives our results. Furthermore, we test whether the inclusion of any continent is crucial for our main findings. Dropping one continent at a time, we observe that the point estimates vary somewhat (0.116 to 0.169 at the extensive margin and

---

[18] The correlation between the Pemberton et al. (2013) and the Cavalli-Sforza (1994) measure is 0.785 (Spolaore and Wacziarg, 2016*a*, 12).

0.049 to 0.125 at the intensive margin) but are all significant, at least at the 10-percent level.

3.5 Identification Concerns

The empirical analysis thus far suggests that country pairs with a larger ancestral distance are less likely to trade and, if they do trade, they trade smaller numbers and lower volumes of goods. When interpreting these results as evidence of a causal relationship of ancestry on trade, we have to address two major concerns. First, the estimation might suffer from an omitted variable bias if there are other political and economic variables correlated with both trade and ancestral distance. Such variables include colonial history (Head, Mayer and Ries, 2010), historical ties (Findlay and O'Rourke, 2007), and geographic characteristics (Nunn and Puga, 2012). We add detailed country- and pair-specific control variables to mitigate such biases.

— Table 4 about here —

A second potential bias might arise due to the simultaneity of migration and trade: individuals have historically migrated to places with established trade relationships. If that is the case, our main treatment variable (genetic distance) would be endogenous. We investigate whether such a bias affects our estimates by using genetic distance based on population data in 1500 AD as instrumental variable (IV). Columns (2) and (5) in Table 4 show that the effect of ancestral distance on trade is even larger in the IV estimation. In a further test, we explicitly take into account migration flows over the last 500 years. Exploiting data on migration flows between 1500 and 2000 from Putterman and Weil (2010), we rerun our main specification for a sample of countries for which the share of indigenous population as of 1500 accounts for at least 50% of today's population. These countries are those that are less affected by migration. Columns (3) and (6) in Table 4 demonstrate that our results hold in a sample of countries with roughly the same ethnic composition as 500 years ago.

3.6 Evidence from U.S. States

Our analysis has uncovered the existence of a relationship between ancestral distance and international trade for both the extensive and intensive margin for a country-level dataset that covers more than 99% of global trade. While the use of cross-national data has the advantage that the results are valid for the near-universe of international trade, it comes at the cost that several sources of heterogeneity cannot be observed. Subnational data, in contrast, has the advantage that several country-specific characteristics, like trade agreements with other countries, can be held constant. To explore the robustness of the relationship between ancestry and trade in a dataset with subnational heterogeneity, we use data from the U.S. Census to derive an alternative measure of ancestral relatedness. In addition, this sensitivity check tests the robustness of our results when using an alternative measure for ancestry that is not based on genetic distance.

The 2000 Census of the US Bureau of Economics and Statistics asks survey participants to name the ancestry group with which they most closely identify. This means that for each U.S. state we can construct the share of population with ancestral relationship to virtually every other country in the world.[19] Figure A.2 in the Appendix shows the largest ancestral populations for every U.S. state at the municipality level. The three largest ancestry groups are Germans (49m), Africans (41m) and Irish (36m). The figure highlights the spatial distribution and clustering of ancestry groups in the United States.[20]

We combine this ancestry information with state-level exports data, generating a data set that includes 50 U.S. states and 27 partner countries. In addition to trade and ancestry information, we also gather data on geodesic distance and common border between each state and partner country, as well as GDP and population in every state.[21] We can then investigate whether US states export more goods to countries with which they are more closely related. We present a first descriptive result in Figure 3.

— Figure 3 about here —

The figure shows that bilateral export at the U.S.-state-partner-country level is positively

---

[19] The reported ancestry does not necessarily coincide with de facto genetic relatedness because individuals with mixed ancestries might choose to only report the one to which they feel most closely related. However, self-reporting still allows us to identify the nurturing aspect of ancestral relatedness, making this measure a valid proxy for intergenerationally transmitted traits.

[20] There is country-level information on African ancestry and thus we aggregate all information for Africa at the continent-level.

[21] For the geographic distance between each U.S. state and the 27 partner countries, we take the air distance calculated with the great circle formula.

correlated with the share of the state population that identifies the partner country as their main ancestor. We then turn to the econometric investigation by building a standard gravity equation for exports of U.S. states. We present the results in Table 5.

— Table 5 about here —

In column (1), we replicate Figure 5 and show that exports and ancestral relatedness are positively correlated. Notably, the ancestral relatedness alone accounts for almost 19% of the variation in exports. In column (2), we add geodesic distance and find that its effect is insignificant and the coefficient of ancestral origin remains virtually unchanged. Columns (3) and (4) add common border and state-level controls, while column (5) repeats the exercise with the inclusion of state-fixed effects. In this case, the coefficient of interest actually increases in magnitude. Finally, columns (6) and (7) use only imports and exports, respectively. In both cases, we obtain almost identical point estimates for the effect of ancestral relatedness on trade.

This exercise at the U.S. state level confirms that ancestral relatedness matters for international trade outcomes even in a setting in which we can hold constant country-specific unobserved heterogeneity. It is worth pointing out that the measure of ancestral relatedness that we use here does not allow us to take into account similarities between ancestors. This means that to identify the effect of ancestry on, say, Connecticut's exports to Italy, we are only considering the share of Connecticut's population having Italian ancestry, disregarding the share of the population identifying as French (and thus genetically close to the Italians). The fact that we still find a strong relationship between exports and ancestry underlines the relevance of ancestral relatedness for international trade. These findings arein line with recent work by Burchardi, Chaney and Hassan (2017) who show an empirical link between migration into the United States and FDI from and to these sending countries.

4 Discussion of Channels

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. Having established a negative association between ancestral distance and bilateral trade flows, the question arises which channels account for this relationship. In this section, we discuss how ancestral distance is linked to numerous proximate determinants of trade. This includes trust, technology institutions, consumption structures, migration networks, and values.

4.1 Trust

In their seminal contribution, Guiso, Sapienza and Zingales (2009) suggest that the effect of ancestral distance on trade might operate through trust which serves as a complement for incomplete contracts. To explore the relationship between trust and ancestral distance, we collected data on bilateral trust from the Eurobarometer, an annual survey conducted by the European Commission in European countries. Overall, our sample includes fifteen countries in which a representative group was asked how much they trust people from 25 nations, including from their own country.[22] This provides us with a total of 360 observations. Similar to Guiso, Sapienza and Zingales (2009), our sample includes mostly European countries but survey participants were also asked how much they trust people from the United States, Russia, and Japan.

— Table 6 about here —

In column (1) of Table 6, we show that bilateral trust is correlated with ancestral distance and that this relationship remains robust when including fixed effects. In a regression with reporter-fixed effects, we find that a one-standard-deviation increase in ancestral distance lowers bilateral trust by about 12 percent of a standard deviation.

---

[22] We use data from the Eurobarometer 46 which included the following countries: Belgium, Denmark, Germany, Greece, Spain, France, Ireland, Italy, Luxembourg, Netherlands, Austria, Poland, Finland, Sweden, and United Kingdom. The survey question was "I would like to ask you a question about how much trust you have in people from various countries. For each, please tell me whether you have a lot of trust, some trust, not very much trust or no trust at all?". We provide further details on the data in the Appendix.

4.2 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 close 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 data set that includes measures of technology 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 data set 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. Interestingly, ancestral distance is positively related to differences in the technological level as suggested by column (2) of Table 6. A one standard deviation increase in ancestral distance increases technological differences by 25.5% of a standard deviation when we account for country-specific fixed effects.

4.3 Institutions

Another channel through which ancestral distance might influence bilateral trade is through political institutions because stable institutions foster trade by mitigating contractual problems (Greif, Milgrom and Weingast, 1994). As suggested by Alesina and Giuliano (2015) and others, the diffusion of political institutions is shaped by the relatedness of populations. More specifically, Spilimbergo (2009) provides evidence that people who study in democratic countries are more likely to acquire democratic values which then promotes democracy in their home countries. To test this link between ancestry and institutions, we use Polity IV data from Marshall, Gurr and Jaggers (2017). The Polity score is computed by subtracting the autocracy score from the democracy score, resulting in a unified polity scale which ranges from +10 (strongly democratic) to -10 (strongly autocratic). The results in column (3) of Table 6 indicate that there is a strong relationship between ancestry and political institutions. The larger the ancestral distance between two nations' populations the larger the difference in political institutions.

4.4 Consumption Structures

A fourth important proximate determinant through which ancestral distance affects trade are consumer 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 consumption structures.

To investigate the relationship between ancestral distance and consumption structures, we define a dummy variable that takes the value one if a country consumes a certain commodity, and zero otherwise.[23] For a given country pair, we then calculate the share of goods that is consumed in both countries. This measure reflects the unconditional probability that two countries both consume a commodity. Column (4) of Table 6 documents that this measure is strongly correlated with ancestral distance. A one standard deviation increase in ancestral distance increases differences in consumption patterns by 12.5% of a standard deviation when we account for country-pair-specific fixed effects.

4.5 Migration Networks

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

---

[23] Note that we use the UNIDO definition of apparent consumption which is given by: apparent consumption = domestic output + total import - total export.

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 as it creates networks that are likely to directly affect international trade by facilitating the matching between buyers and sellers (Rauch, 1999; Rauch and Trindade, 2002). The importance of networks for trade decisions has been documented for Spain (Peri, Requena-Silvente and Davis, 2010), France (Combes, Lafourcade and Mayer, 2005), and Italy (Bratti, De Benedictis and Santoni, 2014), while a recent literature emphasizes that migrant networks might also be important for foreign direct investment (Javorcik et al., 2011; Burchardi, Chaney and Hassan, 2017). 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). Column (5) in Table 6 documents that recent migration stocks are negatively correlated with ancestral distance. A one standard deviation increase in ancestral distance decreases networks the by 12.7% of a standard deviation in the estimation using country-pair fixed effects.

4.6 Values

A final but fundamental proximate determinant of trade is captured by the question whether countries share a common set of values. Individual attitudes towards role of the family, the perception of life, and answers to moral and religious questions are important cultural traits that are transmitted intergenerationally. 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 which includes a total of 341,271 surveyed individuals from 100 countries[24] for the period 1981–2014 with $(89 \times 90)/2$ unique pair observations. We then compute the Manhattan distance based on the answers of 861 questions for which WVS information is available. The average

---

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

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. The entries in column (6) in Table 6 document that this measure is strongly correlated with ancestral distance. A one standard deviation increase in ancestral distance increases differences in values by 57.7% of a standard deviation.

4.7 Ancestral Distance, Culture and Incomplete Contracts

Ancestral distance is strongly related 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 levels of trust between two populations lead to more trade. To formally examine the consequences of trust on trade, Guiso, Sapienza and Zingales (2009) propose an analysis 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 7 about here —

We merged Rauch's commodity classification to our 4-digit trade data and ran our main regression specification using the number of goods traded per country-pair as a dependent variable (in logs). Table 7 reports the results. We find that the effect of ancestral distance is substantially larger for differentiated than for reference-priced products. More interestingly, ancestral distance seems to have no impact on homogeneous products which are traded on organized exchanges and for which there is little uncertainty about a product's quality. These findings are consistent with the notion that ancestral distance operates (though not exclusively) 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 that covers more than 99% of bilateral trade among 172 countries and combine it with detailed information about bilateral ancestral distance between countries. 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 microgeographic and political control variables as well as other measures of cultural similarity, including linguistic and religious distance.

We interpret our findings as evidence for deep-rooted human barriers to international trade. From a policy perspective, one might argue that genetic distances between countries cannot be changed, at least in the short run, and thus it may remains unclear what we learn from the results of our study. We argue that the negative relationship between ancestral distance and trade may help policymakers designing the institutional setting of trade. For example, removing tariff barriers between ancestrally distant countries seems to be particularly efficiency-enhancing because these countries are prima facie unlikely to trade with each other even in the presence of pronounced comparative advantages.

One possible direction for future research is integrating these reduced-form estimates of the human barriers to trade into structural models of trade. An important question in this regard is whether the costs associated with ancestry are simply a fixed market entry cost or rather part of the variable trade costs that increase proportionally to the units of goods shipped from origin to destination country (Melitz, 2003). The large effect for the extensive margin documented in this paper indicate that ancestry may be rather a one-time market entry cost. The answer to this question has effects for the welfare consequences of trade reforms (Arkolakis, Costinot and Rodríguez-Clare, 2012; Melitz and Redding, 2015). Another fruitful area of research is to explore the relationship between ancestral distance and transportation costs over time.

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.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New Trade Models, Same Old Gains?" *American Economic Review*, 102(1): 94–130.

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.

Backus, David K, Patrick J Kehoe, and Finn E Kydland. 1992. "International Real Business Cycles." *Journal of Political Economy*, 100(4): 745–775.

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.

Balistreri, Edward J., and Russell H. Hillberry. 2006. "Trade frictions and welfare in the gravity model: How much of the iceberg melts?" *Canadian Journal of Economics*, 39(1): 247–265.

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.

Bratti, Massimiliano, Luca De Benedictis, and Gianluca Santoni. 2014. "On the pro-trade effects of immigrants." *Review of World Economics*, 150(3): 557–594.

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.

Burchardi, Konrad B., Thomas Chaney, and Tarek A. Hassan. 2017. "Migrants, Ancestors, and Foreign Investments."

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.

Campante, Filipe, and David Yanagizawa-Drott. 2016. "Long-Range Growth: Economic Development in the Global Network of Air Links."

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.*

Combes, Pierre Philippe, Miren Lafourcade, and Thierry Mayer. 2005. "The trade-creating effects of business and social networks: Evidence from France." *Journal of International Economics*, 66(1): 1–29.

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, 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 Andrea Lassmann. 2015. "The Causal Impact of Common Native Language on International Trade: Evidence from a Spatial Regression Discontinuity Design." *The Economic Journal*, 125(584): 699–745.

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.*

Findlay, Ronald, and Kevin H. O'Rourke. 2007. *Power and Plenty.* Princeton, NJ:Princeton University Press.

Fisman, Raymond, Daniel Paravisini, and Vikrant Vig. 2017. "Cultural Proximity and Loan Outcomes." *American Economic Review*, 107(2): 457–492.

Föllmi, Reto, Christian Hepenstrick, and Josef Zweimüller. 2017. "International Arbitrage and the Extensive Margin of Trade between Rich and Poor Countries." *Review of Economic Studies*, forthcoming.

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

Greif, Avner, Paul Milgrom, and Barry R Weingast. 1994. "Coordination, Commitment, and Enforcement." *Journal of Political Economy*, 102(4): 745–776.

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.

Head, Keith, Thierry Mayer, and John Ries. 2010. "The erosion of colonial trade linkages after independence." *Journal of International Economics*, 81(1): 1–14.

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.

Hornok, Cecília, and Miklós Koren. 2015. "Administrative barriers to trade." *Journal of International Economics*, 96(S1): S110–S122.

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.

Isphording, Ingo Eduard, and Sebastian Otten. 2013. "The costs of babylon-linguistic distance in applied economics." *Review of International Economics*, 21(2): 354–369.

Javorcik, Beata S., Çalar Özden, Mariana Spatareanu, and Cristina Neagu. 2011. "Migrant networks and foreign direct investment." *Journal of Development Economics*, 94(2): 231–241.

Kropf, Andreas, and Philip Sauré. 2014. "Fixed costs per shipment." *Journal of International Economics*, 92(1): 166–184.

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

Lo Turco, Alessia, and Daniela Maggioni. 2016. "For God's Sake: The Impact of Religious Proximity on Firms' Exports."

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."

Marshall, Monty G., Ted Robert Gurr, and Keith Jaggers. 2017. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2016. 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.

McCallum, John. 1995. "National Borders Matter: Canada-U.S. Regional Trade Patterns." *American Economic Review*, 85(3): 615–623.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6): 1695–1725.

Melitz, Marc J., and Stephen J. Redding. 2015. "New Trade Models New Welfare Implications." *American Economic Review*, 105(3): 1105–1146.

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.

Peri, Giovanni, Francisco Requena-Silvente, and U C Davis. 2010. "The Trade Creation Effect of Immigrants: Evidence from the Remarkable Case of Spain." *Canadian Journal of Economics*, 43(4): 1433–1459.

Putterman, Louis, and David N. Weil. 2010. "Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality." *Quarterly Journal of Economics*, 125(4): 1627–1682.

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.

Santos Silva, J. M. C., Silvana Tenreyro, and Kehai Wei. 2014. "Estimating the extensive margin of trade." *Journal of International Economics*, 93(1): 67–75.

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.

Spilimbergo, Antonio. 2009. "Democracy and Foreign Education." *American Economic Review*, 99(1): 528–543.

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

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

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*a*. "Ancestry and Development: New Evidence."

Spolaore, Enrico, and Romain Wacziarg. 2016*b*. "Fertility and Modernity." *Tufts University Discussion Papers Series 0779.*

Spolaore, Enrico, and Romain Wacziarg. 2016*c*. "War and Relatedness." *Review of Economics and Statistics*, 98(5): 925–939.

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

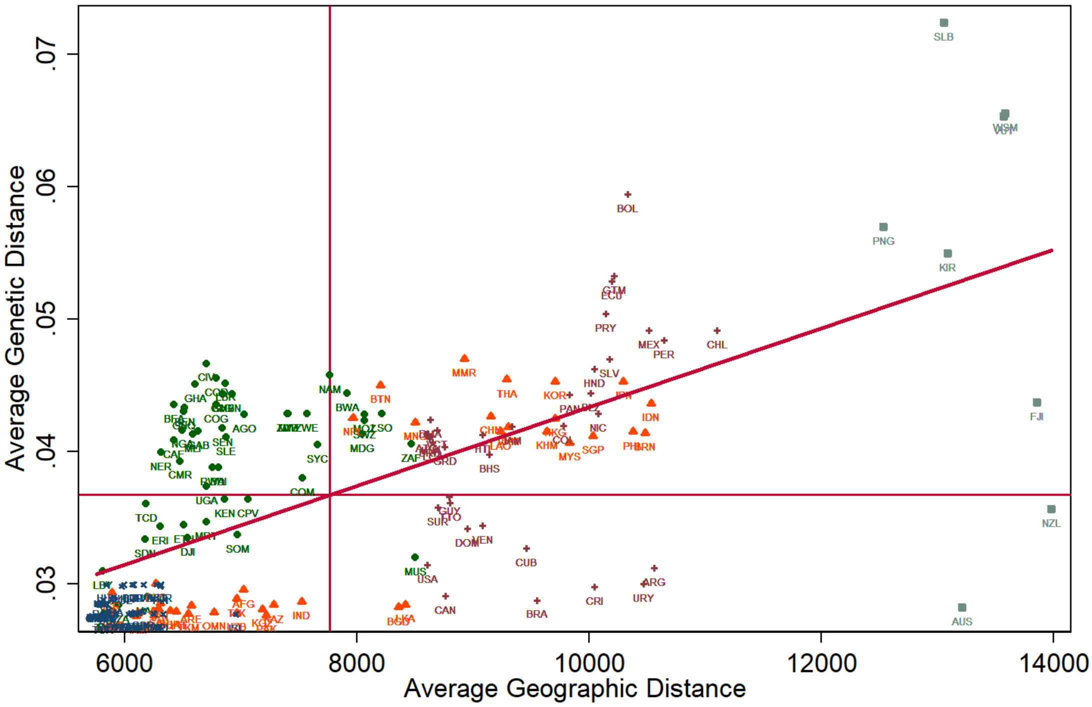
Tinbergen, Jan. 1962. *The World Economy. Suggestions for an International Economic Policy.* New York:Twentieth Century Fund.

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

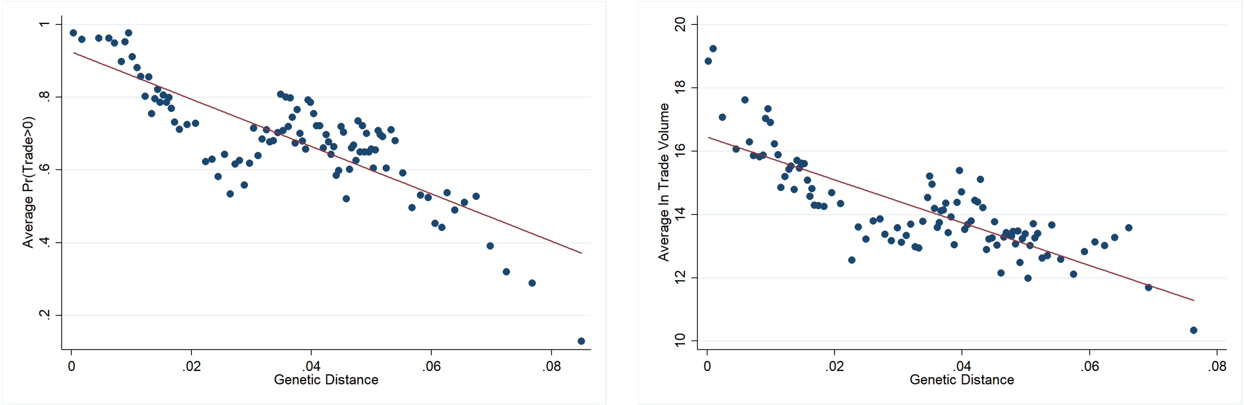
Figure 1: Geographic and Genetic Distance



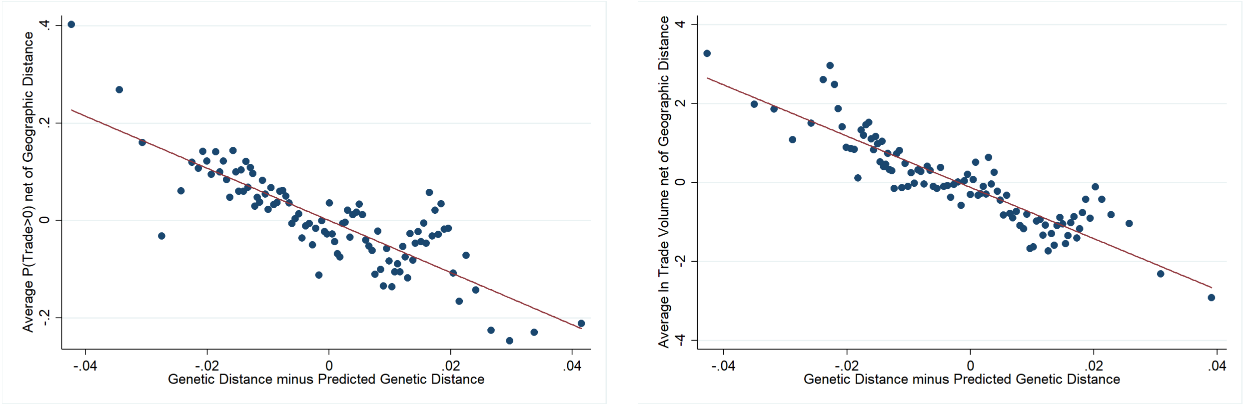
Note: The figure shows each country's average genetic distance to its (potential) trading partners on the vertical axis plotted against each country's average geographic distance to all its (potential) trading partners on the horizontal axis. Countries are colored according to their continent. The horizontal and vertical lines show the average for the genetic (0.037) and geographic distance (7,769 km), respectively. The solid line shows a linear fit with a t-value of 7.25 in a regression with clustered standard errors.

Figure 2: Genetic Distance and Trade

(a) Extensive Margin (b) Intensive Margin

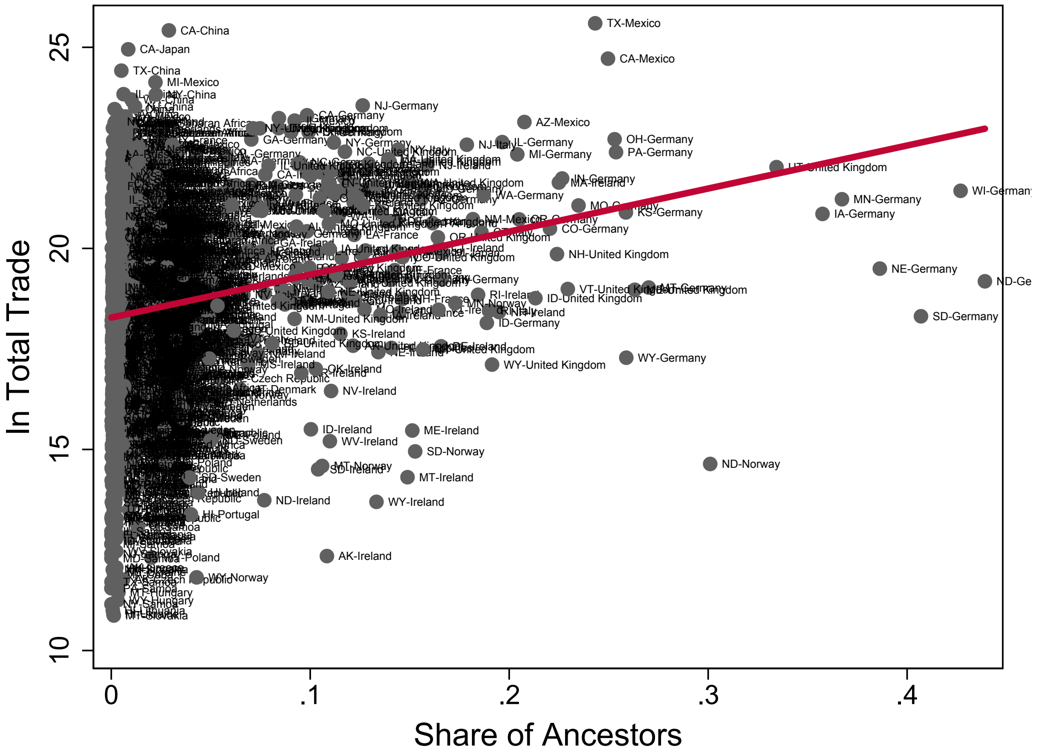


(c) Extensive Margin (d) Intensive Margin



Note: The figures plot the extensive and intensive margin of trade against bilateral genetic distance. In plots (a) and (b), we use the raw data while in plots (c) and (d), we control for geographic distance. In particular, we use on the x-axes the difference between genetic distance and predicted genetic distance based on geographic distance and on the y-axes we use the residuals of trade regressed on geographic distance. We group all 29,412 country pairs into percentiles according to their genetic distance. A linear fit is shown in each plot.

Figure 3: U.S. States Exports and Ancestral Heritage



Note: The figure shows the relationship between total trade at the US-state level and the share of the state population identifying the importer's population as its main ancestral group. Each dot reflects a state-country pair (50 states and 27 partner countries) and the solid line indicates a linear fit.

Table 1: Descriptive Statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Part I: Country-Level Variables:					
GDP (mil. 2005 USD)	359,922.25	1,190,108.92	112.24	1,2975,535	172
GDP per capita	13,984.56	19,383.87	262.52	130,402.97	172
Population (mil.)	34.83	129.19	0.05	1269.97	172
Landlocked	0.22	0.41	0	1	172
Island	0.21	0.41	0	1	172
Latitude	19.02	24.3	-41.81	64.99	172
Longitude	16.95	62.64	-172.17	171.48	172
Part II: Bilateral Variables:					
Genetic Distance (weighted, new)	0.04	0.02	0	0.09	29,412
Genetic Distance (weighted, old)	0.11	0.07	0	0.35	29,412
Linguistic Distance (WALS)	93.73	16.79	0	107.33	28,730
Religious Similarity	0.17	0.23	0	0.99	24,180
Geodesic Distance (biggest city, km)	7,769.03	4,404.82	10.48	19,904.45	29,412
Common Border	0.02	0.13	0	1	29,412
Island (none/one, both)	0.04	0.2	0	1	29,412
Landlocked (none/one, both)	0.05	0.21	0	1	29,412
Colonial Ties	0.01	0.11	0	1	29,412
Common Currency	0.02	0.15	0	1	29,412
Both WTO Members	0.73	0.44	0	1	29,412
Both not WTO Members	0.02	0.14	0	1	29,412
Access to Same Sea	0.16	0.37	0	1	29,412
Common Off. Language	0.16	0.36	0	1	29,412
Same Country in Past	0.01	0.09	0	1	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
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	56,570.2	425,570
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. At the 4-digit level, there are 782 commodities.

Table 2: Main Regression Results with Aggregate Data

I. Extensive Margin					
			Positive Trade Flow		
Mean of dep. variable			0.685		
	(1)	(2)	(3)	(4)	(5)
log Genetic Distance	-0.188***			-0.140***	-0.027***
	(0.005)			(0.006)	(0.006)
log Geodesic Distance		-0.146***	-0.283***	-0.178***	-0.128***
		(0.004)	(0.011)	(0.011)	(0.010)
Common Border			0.006	0.033	-0.041
			(0.037)	(0.037)	(0.040)
Controls	-	-	yes	yes	HMR
Fixed Effects	-	-	-	-	-
Observations	29,358	29,358	29,358	29,358	29,358
R-squared	0.054	0.039	0.082	0.010	0.376

II. Intensive Margin					
			Trade Value		
Mean of dep. variable			200.52		
	(1)	(2)	(3)	(4)	(5)
log Genetic Distance	-0.608***			-0.071***	-0.041
	(0.036)			(0.030)	(0.026)
log Geodesic Distance		-0.947***	-0.669***	-0.623***	-0.635***
		(0.027)	(0.048)	(0.053)	(0.056)
Common Border			0.756***	0.764***	0.610 ***
			(0.091)	(0.088)	(0.094)
Controls	-	-	yes	yes	HMR
Fixed Effects	yes	yes	yes	yes	yes
Observations	29,358	29,358	29,358	29,358	29,358
R-squared	0.609	0.890	0.919	0.919	0.916

Note: The top part of the table shows the result of five Probit regressions using the extensive margin of trade as dependent variable. All coefficients are marginal effects at means. We do not include FE in the regression in columns (1) to (6) using aggregate data because this would exclude all countries that export to all destinations or import from all origin countries, namely South Korea, Singapore, and United States. The results with FE can be found in Table A.3 in the Appendix. 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. Both origin and destination fixed effects (FE) are included in the PPML regressions. The data comprises aggregate bilateral trade flows from the year 2000. Control variables include being an island, being landlocked, access to the same sea, difference in latitude, difference in longitude. Additional control variables follow Helpman, Melitz and Rubinstein (2008) (HMR) and include GDP per capita of exporter and importer, total population, colonial relationship, common currency, common membership in the WTO, common official language, same country in the past, 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

	Positive Trade Flow					
Mean of dep. variable	0.685	0.683	0.691	0.680	0.733	0.686
	(1)	(2)	(3)	(4)	(5)	(6)
log Genetic Distance	-0.140***	-0.136***	-0.140***	-0.145***	-0.121***	
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	
log Geodesic Distance	-0.178***	-0.121***	-0.171***	-0.178***	-0.172***	-0.254***
	(0.011)	(0.011)	(0.011)	(0.012)	(0.011)	(0.012)
Common Border	0.033	-0.027	0.075*	0.031	-0.053*	0.020
	(0.037)	(0.042)	(0.044)	(0.038)	(0.032)	(0.038)
Common Currency		-0.275***				
		(0.027)				
Colonial Ties		0.558***				
		(0.089)				
Economic Integration		0.098***				
		(0.007)				
Both Democratic		0.140***				
		(0.006)				
Linguistic Distance (WALS)			-0.128***			
			(0.029)			
Religious Similarity				0.005***		
				(0.001)		
Genetic Distance (Cavalli-Sforza)						-0.036***
						(0.004)
Year	2000	2000	2000	2000	2005	2000
Controls	yes	yes	yes	yes	yes	yes
Observations	29,358	29,150	28,040	28,242	29,358	29,330
R-squared	0.100	0.128	0.097	0.106	0.101	0.087

II. Intensive Margin

	Trade Value					
Mean of dep. variable	197.667	197.667	202.171	189.340	323.969	200.900
	(1)	(2)	(3)	(4)	(5)	(6)
log Genetic Distance	-0.070**	-0.072***	-0.065**	-0.039	-0.065**	
	(0.030)	(0.022)	(0.033)	(0.028)	(0.027)	
log Geodesic Distance	-0.624***	-0.399***	-0.653***	-0.667***	-0.624***	-0.632***
	(0.054)	(0.057)	(0.063)	(0.050)	(0.051)	(0.049)
Common Border	0.764***	0.492***	0.734***	0.633***	0.637***	0.803***
	(0.088)	(0.086)	(0.094)	(0.086)	(0.079)	(0.080)
Common Currency		-0.259***				
		(0.089)				
Colonial Ties		0.480***				
		(0.096)				
Economic Integration		0.349***				
		(0.029)				
Both Democratic		-0.313***				
		(0.105)				
Linguistic Distance (WALS)			0.191			
			(0.208)			
Religious Similarity				0.010		
				(0.018)		
Genetic Distance (Cavalli-Sforza)						-0.086***
						(0.027)
Year	2000	2000	2000	2000	2005	2000
Controls	yes	yes	yes	yes	yes	yes
Origin FE	yes	yes	yes	yes	yes	yes
Destination FE	yes	yes	yes	yes	yes	yes
Observations	29,296	29,296	27,978	28,180	29,358	29,330
R-squared	0.919	0.947	0.873	0.927	0.888	0.921

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. We do not include FE in the regression in columns (1) to (6) using aggregate data because this would exclude all countries that export to all destinations or import from all origin countries, namely South Korea, Singapore, and United States. 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. Control variables include being an island, being landlocked, access to the same sea, difference in latitude, difference in longitude. 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: Robustness Test for the Endogeneity of Genetic Distance

	Positive Trade Flow			**Trade Value**		
Mean of dep. variable	0.673		0.751		157.024	
	(1)	(2)	(3)	(4)	(5)	(6)
log Genetic Distance	-0.156***	-1.041***	-0.154***	0.166***	-0.524***	-0.073**
	(0.007)	(0.032)	(0.012)	(0.043)	(0.127)	(0.029)
log Geodesic Distance	-0.186***	-0.039	-0.100***	-1.025***	-0.733***	-0.443***
	(0.013)	(0.038)	(0.021)	(0.063)	(0.076)	(0.064)
Common Border	0.022	0.293**	0.076**	0.558***	1.060***	0.442***
	(0.049)	(0.125)	(0.037)	(0.096)	(0.146)	(0.075)
Controls	yes	yes	yes	yes	yes	yes
Method	Probit	IV Probit	Probit	PPML	IV Poisson	PPML
Observations	27,898	27,898	14,472	27,898	27,898	14,472
R-squared	0.091	-	0.101	0.955	-	0.891

Note: The first three columns of the table show the results 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. In columns (3) and (6), we restrict the sample according to data from Putterman and Weil (2010) with share of indigenous population as of 1500 in today's population is larger than 50 percent. Control variables include being an island, being landlocked, access to the same sea, difference in latitude, difference in longitude. Following Table 2, we only add exporter- and importer-fixed effects for the intensive margin. In order to obtain comparable estimates, we restrict the sample in columns (1) and (4) to observations without missing 1500 genetic distance data. 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: Trade of U.S. States and Share of Ancestors

	Total Trade					**Imports**	**Exports**
Mean of dep. variable			18.577			17.518	17.287
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln Share Ancestors	0.357***	0.310***	0.270***	0.248**	0.274***	0.277***	0.278**
	(0.138)	(0.101)	(0.086)	(0.100)	(0.104)	(0.102)	(0.123)
ln Geodesic Distance		-0.729	0.187	-0.126	-0.078	0.198	-0.504
		(0.770)	(1.212)	(0.958)	(0.984)	(1.299)	(0.573)
Common Border			3.287*	1.701	1.564	1.762	1.266
			(1.815)	(1.353)	(1.549)	(1.940)	(1.051)
ln State GDP				0.659			
				(0.473)			
ln State Population				0.436			
				(0.493)			
State FE	-	-	-	-	yes	yes	yes
Observations	1,350	1,350	1,350	1,350	1,350	1,350	1,350
R-squared	0.021	0.095	0.214	0.362	0.456	0.278	0.813

Note: The table shows the result of seven separate PPML regressions. The first five columns use the state-level total trade values as dependent variable, column (6) uses total imports and column (7) uses exports as dependent variable. The data set comprises trade flows between 50 U.S. states and 27 countries worldwide. Note that Sub-Saharan African countries are merged into one country in the original Census data. Standard errors are shown in parentheses and clustered at the state-level. Significance at the 10% level is indicated by *, at the 5% level by **, and at the 1% level by ***.

Table 6: Ancestral Distance and Proximate Determinants of Trade

	(1) Trust	(2) Technology	(3) Political Institutions	(4) Consumption Structures	(5) Migration Networks	(6) Values
Correlation with Ancestral Distance	-0.161	0.038	0.025	0.216	-0.320	0.094
Coefficient on Ancestral Distance	-0.121***	0.255***	0.185***	0.125***	-0.127***	0.577***
	(0.034)	(0.038)	(0.026)	(0.021)	(0.009)	(0.047)
N	360	8,385	21,756	11,556	29,412	6,806

Note: The table shows the raw correlation between ancestral distance and six proximate determinants of international trade in the first row. The second row shows the coefficient of ancestral distance in a regression with the proximate determinant as dependent variable and reporter and partner fixed effects (columns 2, 3, 4, and 6) and reporter fixed effects (columns 1 and 5). Note that we use only reporter fixed effects in column 1 due to the limited sample size. We also use only reporter fixed effects in column 5 because technological difference is symmetric across country pairs and thus we use only one observation per country pair. All variables are standardized such that the coefficient can be interpreted as the percent change of a standard deviation in the dependent variable in response to a one standard deviation increase in the independent variable. Standard errors are shown in parentheses and clustered at the reporter level. Significance at the 10% level is indicated by *, at the 5% level by **, and at the 1% level by ***.

Table 7: Types of Commodities Traded

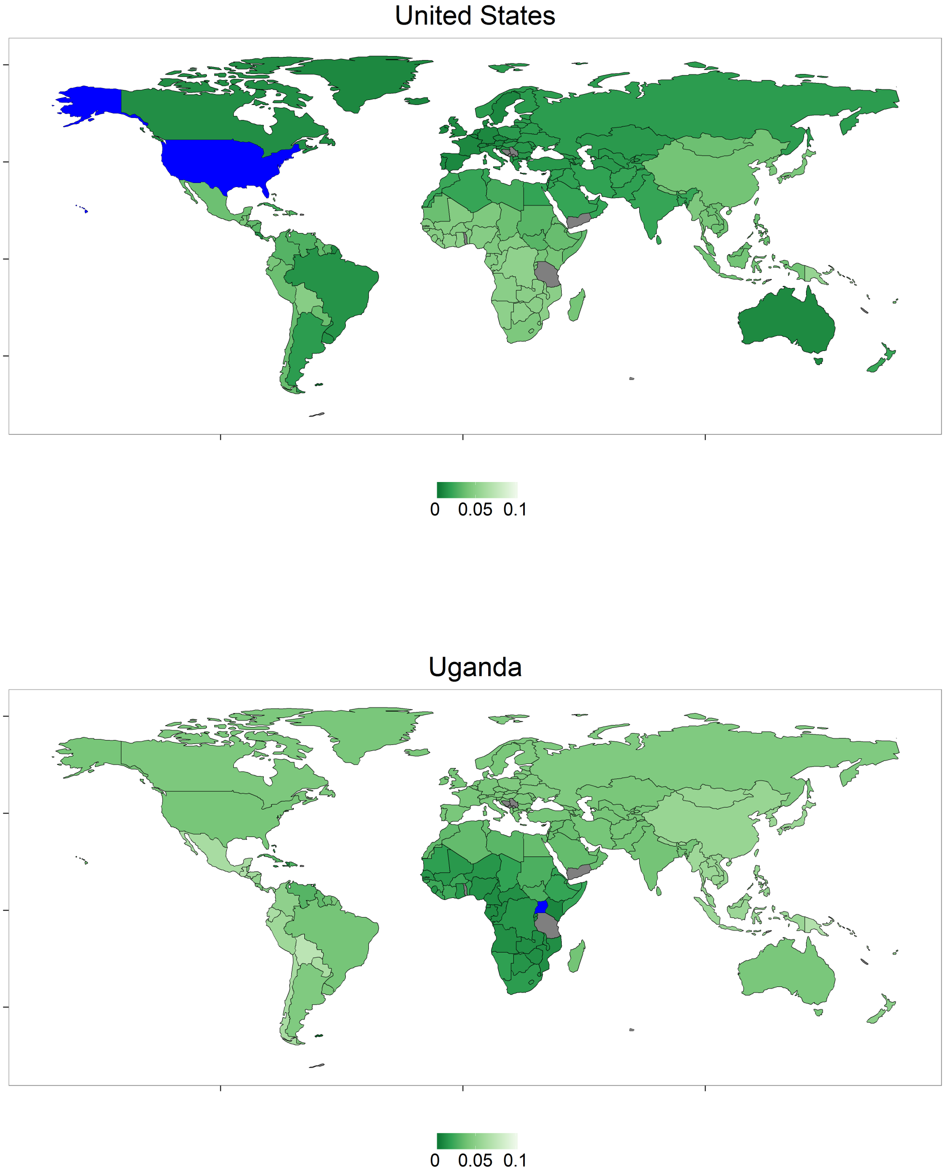
	Number of Commodities Traded			
Mean of dep. variable	69.58	43.84	15.29	4.94
	(1)	(2)	(3)	(4)
log Genetic Distance	0.040***	0.048***	0.024*	0.002
	(0.011)	(0.011)	(0.012)	(0.014)
log Geodesic Distance	-0.541***	-0.522***	-0.583***	-0.581***
	(0.027)	(0.026)	(0.030)	(0.032)
Common Border	-0.011	-0.054	0.007	0.138*
	(0.066)	(0.066)	(0.068)	(0.071)
Commoditiy Types	all	Differentiated	Ref.-Price	Homogenous
Controls	yes	yes	yes	yes
Origin FE	yes	yes	yes	yes
Destination FE	yes	yes	yes	yes
Observations	29,358	29,358	29,358	29,358
R-squared	0.692	0.692	0.665	0.606

Note: The table shows the result of four PPML 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. Control variables include being an island, being landlocked, access to the same sea, difference in latitude, difference in longitude. 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

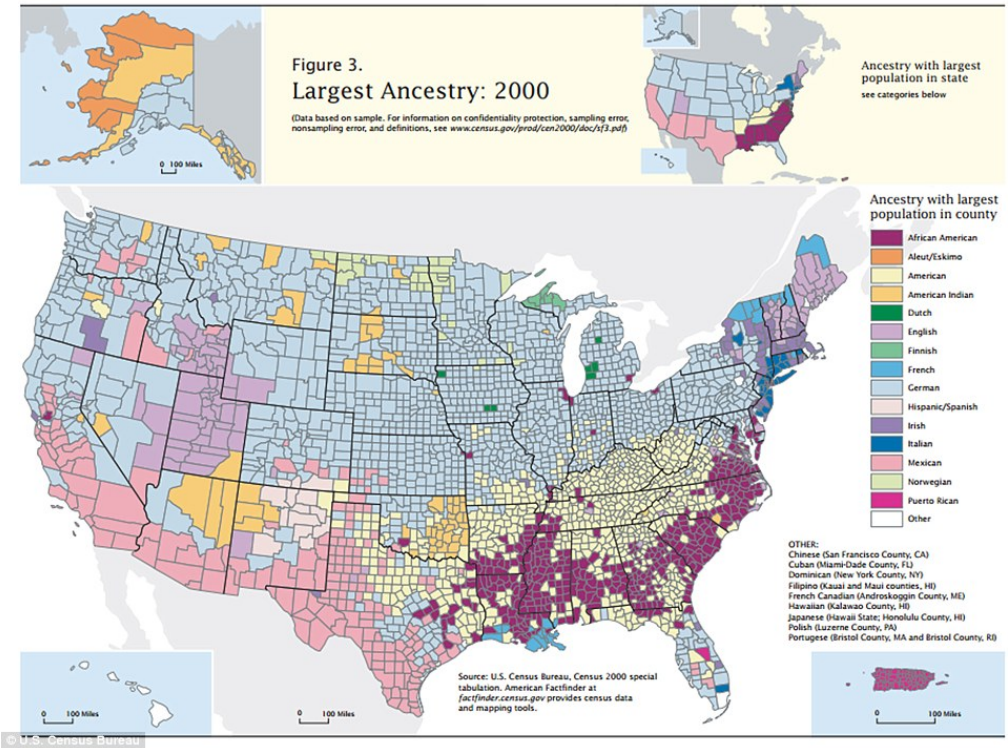
Appendix A: Additional Figures and Tables

Figure A.1: Genetic Distance in a World Map



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: Ancestral Origin in U.S. States



Note: The figure shows the major ancestry groups in each US state by municipality. The largest ancestry groups in the US are: 49m Germans, 41m Africans, 36m Irish, 32m Mexican, 27m English, 18m Italian, 10m Polish, 9m French. Source: U.S. Census Bureau, Census 2000 Special Tabulation, American Factfinder at factfinder.census.gov.

Table A.1: Countries in the Sample

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

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*a*): 0.701% of Trade Volume

Aruba
Bermuda
Bosnia Herzegovina
Br. Virgin Isds
Cayman Isds
Cook Isds
French Polynesia
Greenland
Maldives
Marshall Isds
Palau
Sao Tome and Principe
Serbia and Montenegro
Tanzania
Timor-Leste
Tonga
Turks and Caicos Isds
Yemen

(B) No Linguistic Distance in Isphording and Otten (2013): 0.717% of Trade Volume

Afghanistan
Aruba
Bermuda
Bosnia Herzegovina
Br. Virgin Isds
Brunei
Cayman Isds
Cook Isds
French Polynesia
Greenland
Maldives
Marshall Isds
Palau
Sao Tome and Principe
Serbia and Montenegro
Tanzania
Timor-Leste
Tonga
Turks and Caicos Isds
Yemen

(C) No Religion Data from the Association of Religion Data Archive: 0.764% of Trade Volume

Aruba
Bermuda
Bosnia Herzegovina
Br. Virgin Isds
Cayman Isds
Cook Isds
French Polynesia
Greenland
Hong Kong
Maldives
Marshall Isds
Palau
Sao Tome and Principe
Serbia and Montenegro
Tanzania
Timor-Leste
Tonga
Turks and Caicos Isds
Vietnam
Yemen

Note: The table lists all countries with missing linguistic or religious data separated in two different categories. Panel (A) reports all countries that have no weighted genetic distance data in Spolaore and Wacziarg (2016*a*), Panel (B) lists all countries with missing linguistic distance data provided by Isphording and Otten (2013), and Panel (C) those with missing religious distance data as provided by the Association for Religion Data.

Table A.3: The Extensive Margin of Trade with FE

	Positive Trade Flow				
Mean of dep. variable			0.685		
	(1)	(2)	(3)	(4)	(5)
log Genetic Distance	-0.201***			-0.080***	-0.062***
	(0.008)			(0.009)	(0.009)
log Geodesic Distance		-0.240***	-0.249***	-0.199***	-0.183***
		(0.007)	(0.014)	(0.014)	(0.015)
Common Border			-0.011	-0.009	-0.051
			(0.043)	(0.043)	(0.044)
Controls	-	-	yes	yes	HMR
Origin FE	yes	yes	yes	yes	yes
Destination FE	yes	yes	yes	yes	yes
Observations	25,506	25,506	25,506	25,506	25,506
R-squared	0.477	0.509	0.510	0.514	0.521

Note: The table replicates the results of the first five columns of table 2 with the addition of origin and destination fixed effects (FEs). Whereas the inclusion of FEs reduces the bias of the estimator, this improvement comes at a cost as we lose about 13% of observations, which amount to 75% of the world trade volume. Control variables include being an island, being landlocked, access to the same sea, difference in latitude, difference in longitude. Additional control variables follow Helpman, Melitz and Rubinstein (2008) (HMR) and include colonial relationship, common currency, common membership in the WTO, common official language, same country in the past, 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 A.4: Main Regression Results with 2-Digit Data

I. Extensive Margin

			Positive Trade Flow		
Mean of dep. variable			0.685		
	(1)	(2)	(3)	(4)	(5)
log Genetic Distance	-0.100***			-0.036***	-0.034***
	(0.000)			(0.000)	(0.000)
log Geodesic Distance		-0.132***	-0.110***	-0.088***	-0.075***
		(0.000)	(0.001)	(0.001)	(0.001)
Common Border			0.145***	0.139***	0.106***
			(0.002)	(0.002)	(0.002)
Controls	-	-	yes	yes	HMR
Destination-Com. FE	yes	yes	yes	yes	yes
Origin-Com. FE	yes	yes	yes	yes	yes
Observations	2,025,702	2,025,702	2,025,702	2,025,702	2,025,702
R-squared	0.516	0.536	0.539	0.541	0.545

II. Intensive Margin

			Trade Value		
Mean of dep. variable			200.52		
	(1)	(2)	(3)	(4)	(5)
log Genetic Distance	-0.608***			-0.062***	-0.042***
	(0.013)			(0.013)	(0.013)
log Geodesic Distance		-1.021***	-0.743***	-0.702***	-0.724***
		(0.013)	(0.025)	(0.027)	(0.027)
Common Border			0.714***	0.719***	0.512***
			(0.036)	(0.036)	(0.039)
Controls	-	-	yes	yes	HMR
Destination-Com. FE	yes	yes	yes	yes	yes
Origin-Com. FE	yes	yes	yes	yes	yes
Observations	209,958	209,958	209,958	209,958	209,958

Note: The top part of the table shows the result of five 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. Both origin- and destination-commodity fixed effects (FE) are included in all regressions. The data comprises bilateral trade flows from the year 2000 a the 2-HS-digit level. The PPML regression analyses only use a 50 percent random sample of the data due to computing constraints. Control variables include being an island, being landlocked, access to the same sea, difference in latitude, difference in longitude. Additional control variables follow Helpman, Melitz and Rubinstein (2008) (HMR) and include GDP per capita of exporter and importer, total population, colonial relationship, common currency, common membership in the WTO, common official language, same country in the past, 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 B: Description of Ocean Data Set

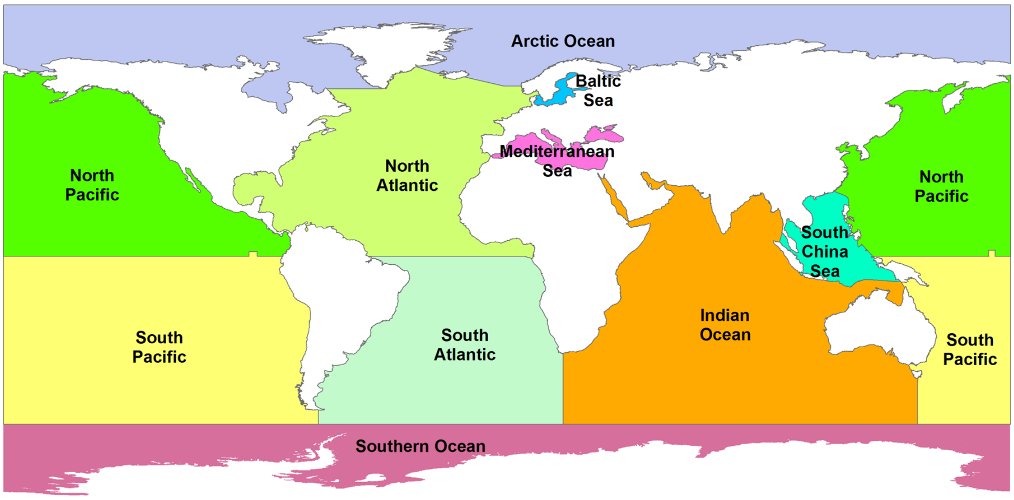
For our analysis, we created a variable indicating whether two countries have access to the same sea. Here, we explain how we have built this measure. First, we selected a list of ten major seas according to the classification by Eakins and Sharman (2010). The authors modified the list of world-seas 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.

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: a variable from the EIA Database (August 10, 2015) by Jeffrey Bergstrand indexing the amount of trade openness, on a scale 1 to 6, between country pairs.

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.

Trust

As described in the text, we use data from the Eurobarometer 46 survey to measure bilateral trust. Our sample includes fifteen countries in which a representative group was asked how much they trust people from the following 25 countries: Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, Luxembourg, Netherlands, Norway, Poland, Portugal, Russia, Slovakia, Spain, Sweden, Switzerland, Turkey, United Kingdom, and United States. For each country-pair, we have the mean response calculated by applying the coefficients 4, 3, 2 and 1 respectively to the various answer codes: a lot of trust, some trust, not very much trust, or no trust at all. Note that "no reply" answers are excluded from the calculation. The mid point is then 2.50: below this level, the negative answers predominate and above, the positive ones. In our sample of 360 observations, the average level of trust is given by 2.63 with a standard deviation of 0.41 and 1.27 at the minimum and 3.65 at the maximum.