I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

Ancestral Characteristics of Modern Populations

Paola Giuliano
Nathan Nunn

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11528

Ancestral Characteristics of Modern Populations

Paola Giuliano
University of California Los Angeles, CEPR, NBER and IZA

Nathan Nunn
Harvard University, NBER and BREAD

MAY 2018

ABSTRACT

Ancestral Characteristics of Modern Populations*

We construct a database, with global coverage, that provides measures of the cultural and environmental characteristics of the pre-industrial ancestors of the world's current populations. In this paper, we describe the construction of the database, including the underlying data, the procedure to produce the estimates, and the structure of the final data. We then provide illustrations of some of the variation in the data and provide an illustration of how the data can be used.

Corresponding author:
Paola Giuliano
Anderson School of Management
University of California, Los Angeles
Los Angeles, CA 90095
USA

E-mail: paola.giuliano@anderson.ucla.edu

1. Introduction

It is now recognized that much of contemporary economic underdevelopment is rooted in history. Evidence has been put forth showing that much of the variation in current economic performance can be explained by historical shocks like colonial rule, forced labour, and the slave trade (Acemoglu, Johnson and Robinson, 2001, Banerjee and Iyer, 2005, Dell, 2010, Nunn, 2008, Michalopoulos and Pappaioannou, 2011). A large number of studies document a remarkable amount of persistence over time, whether one examines economic prosperity, technology, political development, or cultural traits (Comin, Easterly and Gong, 2010, Bockstette, Chanda and Putterman, 2002, Putterman and Weil, 2010, Michalopoulos and Pappaioannou, 2013, Voigtlaender and Voth, 2012, Spolaore and Wacziarg, 2013).[1]

We contribute to this line of research by providing a publicly-accessible database that measures the economic, cultural, political, and environmental characteristics of the ancestors of current population groups.[2] Specifically, we construct measures of the average pre-industrial characteristics of the ancestors of the populations in each country of the world. The database is constructed by combining pre-industrial ethnographic information for approximately 1,300 ethnic groups with information on the current distribution of approximately 7,500 language groups measured at the grid-cell level. We link the ancestral characteristics data with current populations using the languages and dialects spoken. We implicitly assume that the ancestral traits will be transmitted in a manner that is correlated with the transmission of language, which is itself is an important vertically transmitted trait.

The primary source of ethnographic information is the *Ethnographic Atlas*, which provides information on the pre-industrial characteristics of 1,265 ethnic groups.[3] One shortcoming of the sample from the *Ethnographic Atlas* is that European groups are significantly under-represented. This is not because information about these cultures was not available, but because writing had existed for centuries among these groups, a study of the pre-industrial characteristics of these societies was seen as falling within the field of history rather than anthropology. We attempt to

---

[1]Although Acemoglu, Johnson and Robinson (2002) show evidence of a reversal of fortunes among former European colonies, once one examines continuity at the level of societies rather than at the level of geography, then one again observes strong persistence (Putterman and Weil, 2010).

[2]The database is posted on the authors' webpages. Although the url may change over time, currently the database can be accessed at: `https://scholar.harvard.edu/nunn/pages/data-0`.

[3]This source has been widely used in the political economy, economic history, and cultural economics literatures (e.g., Gennaioli and Rainer, 2007, Alesina, Giuliano and Nunn, 2013, Michalopoulos and Pappaioannou, 2013).

correct for this in our database by drawing on three additional sources. The first two sources are data collections, made subsequent to the *Ethnographic Atlas*, that are meant to be appended to the Atlas. One includes 17 ethnic groups from Eastern Europe and the other includes 10 ethnic groups from Siberia (Bondarenko, Kazankov, Khaltourina and Korotayev, 2005, Korotayev, Kazankov, Borinskaya, Khaltourina and Bondarenko, 2004). The third additional source is taken from the *World Ethnographic Sample*, which was assembled by George Peter Murdock (1957a). The sample comprises 565 ethnic groups. Among these, 17 observations, which include many European groups, do not appear in the *Ethnographic Atlas*. We also use this information.

We create three versions of our database. The first uses the standard *Ethnographic Atlas* only. The second also uses data from Bondarenko et al. (2005) and Korotayev et al. (2004). The third uses all available data sources, including the additional ethnic groups from the *World Ethnographic Sample*.

In what follows, we provide a detailed description of the dataset, first describing the underlying data, the procedure used to construct the dataset, and the structure of the final database. We then provide illustrations of some of the variation in the data. We end by providing one empirical exercise that illustrates how the data can be used.

2. Data Construction

A. *Ethnographic Atlas*

The primary data source for our database is the *Ethnographic Atlas*, a world-wide ethnicity-level database constructed by George Peter Murdock that contains ethnographic information on the pre-industrial characteristics of 1,265 ethnic groups.[4] The information has been coded for the earliest period for which satisfactory ethnographic data are available or can be reconstructed. The earliest observation dates are for groups in the Old World where early written evidence is available. For the parts of the world without a written history, the information is from the earliest observers of these cultures, which for some is as late as the 20th century. However, even for these cultures, the data capture as much as possible the characteristics of the ethnic group prior to European contact. For all groups in the dataset, the variables measure characteristics of the

---

[4]The digitized version of Murdock's *Ethnographic Atlas* was released in 1999. The release included 1,267 ethnic groups. However, two ethnic groups appear twice (Chilcotin and Tokelau). Thus, the Atlas includes 1,265 different ethnic groups. For a summary of the life's work of George Peter Murdock, including the *Ethnographic Atlas*, see Spoehr (1985).

societies prior to industrialization. In total, 23 ethnicities are observed during the 17th century or earlier, 16 during the 18th century, 310 during the 19th century, 876 between 1900 and 1950, and 31 after 1950. For nine ethnicities an exact year is not provided.

Although the Ethnographic Atlas is the best and most comprehensive source of global cross-cultural information, it is not without its shortcomings. First, as mentioned, ethnic groups are sampled in different periods of time. Given that most characteristics, within a group, generally remain fairly stable over time, this is not something that prohibits use of the database. However, it is a shortcoming. A second shortcoming, that we discuss in more detail below, is that groups with a written history – namely, European groups – are under-sampled in the database.

B. *Additional Ethnographic Sources*

We supplement the ethnographic data using three additional samples, which help to more completely cover ethnic groups from Europe. Following the release of the *Ethnographic Atlas*, a number of researchers have undertaken work which extends the work of Murdock by including ethnic groups that are missing from his sample. According to Korotayev et al. (2004), one shortcoming of the Atlas is that it does not adequately cover ethnic groups of the former Soviet Union. They attribute this to language barriers since the ethnographic sources are published in Russian. Thus, the authors used Murdock's same procedure to construct a dataset for ten Siberian ethnic groups. A similar initiative was published in 2005 by Bondarenko et al. (2005), but covering seventeen ethnic groups from Eastern Europe. Both groups are measured in the late 19th century. The two sources help greatly the under-representation of the *Ethnographic Atlas* for Eastern European ethnic groups. The sources provide information for important groups, such as the Bashkirs, Estonians, Latvians, and Moldovans. The seventeen ethnic groups that are included in the two additional samples are reported in the first two columns of Table 1.

A final sample that we use to supplement the *Ethnographic Atlas* provides seventeen additional ethnic groups, many of which are from Western Europe. In 1957, prior to the construction of the *Ethnographic Atlas*, George Peter Murdock constructed the *World Ethnographic Sample*, which was published in *Ethnology* (see Murdock, 1957b). Most of the 565 ethnic groups from the *World Ethnographic Sample* later appeared in the *Ethnographic Atlas*, but seventeen ethnic groups did not. These were ethnic groups for which information was more limited. If they had been included in

Table 1: Additional ethnic groups used to supplement the *Ethnographic Atlas*.

Peoples of Siberia	Peoples of Easternmost Europe	World Ethnographic Sample
Nganasan	Bashkir	French (Provence)
Itelmen	Besermyan	Italians (Sicily)
Ungazikmit	Chuvash	English
Mansi	Erzia Mordva	Germans (Prussia)
Evenk	Estonians	Danes (Lolland)
Negidal	Gagauz	Finns
Ulch	Ingrians	Argentinians
Orok	Karelians	Tajik (Mountain)
Oroch	Kazan Tatar	Nepalese (Kiranti)
Udihe	Latvians	Bondo
	Lithuanian Karaim	Aeta (Bataan)
	Lithuanian Tatar	Tagalog
	Livs	Maanyan (Siong)
	Moldovans	Jamaicans
	Udmurt	Cawahib
	Veps	Apalai
	Votes	Bagielli

the *Ethnographic Atlas*, they would have had a number of variables with missing values. Including this source increases the sample to 1,309 ethnic groups.

Although the number of ethnic groups added from the *World Ethnographic Sample* is limited, the additions are particularly important due to the size and importance of the groups that are added. As reported in Table 1, the source provides observations from important Western European groups that are missing from the *Ethnographic Atlas* like the French, Sicilians, English, Lollanders (Danes), Finns, and Prussians.

We construct three versions of the Ancestral Characteristics Database. One using only the original *Ethnographic Atlas*, a second that also uses the Eastern European and Siberian samples from Bondarenko et al. (2005) and Korotayev et al. (2004), and a third that adds to this the additional ethnic groups from the *World Ethnographic Sample*.

C. *Linking Ancestral Characteristics to Populations Today*

We link the ancestral characteristics from the ethnographic samples to current population distributions using the 16th edition of the *Ethnologue: Languages of the World* (Gordon, 2009), a data source that maps the current geographic distribution of over 7,000 different languages and
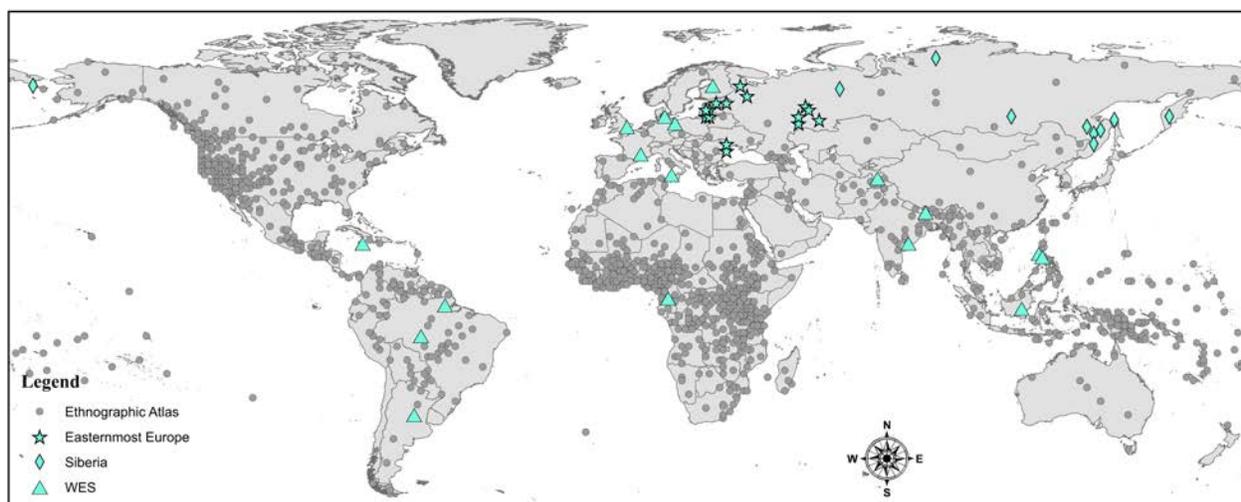
Figure 1: Approximate location of centroids of ethnic groups in the Ethnographic Atlas, as well as the Siberia, Easternmost Europe, and World Ethnographic Samples.

dialects, each of which we manually matched to one of the ethnic groups from the ethnographic data sources.

The *Ethnologue* provides a shape file that divides the world's land into polygons, with each polygon indicating the location of a specific language/dialect as of the date of publication. The raw *Ethnologue* shapefile had to be cleaned to make it functional for our use. First, the original file had some polygons that were partially or fully overlapping. Thus, some locations were assigned multiple languages. We created a shapefile that had mutually exclusive, non-overlapping polygons. When choosing between multiple polygons for a location, we assigned location the larger of the two polygons (based on land area). Second, the original file had 'slivers' – namely, small and narrow polygons that are created due to imprecisions in mapping. These were removed. The final cleaned Ethnologue shapefile is shown in Figure 2.

We combine the cleaned *Ethnologue* shapefile with data on the global distribution of the world's populations taken from the *Landscan 2007* database. The source reports estimates of the world's population in 2007 for 30 arc-second by 30 arc-second (roughly 1km by 1km) grid-cells globally. The database is produced by Oakridge Laboratories in cooperation with the U.S. Government and NASA. Combining these two sources of data provides an estimate of the distribution of populations' mother-tongues and, hence, the ancestral characteristics of populations across the globe today at a 1-km resolution.

By combining these data sources, we are able to construct country-level estimates of the average ancestral characteristics of populations from each modern country. From the ethnographic
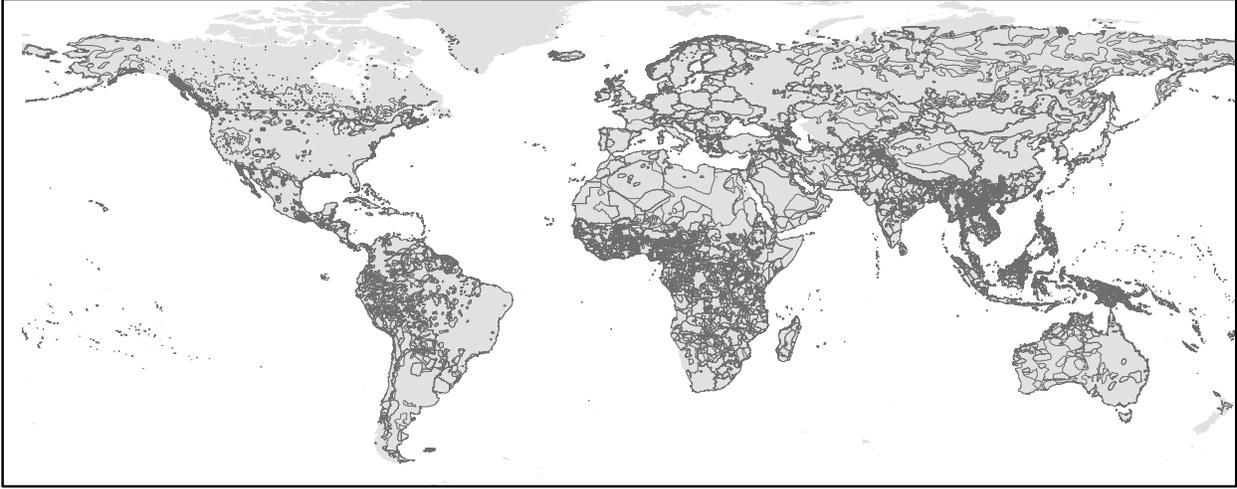
Figure 2: Language and dialect groups from Ethnologue 16.

sources, we know whether each ethnic group had a specific trait p historically. We define I_e^p to be an indicator variable that is equal to one if ethnic group e has characteristic p and zero otherwise. By matching each of the approximately 7,500 *Ethnologue* language polygons (i.e., a language group in a particular location) to one of the approximately 1,300 ethnic groups in from the ethnographic sources, we can determine whether the ancestors of each language group had trait p. We thus have an estimate of the distribution of trait p among individuals across the world, observed at a 1km grid-cell resolution. We combine this with information about the modern country borders to construct location-level averages of the prevalence of trait p among the ancestors of people living in each country.

To be more precise, let $N_{e,i,c}$ denote the number of individuals of ethnicity e living in grid-cell i located in country c. We construct a population-weighted average of I_e^p for all ethnic groups living in country c. Thus, the measure of the fraction of the population with ancestors with a particular characteristic p is given by:

$$\overline{I}_c^p = \sum_e \sum_i \frac{N_{e,i,c}}{N_c} \times I_e^p \tag{1}$$

where N_c is the total number of people living in country c.

D. *The Final Data*

a. Ethnographic Variables

The final database includes all variables that are present in the *Ethnographic Atlas*. The entries in the database use a variant of variable definitions from the original database. The variables in the original database are named $v1$, $v2$, etc. For example, variable $v33$ in the database is a variable that measures the level of jurisdictional hierarchy beyond the local community, which is commonly used as a measure of state centralization and state development (e.g., Nunn, 2007, Gennaioli and Rainer, 2007, Michalopoulos and Pappaioannou, 2013). The values of variable $v33$ take on integer values that indicate one of each of the following six categories: (1) the entry for an ethnicity is missing, (2) there are zero levels of authority beyond the local community, (3) there is one level, (4) there are two levels, (5) there are three levels, (6) there are four levels. In the Ancestral Characteristics database, the information on the levels of political authority of the ancestors of a country's population is represented by six different variables. These are named: $v33\_grp1$, $v33\_grp2$, $v33\_grp3$, $v33\_grp4$, $v33\_grp5$, and $v33\_grp6$. Each variable reports the fraction of a country's population that was connected to an ancestral ethnic group with a particular characteristic. For example, variable $v33\_grp1$ reports the fraction of a country's population with ancestors for which data on jurisdictional hierarchy is missing. Variable $v33\_grp2$ reports the fraction of a country's population with ancestors that had zero levels of jurisdictional hierarchy beyond the local community. Variable $v33\_grp3$ reports the fraction of a country's population with ancestors that had one level of jurisdictional hierarchy. Variable $v33\_grp6$ reports the fraction of a country's population with ancestors that had four levels of jurisdictional hierarchy.

Thus, if a researcher wanted to calculate the fraction of each country's population (with non-missing ancestral data) with ancestors that had more than one level (i.e., two, three or four levels) of jurisdictional hierarchy beyond the local community, the following calculation would be made:

$$\frac{v33\_grp4 + v33\_grp5 + v33\_grp6}{1 - v33\_grp1}$$

The denominator is the fraction of the population for which data on this ancestral characteristic is not missing. The numerator is the fraction of the population with ancestors that had two ($v33\_grp4$), three ($v33\_grp5$) or four ($v33\_grp6$) levels of jurisdictional hierarchy. Of course, the

same variable could alternatively have been calculated with

$$\frac{1 - (v33\_grp1 + v33\_grp2 + v33\_grp3)}{1 - v33\_grp1}$$

If one wanted to calculate the average levels of jurisdictional hierarchy beyond the local community of a country's ancestors (among observations without missing data), this can be calculated as follows:

$$\frac{v33\_grp2 \times 0 + v33\_grp3 \times 1 + v33\_grp4 \times 2 + v33\_grp5 \times 3 + v33\_grp6 \times 4}{1 - v33\_grp1}$$

b. Geographic Variables

The ethnographic sources also provide some information about the geography of each ethnic group's tradition environment. Variables *v95* and *v96* classify the environment of each ethnic group into 14 different categories. Examples of these include: tundra, desert, temperate woodland, sub-tropical bush, etc. However, this information is missing for the vast majority of ethnic groups. For example, data are missing for 867 of the 1265 ethnic groups in the *Ethnographic Atlas*.

For all ethnic groups in our sources, information on the latitude and longitude of the group, reported at integer values of degree, is provided. This can be combined with information about various geographic characteristics to construct attributes of ancestral lands, namely crop suitability, ruggedness, or distance to the coast. In choosing which environmental variables to include, we restricted our focus to those that are by-and-large time invariant. The environment variables that we construct, along with the source of the environmental data are provided below.

- **Climate Zones:** Koppen Geiger climate classification. There are 39 possible categories that are constructed based on climate observed from 1901–1925. The underlying data used to construct this is at a 0.5-degree resolution and is taken from Rubel and Kottek (2010). These are variables *KG_code_11* to *KG_code_62* in the database.

- **Ruggedness:** Terrain ruggedness index (TRI), measured in hundreds of meters. The underlying data are at a 30-arc-second resolution and are from Nunn and Puga (2012). This is variable *avgrug* in the database.

- **Distance to Coast:** Distance to the nearest ice-free coastline, measured in meters. The variable is constructed using information on the ice-free coastline available from Nunn and Puga (2012). This is variable *dist_coast* in the database.

3. Illustrations of the data

A. *Ethnographic Characteristics*

As we have discussed, one of the most widely used variables from the *Ethnographic Atlas* is the level of jurisdictional hierarchy beyond the local community. The variable, which is generally interpreted as a measure of political centralization or political sophistication, measures the level of political authority when one moves beyond the local authority. For example, if the local village chief is the highest level of authority, and he or she does not answer to anyone above them, then the variable would take on a value of zero. If above the chief, there was a district leader, then above this, there was a territory leader, and above this a provincial leader, and above this the paramount chief, then this variable would take on the value of four. The variable has been used by Nunn (2008), Gennaioli and Rainer (2007), and Michalopoulos and Pappaioannou (2013) to measure pre-colonial political centralization in Africa.

In the previous section, we described how the data can be used to calculate the average levels of political authority (beyond the local community) among ancestors of the inhabitants of each country.[5] We display this visually in Figure 3.

Another commonly used variable is the measure of complexity of settlements, which is variable *v30* from the *Ethnographic Atlas*. Ethnic groups are classified as belonging to one of the following nine categories: (1) missing data; (2) nomadic or fully migratory, (3) semi-nomadic, (4) semi-sedentary, (5) compact but not permanent settlements, (6) neighbourhoods of disperse family homesteads, (7) separate hamlets forming a single community, (8) compact and relatively permanent settlements and (9) complex settlements. This variable has been used by a number of scholars as a measure of traditional economic development by assigning each non-missing category an integer value from 1 to 8. Thus, the variable is increasing in settlement complexity (e.g., Alesina et al., 2013, Michalopoulos and Pappaioannou, 2013, Giuliano and Nunn, 2016).[6] We display the average of this variable across countries in Figure 4.

---

[5] The formula for this is: $\frac{v33\_grp2 \times 0 + v33\_grp3 \times 1 + v33\_grp4 \times 2 + v33\_grp5 \times 3 + v33\_grp6 \times 4}{1 - v33\_grp1}$.

[6] This is calculated as: $\frac{v30\_grp2 \times 1 + v30\_grp3 \times 2 + v30\_grp4 \times 3 + v30\_grp5 \times 4 + v30\_grp6 \times 5 + v30\_grp7 \times 6 + v30\_grp8 \times 7 + v30\_grp9 \times 8}{1 - v30\_grp1}$.
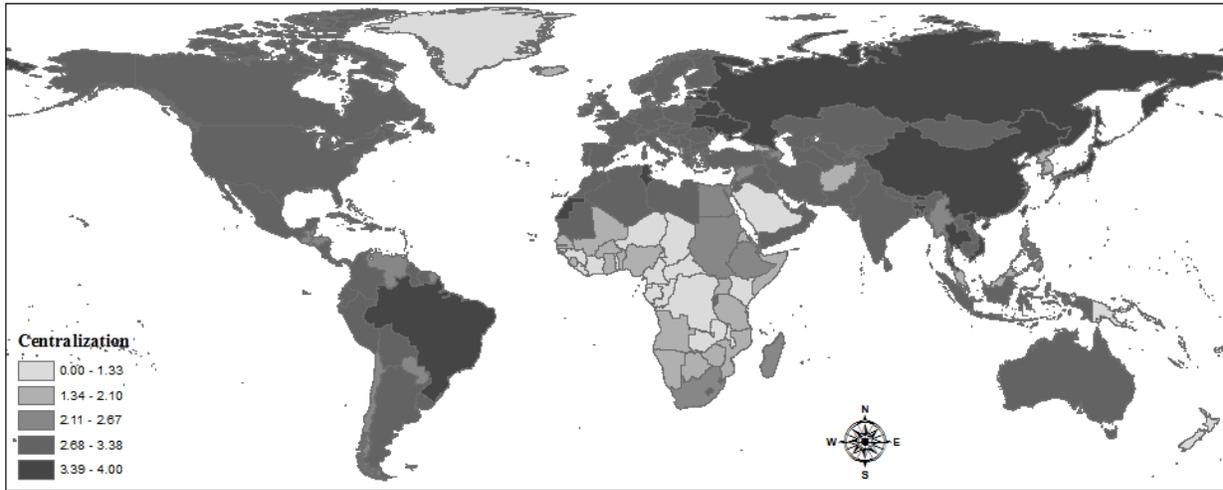
Figure 3: Average levels of jurisdictional hierarchy beyond the local community among each countries ancestors, 0-4: variable *v33*
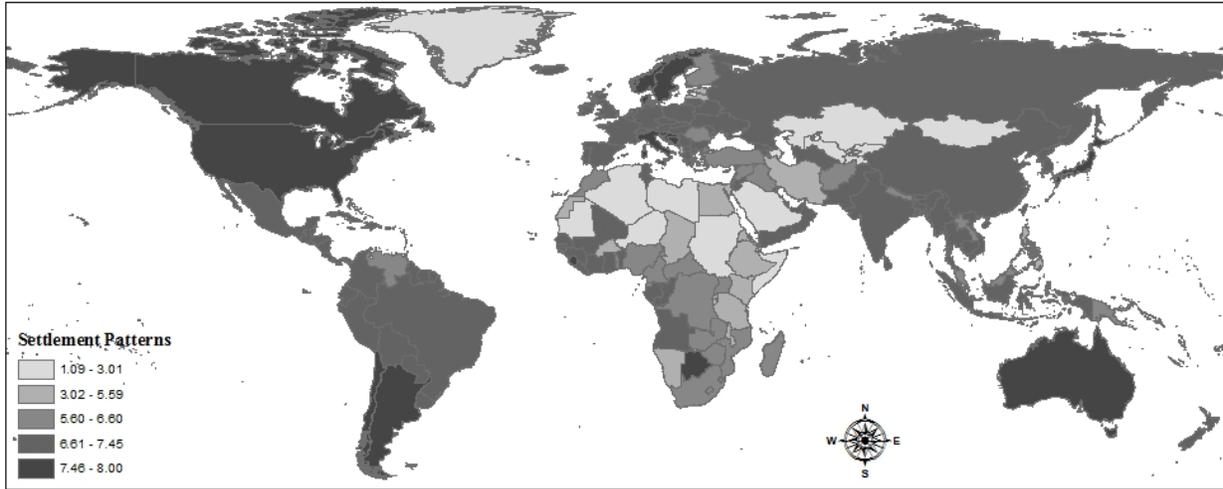

Figure 4: Average complexity of settlement index score among each country's ancestors, 1-8: variable *v30*

Another cultural practice that has received considerable attention in recent research is the practice of bride price, which is a transfer of money and/or other valuable assets that is made at marriage from the groom and/or his parents to the bride's parents. The importance of this tradition for female educational investments has recently been studied by Ashraf, Bau, Nunn and Voena (2017). Corno and Voena (2016) and Corno, Hildebrandt and Voena (2017) study the relationship between the practice of bride price, weather shocks, and age of marriage, while Lowes and Nunn (2017) examine the relationship between the value of the bride price that was paid at marriage and the wellbeing of the wife. Variable *v6* categorizes the marriage customs of pre-industrial societies into the following eight groups: (1) Missing data; (2) *Bride price*, which is also known as bride wealth and is a transfer of a substantial consideration in the form of goods, livestock, or money from the groom or his relatives to the kinsmen of the bride; (3) *Token bride price* is a small or symbolic payment only; (4) *Bride service*, which is a substantive material consideration in which the principal element consists of labour or other services rendered by the groom to the bride's kinsmen; (5) *Gift exchange*, which is a reciprocal exchange of gifts of substantial value between the relatives of the bride and groom, or a continuing exchange of goods and services in approximately equal amounts between the groom or his kinsmen and the bride's relatives; (6) *Female relative exchange*, which is a transfer of a sister or other female relative of the groom in exchange for the bride; (7) *Dowry*, which is a transfer of a substantial amount of property from the bride's relatives to the bride, the groom, or the kinsmen of the latter; and (8) *No significant consideration*, which is an absence of any significant consideration, or giving of bridal gifts only.

The extent to which the ancestors of each country's population practiced bride price is shown in Figure 5. The variable that is mapped is $\frac{v6\_grp2}{1-v6\_grp1}$, where $v6\_grp2$ is the fraction of a population's ancestors who traditionally practiced bride price and $v6\_grp1$ is the fraction of a population's ancestors with missing information for this variable. One can see that the practice was most common among the ancestors of populations living in Africa, the Middle East, and Asia today. One can also see that the custom was practiced by some populations but not all in Zambia and Indonesia. This is the within country variation that is exploited by Ashraf et al. (2017).

The database can be used to calculate the total fraction of the world's current population that have ancestors that traditionally practiced bride price. Doing this, we find that the figure is 33.6%. By comparison, a dowry was the traditional marriage practice among the ancestors of 27.1% of the world's current population. Thus, perhaps surprisingly, bride price, which tends to
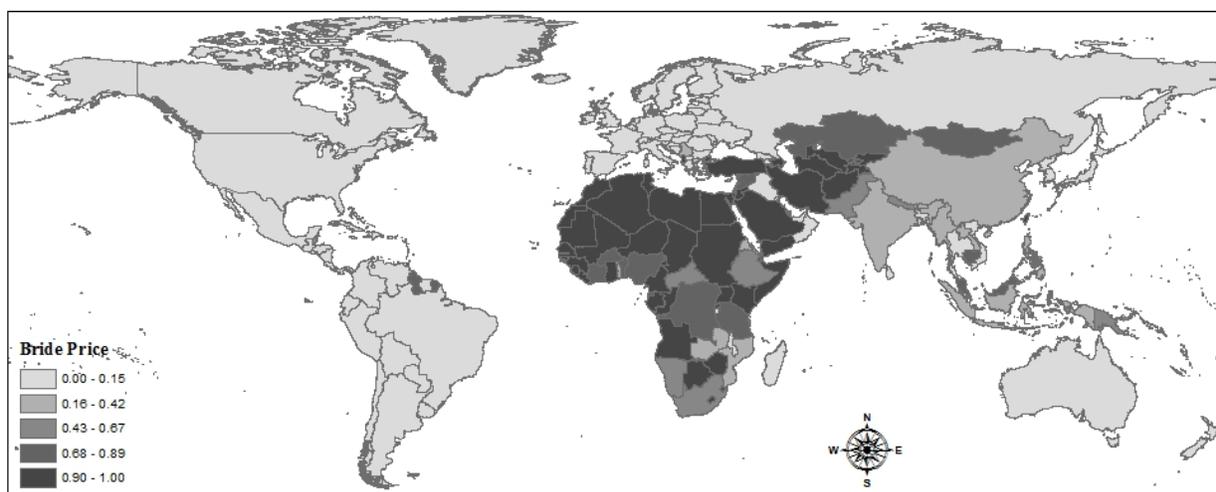
Figure 5: Fraction of a country's population with ancestors that practiced bride price at marriage: variable *v6*

be well-known to general researchers, is actually the more common of the two forms of marriage transfer.

In recent research, Lowes (2017) studies how matrilineal kinship affects spousal cooperation in the southern Democratic Republic of the Congo using a series of behavioral games played with husbands and wives. The key hypothesis is that the structure of broader kinship systems may affect incentives for spouses to cooperate by changing the nature of the relationship between spouses and their extended family members and the extent of husbands' control over their wives. She finds that although matrilineal kinship is associated with less cooperation within the household, women and children seem to benefit in matrilineal systems. Women experience less domestic violence and report greater agency in decision-making, and children of matrilineal women are healthier and better educated. In a follow up study in the same setting, Lowes (2018) tests whether matrilineal kinship closes the gender gap in willingness to compete. In contrast to earlier work from Gneezy, Leonard and List (2009), she does not find any evidence of this.

The fraction of each country's population (without a missing measure) with ancestors that traditionally had matrilineal inheritance (as opposed to patrilineal or cognatic descent) is reported visually in Figure 6. One can see clearly the concentration of the tradition of matrilineal inheritance in Central and Southern Africa around Africa's 'Matrilineal Belt', which is the focus of Lowes (2017, 2018). The tradition can also be seen among the ancestors of those in West Africa, which is the location of the studies by Ferrara (2007) and Ferrara and Milazzo (2017).
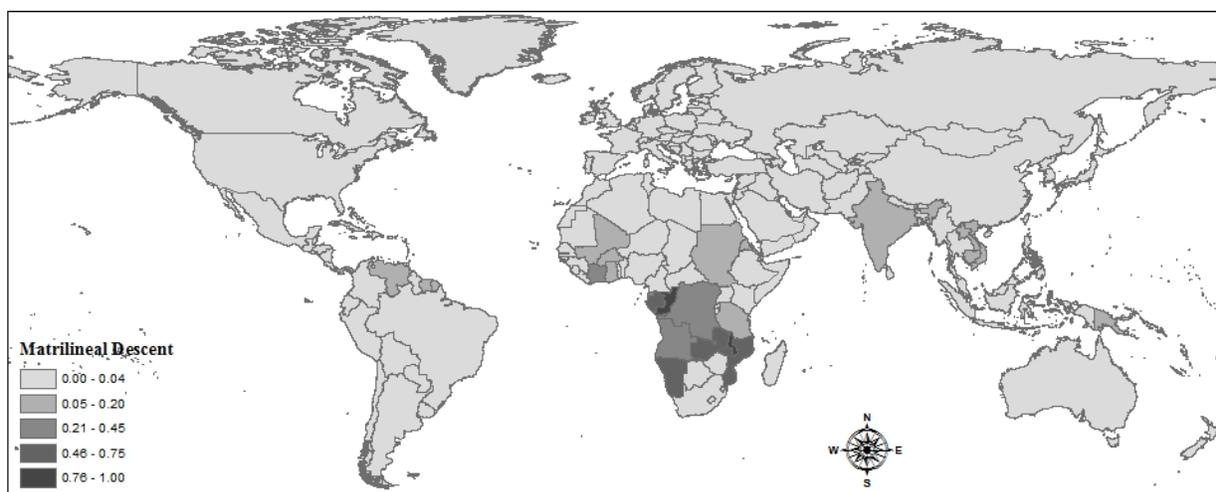
Figure 6: Fraction of a country's population with matrilineal ancestors: variable $v43$

A number of recent studies have also examined other aspects of lineage customs. Benjamin Enke (2017) measures kinship "tightness" using an index constructed using principal components of variables from our database, and studies the relationship between it and a host of economic, psychological and political outcomes. Jonathan Schulz (2017) examines the importance of cousin marriage, which typically coincides with strong kinship ties. He shows that a greater prevalence of cousin marriage today is associated with less democracy and more corruption. Akbari, Bahrami-Rad and Kimbrough (2017) also examine the relationship between cousin marriage and corruption. Moscona, Nunn and Robinson (2017) examine the effects of one particular lineage organization – segmentary lineage systems. They show that, consistent with anthropological accounts, the system is associated with more conflict today.

Motivated by this line of research, we display the cross-country distribution of a traditional preference for cousin marriage in Figure 7. The measure is constructed using variables $v25\_grp1$–$v25\_grp16$. The first category of the set (i.e., $v25\_grp1$) is for missing data. There are then 14 categories for different types of preferred cousin marriage: $v25\_grp2$–$v25\_grp15$. The final variable is $v25\_grp16$, which is for the absence of a preference for cousin marriage. Using these variables, we construct a measure of the proportion of the ancestors of each country (with non-missing data) that traditionally preferred cousin marriage to non-cousin marriage: $\frac{1-v25\_grp1+v25\_grp16}{1-v25\_grp1}$. As Figure 7 shows, cousin marriage has traditionally been the preferred form of marriage among the ancestors of populations from Africa, the Middle East, and Asia. Schulz (2017) has shown that the practice is still common today in much of the world and is associated with less democracy
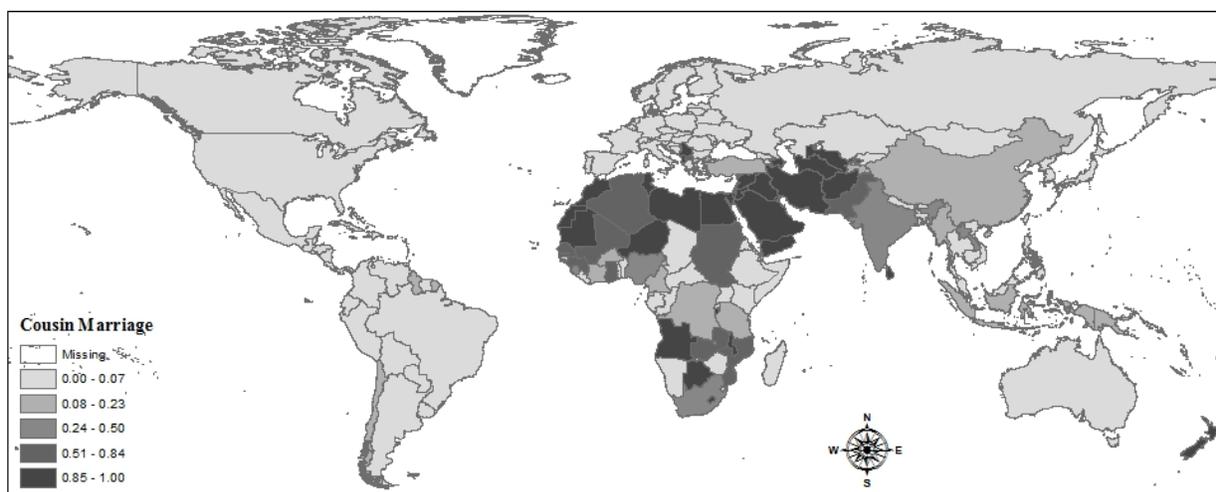
Figure 7: Fraction of a country's population with ancestors that preferred cousin marriage: variable *v25*

and more corruption.

a. Geographic characteristics

We now turn to a discussion of the geographic variables of the database. An ongoing question in the historical development literature is the importance of geography, and how much of its importance is due to its contemporaneous effects relative to its importance due to its historical effects. A large number of studies show that geography can have important effects on current outcomes through historical channels. For example, Alsan (2015) documents the importance of historical climatic conditions on contemporary development working through the historical importance of the TseTse fly. Alesina et al. (2013) show that geo-climatic suitability of the environment for growing specific types of cereals is strongly correlated with views about appropriate gender roles today, not because of a contemporary effects, but because these geographic characteristic affected which cereals could be cultivated historically, which affected whether the plough was adopted and whether women regularly participated in agriculture. Similarly, Nunn and Puga (2012) provide evidence that within Africa ruggedness exerted a positive impact on income today by allowing societies to escape from historical slave trades and their detrimental effects.

Given this mounting evidence, a natural question arises. In general, how important are the historical effects of geography relative to its contemporary effects? Our database, by providing information on the historical environment faced by the ancestors of different populations, can be

Table 2: Contemporary and Ancestral Geographic Characteristics and Contemporary Economic Development

	(1)	(2)	(3)	(4)
	Dep var: Log of real per capita GDP in 2000			
Distance from the Equator				
Contemporary	0.00724			
	(0.0106)			
Ancestral	0.0324***			
	(0.0108)			
Terrain Ruggedness				
Contemporary		0.147	-0.0270	
		(0.0918)	(0.0869)	
Ancestral		-0.380***	-0.449***	
		(0.128)	(0.112)	
Distance from Coast				
Contemporary				0.418
				(0.477)
Ancestral				-1.817***
				(0.577)
Constant	7.443***	8.685***	9.418***	8.961***
	(0.141)	(0.144)	(0.143)	(0.113)
Observations	168	167	119	167
R-squared	0.322	0.051	0.172	0.167

Notes: The unit of observation is a country. Robust standard errors are reported in parentheses. ***, **, and * indicate significance at the 1, 5, and 10% levels.

used to make progress on this question. To illustrate this, we consider three ancestral geographic characteristics: distance from the equator, distance from the coast, and terrain ruggedness. We also examine contemporary measures of these three characteristics, which are taken from Nunn and Puga (2012). We find that the ancestral and contemporary measures of the three geographic characteristics are highly, but not perfectly, correlated. The correlation coefficients between the ancestral and contemporary measures of distance from the equator, distance from the coast, and ruggedness are: 0.90, 0.90, and 0.53, respectively. The correlation between contemporary and ancestral terrain ruggedness is much lower than for distance from the equator or distance from the coast. This is not surprising given that terrain ruggedness can vary significantly over small distances, causing the ancestral and historical measures to differ. By definition, distance from the equator and distance to the coast vary smoothly across space.

We examine the differential ability of the ancestral and contemporary geographic measures to explain variation in countries' real per capita GDP (measured in 2000). The estimates are reported in Table 2. In column 1, we report estimates examining contemporary and ancestral distance from

the equator. Not surprisingly, being further from the equator is positively associated with real per capita GDP. However, what is more surprising is that the ancestral measure appears to be much more strongly correlated than the contemporary measure. This is particularly striking since we would expect the ancestral measure to be more imprecisely measured than the contemporary measure. Column 2 reports analogous estimates for terrain ruggedness. Given the existing evidence for a differential effect of ruggedness within African countries (Nunn and Puga, 2012), in addition to reporting estimates for all countries (column 2), we also report estimates for a sample that excludes countries from Africa (column 3). In both samples, we again find that the ancestral measure provides greater explanatory power than the current measure. The final column of the table reports estimates for distance from the coast (column 4). We find that being further from the coast is associated with lower incomes and that, again, the ancestral measure provides greater explanatory power than the current measure.

Overall, the estimates reported in Table 2, although simply exploratory, provide some indication that geography's greatest effects on current income (at least for the characteristics examined) may be through historical channels. This finding is consistent with a range of previous studies, which in a range of different settings, have identified the historical importance of geography. Examples of such studies include Diamond (1997), Michalopoulos (2012), Fenske (2014), Alsan (2015), and Henderson, Squires, Storeygard and Weil (2018).

4. Concluding comments

We have constructed a database, with global coverage, that provides measures of the cultural and environmental characteristics of the pre-industrial ancestors of the world's current populations. We have provided a detailed description of the dataset. We described the underlying data, the procedure used to construct the dataset, and the structure of the final database. We then provided illustrations of some of the variation in the data and one empirical exercise that illustrated how the data can be used.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 2001, *91*, 1369–1401.

_ , _ , **and** _ , "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution," *Quarterly Journal of Economics*, 2002, *117*, 1231–1294.

Akbari, Mahsa, Duman Bahrami-Rad, and Erik Kimbrough, "Kinship, Fractionalization, and Corruption," 2017. Working paper.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn, "On the Origins of Gender Roles: Women and the Plough," *Quarterly Journal of Economics*, 2013, *28* (2), 469–530.

Alsan, Marcella, "The Effect of the TseTse Fly on African Development," *American Economic Review*, 2015, *105* (1), 503–524.

Ashraf, Nava, Natalie Bau, Nathan Nunn, and Alessandra Voena, "Bride Price and Female Education," 2017. Mimeo, Harvard University.

Banerjee, Abhijit and Lakshmi Iyer, "History, Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in India," *American Economic Review*, 2005, *95* (4), 1190–1213.

Bockstette, Valeri, Areendam Chanda, and Louis Putterman, "States and Markets: The Advantage of an Early Start," *Journal of Economic Growth*, 2002, *7*, 347–369.

Bondarenko, Dmitri, Alexander Kazankov, Daria Khaltourina, and Andrey Korotayev, "Ethnographic Atlas XXXI: Peoples of Easternmost Europe," *Ethnology*, 2005, *44* (3), 261–289.

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

Corno, Lucia and Alessandra Voena, "Selling Daughters: Age of Marriage, Income Shocks, and the Bride Price Tradition," 2016. Mimeo, University of Chicago.

_ , **Nicole Hildebrandt, and Alessandra Voena**, "Age of Marriage, Weather Shocks, and the Direction of Marriage Payments," 2017. Mimeo, University of Chicago.

Dell, Melissa, "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 2010, *78* (6), 1863–1903.

Diamond, Jared, *Guns, Germs, and Steel*, New York: W.W. Norton & Company, 1997.

Enke, Benjamin, "Kinship Systems, Cooperation and the Evolution of Culture," 2017. Working paper, Harvard University.

Fenske, James, "Ecology, Trade, and States in Pre-Colonial Africa," *Journal of the European Economic Association*, 2014, *12* (3), 612–640.

Ferrara, Elianna La, "Descent Rules and Strategic Transfers. Evidence from Matrilineal Groups in Ghana," *Journal of Development Economics*, 2007, *83* (2), 280–301.

_ **and Annamaria Milazzo**, "Customary Norms, Inheritance, and Human Capital: Evidence from a Reform of the Matrilineal System in Ghana," *American Economic Journal: Applied Economics*, 2017, *9* (4), 166–185.

Gennaioli, Nicola and Ilia Rainer, "The Modern Impact of Precolonial Centralization in Africa," *Journal of Economic Growth*, 2007, *12* (3), 185–234.

Giuliano, Paola and Nathan Nunn, "Understanding Cultural Persistence and Change," 2016. Working paper, Harvard University.

Gneezy, Uri, Kenneth L. Leonard, and John A. List, "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society," *Econometrica*, 2009, *77* (5), 1637–1664.

Gordon, Raymond G., *Ethnologue: Languages of the World, 16th Edition*, SIL International, 2009.

Henderson, J. Vernon, Tim Squires, Adam Storeygard, and David Weil, "The Global Distribution of Economic Activity: Nature, History and the Role of Trade," *Quarterly Journal of Economics*, 2018, *133* (1), 357–406.

Korotayev, Andrey, Alexander Kazankov, Svetlana Borinskaya, Daria Khaltourina, and Dmitri Bondarenko, "Ethnographic Atlas XXX: Peoples of Siberia," *Ethnology*, 2004, *43*, 83–92.

Lowes, Sara, "Kinship Systems, Gender Norms, and Household Bargaining: Evidence from the Matrilineal Belt," 2017. Working paper, Harvard University.

⏤ , "Matrilineal Kinship and Gender Differences in Competition," 2018. Working paper, Bocconi University.

⏤ **and Nathan Nunn**, "Bride Price and the Wellbeing of Women," 2017. U.N. WIDER Working Paper 131/2017.

Michalopoulos, Stelios, "The Origins of Ethnolinguistic Diversity," *American Economic Review*, 2012, *102* (4), 1508–1539.

⏤ **and Elias Pappaioannou**, "The Long-Run Effects of the Scramble for Africa," 2011. NBER Working Paper 17620.

⏤ **and** ⏤ , "Pre-Colonial Ethnic Institutions and Contemporary African Development," *Econometrica*, 2013, *81* (1), 113–152.

Moscona, Jacob, Nathan Nunn, and James A. Robinson, "Social Structure and Conflict: Evidence from sub-Saharan Africa," 2017. Working paper, Harvard University.

Murdock, George Peter, "World Ethnographic Sample," *American Anthropologist*, 1957, *59* (4), 664–687.

⏤ , "World Ethnographic Sample," *Ethnology*, 1957, *59* (4), 664–687.

Nunn, Nathan, "Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade," *Quarterly Journal of Economics*, 2007, *122* (2), 569–600.

⏤ , "The Long-Term Effects of Africa's Slave Trades," *Quarterly Journal of Economics*, 2008, *123* (1), 139–176.

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

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

Rubel, Franz and Markus Kottek, "Observed and Projected Climate Shifts 1901–2100 Depicted by World Maps of the Koppen-Geiger Climate Classificaiton," *Meteorologische Zeitschrift*, 2010, *19* (2), 135–141.

Schulz, Jonathan F., "The Churches' Bans on Consanguineous Marriages, Kin-Networks and Democracy," 2017. Working paper.

Spoehr, Alexander, "George Peter Murdock (1897–1985)," *Ethnology*, 1985, *24* (4), 307–317.

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

Voigtlaender, Nico and Hans-Joachim Voth, "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany," *Quarterly Journal of Economics*, 2012, *127* (3), 1339–1392.