The Diffusion of Development

Enrico Spolaore

Tufts University

Romain Wacziarg*

Stanford University

NBER and CEPR

First draft: March 2005

This draft: November 2005

Abstract

This paper provides a framework relating the vertical transmission of characteristics across generations - such as the transmission of culture, language, norms and values - to differences in income per capita across countries. We propose a novel way to measure cultural barriers using genetic distance, i.e. coancestor coefficients from population genetics. Our econometric methodology addresses the problem of spatial correlation inherent in having pairwise income differences as our dependent variable. We find a significant effect of genetic distance on income differences, even when controlling for geographical distance, differences in latitude, and other cultural and geographic distance measures. These results hold not only for contemporary income differences, but also for income differences measured as of 1500 and 1700. We uncover similar patterns of coefficients for the proximate determinants of income differences (differences in human capital, institutions, investment rates and population growth). Our results suggest an important role for vertically transmitted characteristics in the diffusion of development.

1 Introduction

What explains the vast differences in income per capita that are observed across countries? What prevents poor countries from reaching the level of income of rich ones? In recent years a large empirical literature has addressed these questions using cross-country regressions, in which the level of development, measured by income per capita, is regressed on a set of determinants.[1] In this paper we propose a complementary approach. Our goal is to identify the determinants of *differences* in income per capita. Rather than addressing the question "what makes a country richer or poorer?" we address the related question "what barriers account for the *income gap* between two countries?".[2] We use income differences between pairs of countries as our dependent variable in order to investigate what measures of distance between countries are empirically relevant to account for differences in development.

Countries can be distant or close to each other along multiple dimensions. Measures of geographical distance (geodesic distance, differences in latitude, etc.) are relatively straightforward and easy to measure. By contrast, differences along cultural and historical dimensions have proven more elusive to define and measure, and have not been studied systematically. How can we tell whether two populations are close or far in terms of their long-term cultural and societal characteristics? Direct measures of language, religion, habits, values, etc., have been proposed and used by scholars, but a metric summarizing overall distance between populations along those dimensions has so far been lacking.

In this paper we propose a novel way of measuring overall distance in long-term historical and cultural characteristics, by exploiting the fact that the bulk of those characteristics are transmitted

---

[1] Recent contributions to this literature include Hall and Jones (1999), Acemoglu, Johnson and Robinson (2001), Easterly and Levine (2003), Alcalá and Ciccone (2004), Rodrik, Subramanian and Trebbi (2004), Glaeser, La Porta, Lopez-de-Silanes and Shleifer (2004) among others.

[2] There is a voluminous literature on cross-country income convergence, dating back to Baumol (1986). In the neoclassical literature, convergence occurs because the marginal return to capital is higher in countries farther from their steady-state, which depends, among other things, on the level of technology (the "A" parameter). These theories say little about what factors affect the level of technology or its growth rate. In contrast, we seek to characterize the factors that prevent or facilitate the diffusion of productivity-enhancing innovations across countries. In this respect, our paper is closer to the approach in Barro and Sala-i-Martin (1997), where technological diffusion drives convergence. A recent analysis of policy-induced constraints on the diffusion of technology is provided by Parente and Prescott (2002). However, we go beyond this by considering more broadly the barriers to the diffusion of fundamental technological and institutional characteristics in the long run, i.e. the prime determinants of development.

"vertically" across generations.[3] Hence, on average, the distance between two populations in the set of vertically transmitted characteristics (culture, habits, etc.) is a function of the distance in time from the two populations' last common ancestors.[4]. By looking at how populations differ in various DNA neutral markers, geneticists have provided estimates of genetic distance ("coancestor coefficients") that capture this degree of genealogical relatedness between populations.[5] A lower genetic distance means that two populations share more recent common ancestors, have had less time to diverge in their vertically transmitted characteristics, and are therefore more likely to share similar languages, habits, norms, preferences, traditions, and so on.[6] In this paper, we will use measures of genetic distance provided by Cavalli-Sforza et al. (1994) in order to capture the overall distance of populations in long-term vertically-transmitted characteristics, and will show how this distance can help explain differences in income per capita.[7]

Our approach does not imply that the vertical transmission of long-term characteristics is itself genetic. In fact, our approach is consistent with the view, shared by most geneticists and anthropologists, that the strong link between genetic distance and many important characteristics in human populations is mainly due to *cultural* (non-genetic) transmission across generations.[8]

---

[3]Evolutionary models of cultural transmission have been developed, by Cavalli-Sforza and Feldman (1981) and Boyd and Richerson (1985). Economic models of cultural transmission from parents to children have been provided by Bisin and Verdier (2000, 2001). Galor and Moav (2003) present an innovative theory of long-term economic growth in which a key role is played by evolutionary changes in preference parameters that are genetically transmitted across generations. For an in-depth discussion of these issues, see also Galor (2005).

[4]That is, how far two populations are from their "concestors," to use Richard Dawkins' (2004) terminology.

[5]"Neutral" genetic markers are those that have *no effect* on selective fitness. Ideally, measures of human genetic distance are calculated using neutral genetic markers. The classical reference on the neutral theory of molecular evolution is Kimura (1968). A recent textbook reference on human evlutionary genetics is Jobling et al. (2004).

[6]As we will see, in general the relationhsip between genetic distance and distance in the set of verticall-transmitted characteristics should not be expected to hold exactly, but *on average*. It is possible for two populations sharing a more recent common ancestor to show more "divergence" than two populations which are less closely related, but on average genetic distance and differences in vertically transmitted characteristics tend to go hand in hand.

[7]The only other economists who have used this variable, as far as we know, are Guiso, Sapienza and Zingales (2004), who use genetic distance between European populations as an instrument for a measure of trust in order to explain bilateral trade flows. This is quite different from our application, as we are interested in explaining income differences, not trade flows. Their results are consistent with our interpretation of genetic distance as related to cultural barriers.

[8]For example, see Cavalli-Sforza and Cavalli-Sforza (1995, chapter 8).

Language is an obvious example. There is no gene for speaking Japanese or Italian, but people who speak the same language on average tend to be closely related genetically, because most children learn their language from their parents.[9] In the theoretical part of this paper we present a simple framework that shows how the vertical transmission of (non-genetic) cultural characteristics across generations can be sufficient to explain a positive correlation between genetic distance and differences in income per capita.

More generally, in our theoretical section we provide a framework illustrating how vertically-transmitted characteristics affect income differences across countries. An important assumption of our model is that vertically-transmitted characteristics affect *barriers* to the (horizontal) transmission of technological and institutional innovations. For example, all other things being equal, people who share similar languages and cultures may be more likely to learn from each other. In our basic theoretical framework, different populations improve their productivity by adopting innovations (new technology, new institutions) directly or through imitation and adaptation of innovations from other societies. Technology and institutions are transmitted along with other cultural characteristics. Income differences across societies emerge and persist when there are barriers to the horizontal diffusion of innovations, related to distance in vertically-transmitted characteristics. In our theoretical part we also extend the framework to allow for a direct effect of cultural characteristics on the probability of achieving productivity-enhancing innovations, and provide a general discussion of the different channels linking vertically-transmitted characteristics to income differences.

In the empirical part of this paper we develop a novel methodology for estimation and inference

---

[9]For a discussion of the very close connection between genetic distance and linguistic classifications, see Cavalli-Sforza, Menozzi and Piazza (1994, pp. 96-105). While we also use differences in language directly in our regressions, we believe that genetic distance makes for a more comprehensive and informative measure of overall cultural distance, for several reasons: Genetic distance is *exogenous* with respect to incomes per capita, since individuals and entire populations may change their languages because of economic and political factors (conquests, globalization), but cannot change their genes (yet). Moreover, language or religion are only one among a complex web of vertically-transmitted characteristics that populations transfer across generations. Societies, even when they speak the same language, differ in numerous other cultural characteristics that affect their view of the world, values, attitudes, ability to interact and communicate with other groups, etc. Genetic distance, being correlated with the total set of vertically transmissible characteristics, seems a more promising *general* long-term measure of cultural and societal distance than any other available variable. An additional advantage is that genetic distance is measured as a continuous variable, while other linguistic and cultural classifications typically do not provide a continuous measure of distance.

with respect to the determinants of pairwise income differences across countries. We address a serious econometric problem arising naturally from our empirical strategy: since our dependent variable is the absolute value of the difference in income per capita between two countries, the covariance matrix of the residuals will feature spatial correlation: observations that include a common country will not be independent. We include a set of common-country fixed-effects to address this issue. We find a significant effect of coancestor coefficients (genetic distance) on income differences, even when controlling for geographical distance, differences in latitude, and other characteristics. This suggests an important role for vertically transmitted characteristics in the diffusion of development. Our results, moreover, hold not only for contemporary income differences, but also for income differences in 1500 and 1700.

In our empirical analysis we also examine directly the proximate determinants of income differences: differences in institutional outcomes (expropriation risk), population growth, human capital, investment rates and openness to trade. The same distance measures that account for differences in income levels also account for differences in some of these underlying variables, and generally in similar ways. This suggests that vertically-transmitted characteristics captured by genetic distance affect income differences through their effects on key proximate determinants of development, such as institutions, human capital, and population growth.

This paper is organized as follows. In Section 2 we present a simple analytical framework linking genetic distance (i.e., distance between two populations in numbers of generations from their last common ancestor), cultural distance, barriers to the diffusion of innovation, and differences in income per capita. Section 3 discusses the data and the empirical methodology. In Section 4 we present our empirical results, in which, consistently with our theoretical framework, genetic distance explains differences in income per capita and in its proximate determinants. Section 5 concludes.

2   The Diffusion of Development: A Simple Framework

2.1   The Basic Model

In this section we develop a framework to study the diffusion of technological and institutional innovations across societies. We sketch a simple model in which:

a) Innovations are transmitted vertically (across generations within a given population) and horizontally (across different populations).

b) The horizontal diffusion of innovations is not instantaneous, but is a function of barriers to technological and institutional diffusion.

c) Barriers to technological and institutional diffusion across societies are a function of how far societies are from each other as a result of divergent historical paths.

Productive knowledge is summarized by a positive real number A_{it}. We assume a linear technology $Y_{it} = A_{it}L_{it}$, where L_{it} is the size of the population, which implies that income per capita is given by $y_{it} \equiv Y_{it}/L_{it} = A_{it}$.

For simplicity, we summarize all other relevant characteristics of a society (cultural habits and traditions, language, etc.) as a point on the real line. That is, we will say that at each time t a population i will have "cultural" characteristics q_{it}, where q_{it} is a real number.[10] These characteristics are transmitted across generations with variations. Over time, characteristics change (vocabulary and grammar are modified, some cultural habits and norms are dropped while new ones are introduced, etc.). Hence, at time $t+1$ a population i will have different characteristics, given by:

$$q_{it+1} = q_{it} + \eta_{it+1} \tag{1}$$

where q_{it} are the characteristics inherited from the previous generations, while η_{it+1} denotes cultural change.

By the same token, the dynamics of productive knowledge includes vertical transmission across generations as well as changes (innovations), that is:

$$A_{it+1} = A_{it} + \Delta_{it+1} \tag{2}$$

where Δ_{it+1} denotes change in productivity due to technological and institutional innovations. Changes may take place because of original discovery by agents that belong to population i and/or because of successful imitation/adaptation of innovations that were discovered elsewhere. The diffusion of technological and institutional innovations can be viewed as a special case of cultural transmission.

We are interested in the long-run process of vertical and horizontal transmission of culture across populations at different genealogical distances from each other - that is, with different distances

---

[10]Of course, this is a highly simplified and reductive way of capturing cultural differences. In general, culture is a highly elusive and multi-faceted concept. In a well-known survey over fifty years ago Kroeber and Kluckhohn (1952) listed 164 definitions of culture proposed by historians and social scientists. See also Boyd and Richerson (1985).

from their last common ancestor. To capture these relationships in the simplest possible way, we will assume the following intergenerational structure. At time 0, there exists only one population, with cultural characteristics q_0 (normalized to zero) and productive knowledge A_0.[11] At time 1 the population splits in two distinct populations (population 1 and population 2). At time 2, population 1 splits in two populations (populations 1.1 and 1.2), and population 2 splits in two populations (populations 2.1 and 2.2). This structure provides us with the minimum number of splits we need to have variation in genealogical distances between populations at time 2. We can measure genealogical distance between populations by the number of genealogical steps one must take to reach the closest common ancestor population. Let $d(i, j)$ denote the genealogical distance between populations i and j. Populations 1.1 and 1.2 have to go back only one step to find their common ancestor (population 1), while populations 1.1 and 2.1 have to go back two steps to find their common ancestor (population 0), as illustrated in Figure 1. Therefore, we have:

$$d(1.1, 1.2) = d(2.1, 2.2) = 1 \tag{3}$$

and:

$$d(1.1, 2.1) = d(1.1, 2.2) = d(1.2, 2.1) = d(1.2, 2.2) = 2 \tag{4}$$

What is the relationship between genealogical distance, cultural change and technological change? In order to explore these issues, it is useful to consider the following benchmark assumptions:

A1. At each time t two populations i and j with $A_{it} = A_{jt}$ face an identical probability π_t of discovering an *original* innovation that would increase productive knowledge by Δ_t.

This assumption means that cultural characteristics q_{it} *per se* do not have a *direct* effect on the rate of technological progress: two populations with different cultural characteristics but identical levels of productive knowledge face identical probabilities of expanding the technological frontier. In other words, cultural characteristics are assumed to be *neutral* with respect to the process of innovation.[12]

By contrast, it is reasonable to assume that the process of *imitating* somebody else's innovation is a function of the "cultural distance" between the innovator and the imitator. That is, we assume:

---

[11] In this analysis we will abstract from differences in size across populations, and assume that all populations have identical size.
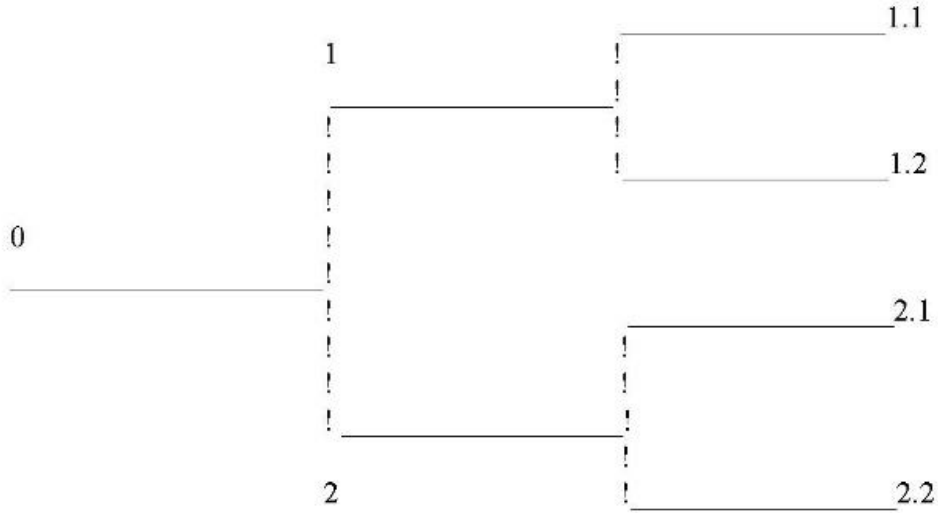
[12] We will relax this assumption below.

Figure 1: Population Tree.

A2. If an innovation is introduced by some population i with cultural characteristics q_i, the extent to which a population j, with cultural characteristics q_j, can increase its own technological knowledge through the imitation and adaptation of population i's innovation will depend on the "cultural distance" between the two populations, that is, on $|q_j - q_i|$.

Specifically, we assume that if a population i with characteristics q_i increases its productive knowledge through an innovation of size Δ_i, a population j with characteristics q_j can increase its own productive knowledge by:

$$\Delta_j = (1 - \beta|q_j - q_i|)\Delta_i \tag{5}$$

where the parameter β measures the barriers to the diffusion of innovations associated with differences in the two populations' characteristics (languages, cultural habits and traditions, etc.). In a general sense, we can interpret these barriers as stemming from difficulty in communication, translation and understanding. Technological and institutional innovations may take specific forms when developed by population i within its cultural setting q_i. Those forms may be difficult to adapt and "translate" into population j's different cultural setting q_j.

But how do different populations end up with differing cultural characteristics? For the purposes of this analysis, we will consider a simple model of cultural divergence ("mimetic drift"):

7

A3. Cultural transmission follows a random walk, in which cultural characteristics are transmitted vertically across populations, while "cultural change" is white noise.

Clearly, this is a highly stylized approximation of more complex phenomena, but it does provide a simple way to capture the dynamics of changes in "neutral" cultural characteristics. Specifically, we will assume that for each population i cultural characteristics are given by:

$$q_i = q_{i'} + \eta_i \tag{6}$$

where $q_{i'}$ are the characteristics of the closest ancestor (population 0 for populations 1 and 2, population 1 for populations 1.1 and 1.2, population 2 for populations 2.1 and 2.2), and η_i is equal to $\eta > 0$ with probability $1/2$ and $-\eta$ with probability $1/2$.

When cultural characteristics follow the above process, we can immediately show that on average cultural distance between two populations is increasing in their genealogical distance. Specifically, in our example, the expected cultural distance between populations at a genealogical distance $d(i, j) = 1$ is:

$$E\{|q_j - q_i| \mid d(i, j) = 1\} = \eta \tag{7}$$

while populations at a genealogical distance $d(i, j) = 2$ have twice the expected cultural distance:

$$E\{|q_j - q_i| \mid d(i, j) = 2\} = 2\eta \tag{8}$$

The above relationships imply that, on average, populations that are closer genealogically will also be closer culturally:

$$E\{|q_j - q_i| \mid d(i, j) = 2\} - E\{|q_j - q_i| \mid d(i, j) = 1\} = \eta > 0 \tag{9}$$

This is not a deterministic relationship: it is possible that two populations who are genealogically far may end up with *more similar cultures* than two populations which are more closely related. But that outcome is less likely to be observed than the opposite. In summary, we have:

Result 1

On average, greater genealogical distance is associated with greater cultural distance.

We are now ready to study the relationship between diffusion of innovations, cultural change, and genealogical distance within our framework.

First of all, consider the case in which inter-population barriers to the horizontal diffusion of innovations are prohibitive. In other terms, consider the case in which there is no horizontal

transmission of innovations, but just vertical transmission. To fix ideas, suppose that at time $t = 1$, each of the two existing populations (1 and 2) could independently increase its inherited productivity A_0 by $\Delta > 0$ with probability π. Assuming that no other innovation takes place at time 2, what are the expected differences in income across populations at time 2?

Populations with the same closest ancestor will inherit the same productive knowledge (either A_0 or $A_0 + \Delta$) and will not differ in income per capita. That is:

$$E\{|y_j - y_i| \mid d(i,j) = 1\} = 0 \tag{10}$$

On the other hand, populations 1 and 2 will transfer different technologies to their descendants if and only if one of the two population has successfully innovated at time 1 while the other population has not. This event takes place with probability $2\pi(1 - \pi)$. Hence, expected income differences across populations with genealogical distance equal to 2 are given by:

$$E\{|y_j - y_i| \mid d(i,j) = 2\} = 2\pi(1 - \pi)\Delta \tag{11}$$

Not surprisingly, when technological innovations diffuses only via vertical transmission, income differences are strongly correlated with genealogical distance:

$$E\{|y_j - y_i| \mid d(i,j) = 2\} - E\{|y_j - y_i| \mid d(i,j) = 1\} = 2\pi(1 - \pi)\Delta > 0 \tag{12}$$

The relationship is stronger the higher is the variance of innovations across population (which, in our example, is measured by $\pi(1 - \pi)$, and it is highest at $\pi = 1/2$.

By contrast, if there were *no barriers* to the horizontal transmission of innovations across populations, all societies would have the same income per capita independently of their genealogical distance.[13] In general, genealogical distance matters for income differences if and only if there are barriers to horizontal diffusion. Let us consider the case in which barriers are positive but not prohibitive, that is:

$$\Delta_j = (1 - \beta|q_j - q_i|)\Delta_i \tag{13}$$

with $\beta|q_j - q_i| < 1$ for all q_i and q_j. Now, populations 1 and 2 will *not* end up with the same technology if and only if a) only one of the two populations finds the innovation (an event with probability $2\pi(1-\pi)$), and b) the two populations are culturally different - that is, one experienced

---

[13] For all populations at time 1 and 2, we would have $y = A_0 + \Delta$ with probability $1 - (1 - \pi)^2$ and $y = A_0$ with probability $(1 - \pi)^2$.

a cultural change equal to η while the other experienced $-\eta$ (an event with probability 1/2). If both a) and b) hold (an event with probability $\pi(1-\pi)$), one of the two populations will have productivity equal to $A_0 + \Delta$ while the other will have productivity equal to $A_0 + (1-2\beta\eta)\Delta$. If no additional diffusion can take place at time 2 (that is, if horizontal transmission is possible only for contemporaneous innovations), we have:

$$E\{|y_j - y_i| \mid d(i,j) = 2\} - E\{|y_j - y_i| \mid d(i,j) = 1\} = 2\pi(1-\pi)\beta\eta\Delta > 0 \qquad (14)$$

The above equation shows that *income differences are increasing in genealogical distance if and only if there are positive barriers to diffusion ($\beta \neq 0$) and populations diverge culturally over time ($\eta \neq 0$).*

In the above example we have assumed that horizontal diffusion of the innovation introduced at time 1 takes place only contemporaneously - that is, at time 1. The analysis can be extended to allow for further horizontal transmission at time 2.

Consider the case in which at time t two populations (say, 1.1 and 1.2) have "inherited" technology $A_0 + \Delta$ by vertical transmission while the other two populations (say, 2.1 and 2.2) have inherited $A_0 + (1-2\beta\eta)\Delta$. From population 2.1's perspective, the unadopted innovation from period 1 is:

$$[A_0 + \Delta] - [A_0 + (1-2\beta\eta)\Delta] = 2\beta\eta\Delta \qquad (15)$$

If we consider this situation as equivalent to the case in which populations 1.1 and 1.2 come up with a *new* innovation of size $2\beta\eta\Delta$, we can model the adoption of that "innovation" by population 2.1 as:

$$\Delta_{2.1} = [1 - \beta|\min\{|q_{2.1} - q_{1.1}|, |q_{2.1} - q_{1.2}|\}]2\eta\beta\Delta \qquad (16)$$

where the expression $\min\{|q_{2.1} - q_{1.1}|, |q_{2.1} - q_{1.2}|\}$ captures the fact that population 2.1 will adopt the innovation from the population that is culturally closer.

In this case, the expected income gap between populations at different genealogical distance is given by:

$$E\{|y_j - y_i| \mid d(i,j) = 2\} - E\{|y_j - y_i| \mid d(i,j) = 1\} = \pi(1-\pi)\beta^2\eta^2\Delta > 0 \qquad (17)$$

which, again, implies a positive correlation between differences in income per capita and genealogical distance, as long as $\beta \neq 0$ and $\eta \neq 0$.

An analogous equation can be obtained for innovations that occur in period 2. If the four populations inherit identical technologies from period 2 and each population can find an innovation of size Δ in period 2 with probability π we have:

$$E\{|y_j - y_i||d(i,j) = 2\} - E\{|y_j - y_i||d(i,j) = 1\} = 2\pi^2(1-\pi)^2\beta\eta\Delta > 0 \qquad (18)$$

We can summarize the analysis above as:

Result 2

Income differences across populations are increasing in genealogical distance if and only if there are positive barriers to the diffusion of innovations ($\beta \neq 0$) and populations diverge culturally over time ($\eta \neq 0$).

2.2   An Extension: Non-Neutral Vertically-Transmitted Characteristics

The above results have been obtained under the assumption that cultural characteristics are neutral - that is, they have no direct effect on the production function and on the process of innovation itself, but only on the process of horizontal diffusion of innovations. The assumption can be relaxed by allowing a direct effect of cultural characteristics on the probability of innovating. Specifically, assume that population i's probability of finding an innovation is given by:

$$\pi_i = \pi + \phi q_i$$

This means that a higher q_i is associated with more innovations and a lower q_i with less innovations. The analysis above can be viewed as the special case $\phi = 0$. Under this more general assumption equation (14) becomes:

$$E\{|y_j - y_i||d(i,j) = 2\} - E\{|y_j - y_i| \mid d(i,j) = 1\} = 2[\pi(1-\pi) + \phi^2\eta^2]\beta\eta\Delta \qquad (19)$$

This equation shows that the larger the direct impact of cultural characteristics on the probability of innovating, the stronger the relationship between expected income differences and genealogical distance, *provided there are barriers to diffusion ($\beta \neq 0$) and cultural heterogeneity ($\eta \neq 0$).* In other words, a direct effect of cultural characteristics on the innovation process strengthens the relationship between genealogical distance and income gaps, as long as there are barriers to the diffusion of innovations, consistent with Result 2 above.

2.3 Vertical Transmission of Characteristics and Income Differences: A General Taxonomy

In the analytical framework presented above, we have illustrated a mechanism of development diffusion implying a positive correlation between genealogical distance (i.e., distance from last common ancestors) and income differences. The central feature of the framework is the link between genealogical distance and the vertical transmission of characteristics across generations. In our basic model, we show how differences in *neutral* characteristics (that is, characteristics that *do not have a direct effect* on productivity and innovations) can explain income differences by playing a role as *barriers* to the diffusion of innovation across populations. We then extended the model to include possible *direct* effects of vertically transmitted characteristics on productivity. Specifically, in our framework we considered a direct effect of different characteristics on the probability of adopting productivity-enhancing innovations. We have seen how direct effects increase the magnitude of the correlation between genealogical distance and income differences, but are not necessary for the existence of a positive correlation: *barrier effects* due to *neutral* vertically-transmitted characteristics are sufficient to explain a positive correlation between genealogical distance and income differences.

In our basic framework we have modeled the transmission of characteristics as "cultural" - that is, *not* directly related to the transmission of DNA from parents to children. We have done that for two reasons. One reason is conceptual: to provide a model that clearly shows how a direct link from DNA-transmitted characteristics to economic outcomes is *not necessary* for our results,as long as vertical transmission of cultural characteristics takes place among genetically-related individuals (typically, parents and children). The second reason is substantial. Our focus is on income differences across different populations of Homo Sapiens Sapiens, taking place over a relatively short period in terms of genetic evolution, and we expect that over that time frame divergence in cultural characteristics have played a more important role than direct changes in non-neutral genetic materials across populations. Hence, we wanted to show a model that is consistent with our priors that "genealogical distance" between populations - e.g. how far populations are in terms of common ancestors - can help explain income differences because it proxies for divergence in (vertically-transmitted) *cultural* characteristics.

However, in principle, the insights from our framework can be generalized to include a broader set of channels through which characteristics are vertically transmitted across generations. In general, characteristics can be transmitted across generations through DNA (genetic transmission, or

12

GT - e.g. eye color) or through pure cultural interactions (cultural transmission, or CT - e.g., a specific language). Moreover, from a hypothetical perspective, vertically transmitted characteristics, whether transferred through GT or CT, may affect income differences because of a direct (D) effects on productivity or because they constitute barriers (B) to the transmission of innovations across populations. Hence, in general one can identify four possible combinations of mechanisms through which vertically transmitted characteristics may affect income differences: a GT direct effect, a GT barrier effect, a CT direct effect, and a CT barrier effect.[14] Figure 2 summarizes the four possibilities.

	Direct Effect (D)	Barrier Effect (B)
Genetic Transmission (GT)	Quadrant I	Quadrant II
Cultural Transmission (CT)	Quadrant III	Quadrant IV

Figure 2 - Taxonomy for the vertical transmission of characteristics

For instance, vertically-transmitted characteristics affecting the trade-off between quality and quantity of children in the theoretical framework proposed by Galor and Moav (2002) would be examples of GT direct effects (Quadrant I). GT barrier effects (Quadrant II) could stem from visible genetically-transmitted characteristics (say, physical looks) that do not affect productivity directly, but introduce barriers to the diffusion of innovations and technology by reducing exchanges and learning across populations that perceive each other as different.[15] Direct effect of cultural characteristics have been emphasized in a vast sociological literature that goes back at least to Max Weber.[16] A recent empirical study of the relationship between cultural values and economic outcomes that is consistent with the mechanisms of Quadrant III is provided by Tabellini (2004). The link between cultural characteristics and barriers (Quadrant IV) is at the core of our basic model, while our extension to non-neutral cultural characteristics may be interpreted as an example from Quadrant III.

---

[14] It is important to notice that these conceptual types should not be viewed as completely separable, but rather as points on a logical continuum, which may involve a mix of them. For example, the ability to digest milk as an adult (lactose tolerance) is genetically transmitted, and may interact with culturally-transmitted characteristics in affecting the impact of innovations in animal domestication on a population's standards of living and productivity.

[15] This effect is related to recent work by Guiso et al. (2004), who argue that differences in physical characteristics may affect the extent of trust across populations. Visible differences across ethnic groups play an important role in the analysis of ethnic conflict by Caselli and Coleman (2002).

[16] More recent references can be found in the edited volume by Harrison and Huntington (2000).

As we will see in the following section, genetic distance can be used to measure genealogical distance (e.g., distance from the last common ancestors), and hence to estimate the relationships between vertically-transmitted characteristics and the determinants of differences in income per capita. Measures of genetic distance are constructed using neutral genetic markers that do *not* affect phenotypic characteristics under strong natural selection. Hence, our empirical measures of genetic distance provide particularly good proxies for genealogical distance related to cultural distance (Quadrants III and IV).

3 Data and Empirical Methodology

3.1 Data

Genetic Distance   Since the data on genetic distance that we use as a measure of distance in vertically-transmitted characteristics is not commonly used in the economics literature, it is worth spending some time describing it. Genetic distance measures the genetic similarity of two populations. The basic unit of analysis is the *allele*, or the variant taken by a gene. By sampling populations for specific genes, geneticists have compiled data on allele frequencies, i.e. the proportion of the population with a gene of a specific variant.[17] Differences in allele frequencies are the basis for computing summary measures of distance based on aggregated differences in allele frequencies across various genes (or *loci* on a chromosome). Specifically in this paper, following Cavalli-Sforza et al. (1994), we will use measures of F_{ST} distance, also known as coancestor coefficients (Reynolds et al., 1983). F_{ST} distances, like most measures of genetic diversity, are based on indices of heterozygosity, the probability that two genes at a given locus, selected at random from the relevant populations, will be different (heterozygous). The construction of F_{ST} distances can be illustrated for the simple case of two populations (a and b) of equal size, one locus, and two alleles (1 and 2). Let p_a and q_a be the gene frequency of allele 1 and allele 2, respectively, in population a.[18] The probability that two randomly selected genes at a given locus are *identical* within the population ("homozygosity") is $p_a^2 + q_a^2$, and the probability that they are different ("heterozygosity") is:

$$h_a = 1 - p_a^2 + q_a^2 = 2p_a q_a \tag{20}$$

---

[17] Allele frequencies for various genes and for most populations in the world can be conveniently searched online at http://alfred.med.yale.edu/

[18] Therefore we have $p_a + q_a = 1$ and $(p_a + q_a)^2 = p_a^2 + q_a^2 + 2p_a q_a = 1$.

By the same token, heterozygosity in population b is:

$$h_b = 1 - p_b^2 + q_b^2 = 2p_bq_b \tag{21}$$

where p_b and q_b be the gene frequency of allele 1 and allele 2, respectively, in population b. The average gene frequencies of allele 1 and 2 in the two populations are, respectively:

$$\overline{p} = \frac{p_a + p_b}{2} \tag{22}$$

and:

$$\overline{q} = \frac{q_a + q_b}{2} \tag{23}$$

Heterozygosity in the *sum* of the two populations is:

$$h = 1 - \overline{p}^2 + \overline{q}^2 = 2\overline{pq} \tag{24}$$

By contrast, *average* heterozygosity is measured by:

$$h_m = \frac{h_a + h_b}{2} \tag{25}$$

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

$$F_{ST} = 1 - \frac{h_m}{h} = 1 - \frac{p_aq_a + p_bq_b}{2\overline{pq}} \tag{26}$$

If the two populations have identical allele frequencies ($p_a = p_b$), F_{ST} is zero. On the other hand, if the two populations are completely different at the given locus ($p_a = 1$ and $p_b = 0$, or $p_a = 0$ and $p_b = 1$) , F_{ST} takes value 1. In general, the highest the variation in the allele frequencies across the two populations, the higher is their F_{ST} distance. The formula can be extended to account for L alleles, S populations, different population sizes, and to adjust for sampling bias. The details of these generalizations are provided in Cavalli-Sforza et al. (1994, pp. 26-27).[19]

These measures of genetic distance have been devised mainly to reconstruct "phylogenies" (or "family trees") of populations. F_{ST} (which is also known as the coancestor coefficient) can be interpreted as the distance to a common ancestor of two populations. Thus, in effect, genetic distance is related to how long two populations have been isolated from each other.[20] If two

---

[19] For a general discussion of measures of genetic distances, see also Nei (1987).

[20] "Isolation" here refers to the bulk of the genetic heritage of a given population. Small amounts of interbreeding between members of different populations do not change the big picture.

populations split apart as the result of outmigration, their genes start to change as a result of genetic drift (randomness) and natural selection. When calculating genetic distances in order to study population history and phylogenesis, geneticists concentrate on "neutral" characteristics that are not affected by strong directional selection occurring only in some populations and environments (Cavalli-Sforza et al., 1994, p. 36).[21] In other words, the term "neutral markers" refers to genes affected only by drift, not natural selection.

If the populations are separated, this process of change will take them in different directions, raising the genetic distance between them. The longer the period for which the separation lasts, the greater will genetic distance become. More specifically, the rate of evolution is the amount of evolutionary change, measured as genetic distance between an ancestor and a descendant, divided by the time in which it occurred. If drift rates are constant, genetic distance can be used as a "molecular clock" - that is, the time elapsed since the separation of two populations can be measured by the genetic distance between them. Figure 3, from Cavalli-Sforza et al. (1994), illustrates the process through which different human populations have split apart over time. Heuristically, genetic distance between two populations is captured by the horizontal distance separating them from the next *common* node in the tree.

Therefore, in this paper we will use F_{ST} distance as a measure of "genealogical distance" between populations. Consistent with our theoretical framework, we expect that a larger F_{ST} distance should reflect a longer separation between populations, and hence, on average, a larger difference in vertically-transmitted characteristics. The data itself is from Cavalli-Sforza et al. (1994), p. 75-76: we focus on the set of 42 world populations for which they report all bilateral distances, based on 120 alleles.[22] These populations are aggregated from subpopulations characterized by a high level of genetic similarity. However, measures of bilateral distance among these subpopulations are

---

[21] The classic reference for the "neutral theory" of molecular evolution is Kimura (1968). For more details on the neutral theory, the "molecular clock" hypothesis, and the construction and interpretation of measures of genetic distance, a recent reference is Jobling et al. (2004). The fact that genetic distance is calculated with respect to "neutral" genetic markets implies that genetic distance can provide an especially useful measure for the channels of Quadrants III and IV in Figure 2 (Section 2.3).

[22] Cavalli-Sforza et al. (1994) also provide a different measure of genetic distance (Nei's distance). Nei's distance, like F_{ST}, measures differences in allele frequencies across a set of specific genes between two populations. F_{ST} and Nei's distance have slightly different theoretical properties, but the differences are unimportant in practice as they are very highly correlated, and the choice of measures does not impact our results (as we show below).

.000 .006 .012 .018 .024 .030

Mbuti Pygmy
West African
Bantu
Nilotic
Bushmen
Ethiopian

African

North African
S.W. Asian
Iranian
European
Sardinian
Indian
S.E. Indian
Lapp

Caucasoid

Samoyed
Mongol
Tibetan
Korean
Japanese
Ainu

N.E. Asian

Siberians
Eskimo
Chukchi

Arctic

S. Amerind
C. Amerind
N. Amerind
N.W. Amerind

American

North
Eurasian

Asian

H. sapiens

S. Chinese
Mon Khmer
Thai
Indonesian
Malaysian
Philippine

Mainland & Insular
S.E. Asia

Polynesian
Micronesian
Melanesian

Pacific
Islands
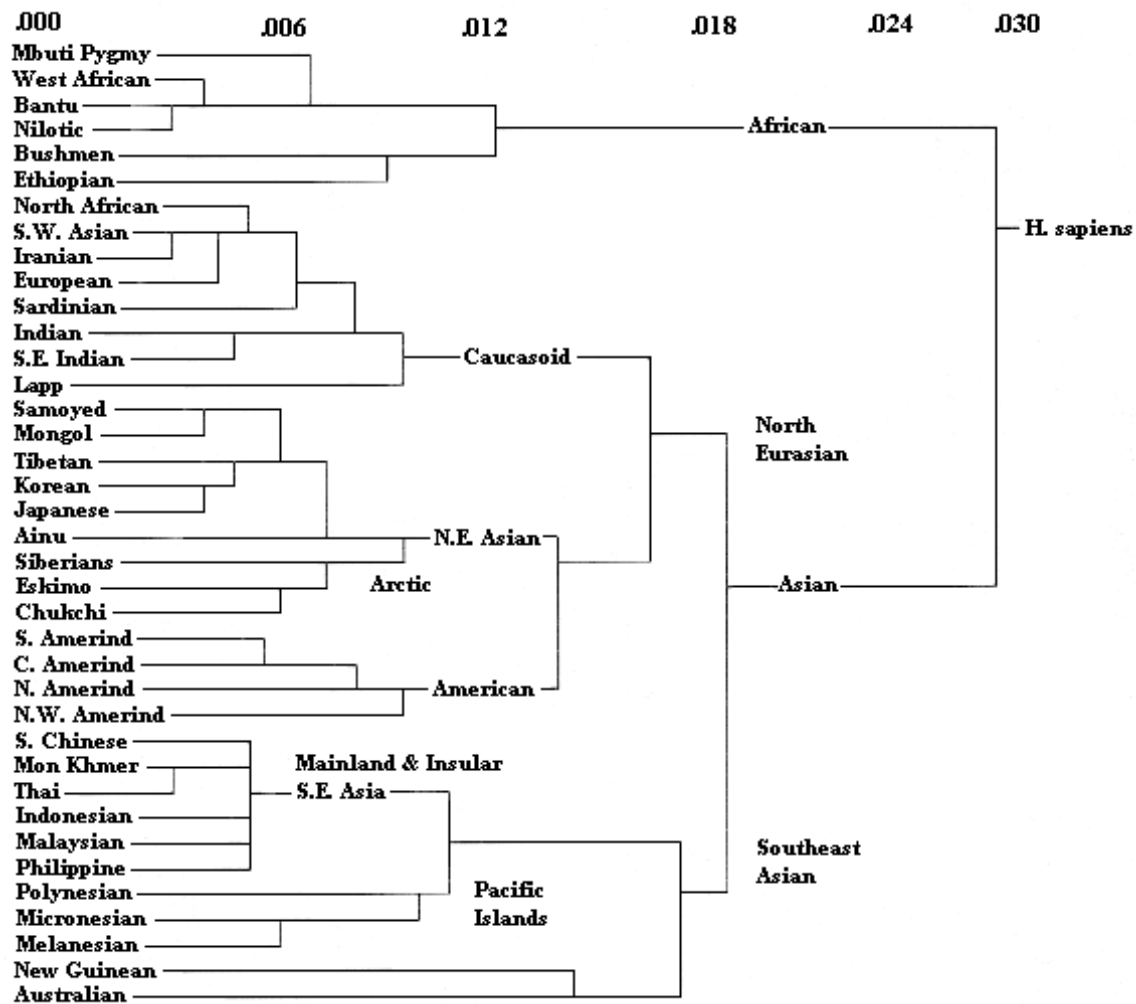
New Guinean
Australian

Southeast
Asian

Figure 2: Genetic distance among 38 populations. Source: Cavalli-Sforza et al., 1994.

available only regionally, not for the world as a whole.[23] Among the set of 42 world populations, the greatest genetic distance observed is between Mbuti Pygmies and Papua New-Guineans, where the F_{ST} distance is 0.4573, and the smallest is between the Danish and the English, for which the genetic distance is 0.0021.[24] The mean genetic distance among the 861 available pairs is 0.1338.

Genetic distance data is available at the population level, not at the country level. It was thus necessary to match populations to countries. We did so using ethnic composition data from Alesina et al (2003). In many cases, it was possible to match ethnic group labels with population labels from Cavalli-Sforza et al. (1994). The was supplemented with information from Encyclopedia Britannica when the mapping of populations to countries was not achievable from ethnic group data. Obviously, many countries feature several ethnic groups. We matched populations to the dominant ethnic group, i.e. the one with the largest share of the country's population.[25]

The ethnic composition in Alesina et al. (2003) refers to the 1990s. This is potentially endogenous with respect to current income differences if the latter are persistent and if areas with high income potential tended to attract immigration since 1500. This would be the case for example under the view that the Europeans settled in North America because of a favorable geographical environment.[26] In order to construct genetic distance between countries as of 1500 (in an effort to obtain a variable that is more exogenous than current genetic distance), we also mapped populations to countries using their ethnic composition as of 1500, i.e. prior to the major colonizations of modern times. Thus, for instance, while the United States is classified as predominantly populated with "English" people for the current genetic distance, it is classified as being populated with "North Amerindians" for the 1500 genetic distance. This distinction affected mostly countries that were colonized by Europeans since 1500 to the point where the main ethnic group is now of Eu-

---

[23] In our empirical work, we make use of the more detailed data for Europe in order to extend our results.

[24] Among the more disaggregated data for Europe which we also compiled, the smallest genetic distance (equal to 0.0009) is between the Dutch and the Danish, and the largest (equal to 0.0667) is between the Lapp and the Sardinians. The mean genetic distance across European populations is 0.013. As can be seen, genetic distances are roughly ten times smaller on average across populations of Europe than in the World dataset. However, we still find that they significantly predict intra-Europe income differences.

[25] We have also computed measures of weighted genetic distance by using the data on each ethnic group within a country. However, we have not used these series yet in our empirical work. Weighted genetic distances are very highly correlated with unweighted ones, so for practical purposes it should not make a big difference which one we use.

[26] In fact, income differences are not very persistent at a long time horizon such as this - see Acemoglu et al. (2002).

ropean descent (New Zealand, Australia, North America and some countries in Latin America...). Genetic distance in 1500 can be used as a convenient instrument for current genetic distance, or can be entered in our regressions in its stead.

Geographic Distance. In addition to genetic distance, we also used several measures of geographic distance. The first is a measure of the greater circle (geodesic) distance between the major cities of the countries in our sample. This comes from a new dataset compiled by researchers at CEPII.[27] This dataset features various measures of distance (between major cities, between capitals, weighted using several distances between several major cities, etc.), all of which bear correlations that exceed 99%. The dataset includes other useful controls such as whether pairs of countries share the same primary or official language, whether the countries are contiguous, whether they had a common colonizer, etc. We used some of these controls in our regressions.

The second measure of geographic distance that we use is latitudinal distance - i.e. simply the absolute value of the difference in latitude between the two countries in each pair: $G_{ij}^L = |\text{latitude}_i - \text{latitude}_j|$. Latitude could be associated with climactic factors that affect income levels directly, as in Sachs (2001). Latitude differences would also act as barriers to technological diffusion: Diamond (1997) suggests that barriers to the transmission of technology are greater along the latitude direction than along the longitude direction, because similar longitudes share the same climate, availability of domesticable animal species, soil conditions, etc. We should therefore expect countries at similar latitudes to also display similar levels of income. All variables are entered in natural logarithms in the regressions.

Summary Statistics Table 1, Panel a displays correlations between our various measures of distance.[28] Perhaps surprisingly, these correlations are not as high as we might have expected. For instance, the correlation between geodesic distance and F_{ST} genetic distance is only 35% - though unsurprisingly it rises to 54.6% if genetic distance is measured based on populations as they were in 1500 (the colonization era acted to weaken the link between genetic distance and geographic distance by shuffling populations across the globe). The correlation between alternative measures of genetic distance, on the other hand, tends to be quite high: the F_{ST} measure and the Nei measure

---

[27] The data is available free of charge at http://www.cepii.fr/anglaisgraph/bdd/distances.htm.

[28] Panel b presents means and standard deviations of the main variables, allowing us to assess the quantitative magnitudes of the effects estimated in the regressions that follow.

of genetic distance bear a correlation of 94%, so it should not matter very much which one we use. Interestingly, latitudinal distance and genetic distance are completely uncorrelated. Finally, our various measures of distance bear positive correlations between 10% and 20% with the absolute value of log income differences in 1995.[29] Together, these correlations suggest it may be possible to identify separately the effects of geographic and cultural barriers on the long-run diffusion of development.

3.2 Empirical Methodology

3.2.1 Specification

Since our empirical methodology is new, we describe it in some detail. We depart from the usual approach, which consists of regressing income levels or income growth on a set of determinants. Instead, we consider a specification in which the absolute *difference* in income between pairs of countries (or other dependent variables such as measures of institutions and human capital) is regressed on measures of distance between the countries in this pair. We computed income differences between all pairs of countries in our sample for which income data was available, i.e. $10,939$ pairs of countries. All our variables are entered in logs. Thus our baseline specification is:

$$|\log y_i - \log y_j| = \beta_0 + \beta_1 \log G_{ij}^D + \beta_2 \log G_{ij}^G + \beta_3 G_{ij}^L + \beta_4' X_{ij} + \varepsilon_{ij} \qquad (27)$$

where G_{ij}^D is genetic distance, G_{ij}^G is geodesic distance, G_{ij}^L is latitudinal distance, X_{ij} is a set of controls capturing other types of barriers and ε_{ij} is a disturbance term to be further discussed below.

The reason our empirical specification must involve income differences rather than a single country's income level on the left hand side is that this makes the use of bilateral measures of barriers possible. There is no other way to quantify the impact of barriers, which are inherently of a bilateral nature, on income differences. Conceptually, therefore, we depart in a major way from existing methodologies: our regression is not "directional" in the sense that the right-hand side variable takes on the same value for each country in the pair, i.e. our specification is not simply obtained by differencing levels regressions across pairs of countries.[30]

---

[29] The data on per capita income is purchasing power-parity adjusted data from the World Bank, for the year 1995. We also used data from the Penn World Tables (Summers, Heston and Aten, 2002), which made little difference. We focus on the World Bank data for 1995 as this allows us to maximize the number of countries in our sample.

[30] This, obviously, would result in adding no new information relative to the levels regressions themselves. Our

We should stress that equation (27) is a reduced form. That is, differences in income are presumably the result of differences in institutions, technologies, human capital, savings rates, etc., all of which are possibly endogenous with respect to income differences. Whether income differences are caused by these factors is the subject of a vast literature but is not primarily the subject of this paper. This paper is concerned with *barriers* to the diffusion of these more proximate causes of income differences: barriers work to explain differences in income presumably because they affect the adoption of technologies, norms for human and physical capital accumulation, the adoption of institutions conducive to differential economic performance, etc.. In Section 5 we will relate barriers to differences in human capital, institutions, investment rates, population growth rates and openness. Evaluating the role of genetic distance in affecting income differences through these various channels, however, does not form the core of our argument.

3.2.2 Estimation

We estimate equation (27) using a new methodology. In principle, if one is willing to assume that the measures of barriers are exogenous, equation (27) can be estimated using least squares. However, in this case usual methods of inference will be problematic. Consider three countries, 1, 2 and 3. Observations on the dependent variable $|\log y_1 - \log y_2|$ and $|\log y_1 - \log y_3|$ will be correlated by virtue of the presence of country 1 in both observations. Conditioning on the right-hand side variables (which are bilateral in nature) should reduce cross-sectional dependence in the errors ε_{12} and ε_{13}, but we are unwilling to assume that observations on the dependent variable are independent conditional on the regressors.[31] In other words, simple least squares standard errors will lead to misleading inferences due to spatial correlation.

Before proceeding, we note the following observations and conventions: with N countries, there are $N(N-1)/2$ distinct pairs. Denote the observation on absolute value income differences between country i and country j as dy_{ij}. Pairs are ordered so that country 1 appears in position i and is

---

methodology is more akin to gravity regressions in the empirical trade literature than to levels or growth regressions in the literature on comparative development.

[31] Another feature that reduces the dependence across pairs is the fact that the dependent variable involves the *absolute value* of log income differences. Simple simulations show that under i.i.d. Normal income draws with moments equal to those observed in our sample (Table 1b), the correlation between absolute value differences in income for any two pair containing the same country will be about 0.22. Without taking absolute values, it is straightforward to notice that the correlation would be exactly 0.5.

matched with all countries from $2...N$ appearing in position j. Then country 2 is in position i and is matched with $3...N$ appearing in position j, and so on. The last observation has country $N-1$ in position i and country N in position j. We denote the non-zero off-diagonal elements of the residual covariance matrix by σ_m where m is the country common to each pair.

A simple example when the number of countries is $N = 4$ is illustrative. In this case, under our maintained assumption that the error covariances among pairs containing a common country m are equal to a common value σ_m, the covariance matrix of the vector of residuals ε is of the form:

$$\Omega = cov \begin{pmatrix} \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{34} \end{pmatrix} = \begin{pmatrix} \sigma_\varepsilon^2 & & & & & \\ \sigma_1 & \sigma_\varepsilon^2 & & & & \\ \sigma_1 & \sigma_1 & \sigma_\varepsilon^2 & & & \\ \sigma_2 & \sigma_3 & 0 & \sigma_\varepsilon^2 & & \\ \sigma_2 & 0 & \sigma_4 & \sigma_2 & \sigma_\varepsilon^2 & \\ 0 & \sigma_3 & \sigma_4 & \sigma_3 & \sigma_4 & \sigma_\varepsilon^2 \end{pmatrix}$$

In this context, controlling for a common-country fixed effect should account for the correlated part of the error term. For this we rely on well-known results cited in Case (1991), showing that fixed effects soak up spatial correlation, though in a context quite different from ours: we do not have longitudinal data, and the panel nature of our dataset comes from the fact that each country is paired with all the other countries in the dataset.[32] Following this observation, we model:

$$\varepsilon_{ij} = \sum_{k=1}^{N} \gamma_k \delta_k + \nu_{ij} \tag{28}$$

where $\delta_k = 1$ if $k = i$ or $k = j$, $\delta_k = 0$ otherwise, and ν_{ij} is a well-behaved disturbance term. We treat δ_k as fixed effects, i.e. we introduce in the regression a set of N dummy variables δ_k each taking on a value of one $N-1$ times.[33] Given our estimator, the effect of the right-hand side

---

[32] Note that simply treating ε_i and ε_j as fixed effects, by including corresponding dummy variables in the regression, will not fully address our concern. This is because, with the exception of country 1 and country N, all countries will appear either in position i or in position j in different observations, inducing spatial correlation between these pairs. In the example above, for instance, country 2 appears in position i in observation 1, and in position j in observation 4, inducing spatial dependence between ε_{12} and ε_{23}. Simple country fixed effects would not soak up this dependence.

[33] The inclusion of fixed effects did not greatly alter the signs or magnitudes of the estimates of the slope coefficients on our variables of interest, compared to simple OLS estimation. In contrast, in line with our expectations, our common-country fixed effects technique resulted in standard errors that were quite different from (and generally larger than) the (wrong) ones obtained with simple OLS.

variables is identified off the variation within each country, across the countries with which it is matched.

To summarize, for each country we create a dummy variable equal to 1 if the country appears in a given pair. We then include the full set of N dummy variables in the regression. The inclusion of these fixed effects soaks up the spatial correlation in the error term resulting from the presence of each country multiple times in various country pairs. In addition, our standard errors are heteroskedasticity-consistent (i.e. we correct standard errors to account for the fact that the diagonal elements of Ω might differ).[34]

4 Empirical Results

4.1 Baseline Results

The baseline estimates of equation (27) are presented in Table 2. Columns (1) through (4) feature our measures of distance entered one by one. The results are in line with expectations: greater distance, whether genealogical, latitudinal or geodesic, is significantly associated with greater income differences. The unconditional results using F_{ST} genetic distance (column 3) suggest that a one standard deviation increase in genetic distance is associated with a 0.256 increase in income differences - 20.19% of this variable's mean. Columns (4) and (5) shows that it matters little whether we use Nei genetic distance rather than F_{ST} genetic distance. In fact, the impact of a standard deviation difference in Nei genetic distance is almost the same as that of F_{ST} distance. We will focus on F_{ST} genetic distance for the remainder of this paper, since it has a clear interpretation as genealogical distance (it is referred to in the literatures as the "coancestor coefficient").

Column (5) and (6) enters all three measures of distance together, for the Nei and F_{ST} distance measures respectively. Interesting results emerge. First, the magnitude of the coefficient on geodesic

---

[34]There are obviously several ways to address the issue of spatial correlation in our context. An alternative we considered would be to do feasible GLS by explicitly estimating the elements of Ω, and introducing the estimated Ω as a weighing matrix in the second stage of the GLS procedure. While apparently straightforward, this is computationally demanding as the dimensionality of Ω is large - in our application we over $10,000$ country pairs with available data on the variables of interest, and up to 155 covariance terms to estimate - so we leave this for future research. We also pursued several bootstrapping strategies based on selecting subsamples such that the problem of spatial correlation would not occur, generating results very similar to those we present here. Details and results from these bootstraps are available upon request. In contrast to these alternatives, our approach is computationally easy to implement.

distance falls by one third, suggesting that it was capturing at least in part the effect of genetic distance. Second, the coefficient on latitudinal distance becomes very small in magnitude and only borderline significant. Hence, genetic distance swamps any effect of differences in latitude on the diffusion of development. Once controlling for genetic distance, the effect of a one standard deviation change in latitudinal distance on income differences equals only 1.59% of the latter variable's mean: the vast income differences observed unconditionally across latitudes (column 2) are not primarily geographical, but linked to vertically transmitted characteristics captured by genetic distance.

Column (7) introduces a set of control variables that might proxy for different sets of barriers, linguistic, historical and geographic (the X_{ij} variables in equation (27)). Some of these variables, such as the one reflecting linguistic similarity and colonial history, are less exogenous with respect to income differences than the distance measures already considered, so results should be interpreted cautiously. Several lessons emerge. First, the signs of the coefficients are largely as expected. Contiguous countries tend to have more similar income levels, as do countries that were ever a single political entity. Countries that have had a common colonizer share more similar income levels, though the effect is small in magnitude. Unsurprisingly, income differences are greater when countries were ever in a colonizer-colonized relationship. On the other hand, having substantial fractions of the populations in each country speaking the same language bears a very small effect on income proximity, and the sign is the opposite of that expected. Second, these coefficients are estimated quite precisely. Third, and perhaps most importantly, the inclusion of these controls does not modify the estimates on our main variables of interest: the distance measures, in particular genetic distance, hardly change at all.

Finally, in column (8), we assess whether the effects of genetic distance might depend on geodesic distance, by adding the interaction between the two variables. We find strong evidence of such an interaction effect: genetic distance matters less for income differences when countries are far apart geographically, and the effect is statistically significant. We can interpret this result as evidence in favor of the "barrier" interpretation for genetic distance: the effect of genetic distance falls as physical distance increases, so these two types of barriers act as substitutes for each other. We would not observe this pattern of coefficients if genetic distance mattered only as the result of a genetic trait bearing a direct effect on income levels.

Together, these results provide considerable evidence consistent with the model of cultural barriers introduced in Section 2: genetic distance enters with a positive and statistically significant

sign in all specifications. Moreover, we find only limited empirical support for the idea that income differences are less pronounced along similar latitudes than across latitudes, once we control for genetic distance.

4.2 Extensions and Robustness

Historical Income Data. Table 3 considers various extensions and robustness checks on our baseline reduced form regressions. The first set of tests we perform examine whether the pattern uncovered for income differences in 1995 held for earlier periods in history. Our basic argument about the importance of cultural barriers for the diffusion of development should hold for other time periods as well, particularly for the period preceding the Industrial Revolution. Unfortunately, internationally comparable income data for this period are available for only a few countries. We used income per capita data in 1500 and 1700 from Maddison (2003), for up to 29 countries (27 for 1500), and repeated our basic reduced form regression.[35] On the right-hand side, we use genetic distance for populations as they were in 1492, prior to the discovery of the Americas and the great migrations of modern times.[36] The results are presented in columns (1) and (2). Much to our surprise, genetic distance again came out statistically significant and positive, despite the small sample. The magnitude of the coefficient is smaller than in Table 2, but so is the average difference in log incomes to be explained (as the dispersion of income was much lower in those time periods): a one standard deviation change in genetic distance is associated with a 0.122 change in the pairwise difference in log income in 1500, corresponding to 38.15% of that variable's mean - a much larger magnitude than in the 1995 data. Similar results hold for income differences computed for 1700 - a one standard deviation change in F_{ST} genetic distance is associated with a change in the difference in log income of 36.20% of its mean. Given the paucity of historical income data, these pre-Industrial Revolution results are quite remarkable.

Finally, we also replicated our results using income in 1960 (column 3). Here, the coefficient is

---

[35] These countries are Australia, Austria, Belgium, Brazil, Canada, China, Denmark, Egypt, Finland, France, Germany, Greece, India, Indonesia, Ireland, Italy, Japan, Korea, Mexico, Morocco, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States. There were 275 pairs with available data for 1500 income, and 328 pairs for 1700. A noteworthy feature of this sample is that it contains no countries in Sub-Saharan Africa.

[36] This variable is constructed using the Cavalli-Sforza data, but matching countries to populations as they were in 1500, not as they are today, as explained in Section 3.2.

smaller than the one obtained for 1995, though it remains highly significant statistically. Again, the average pairwise difference in log income per capita was lower in 1960 than it was in 1990: so a one standard deviation change in log F_{ST} genetic distance is associated with a 0.178 change in the difference in log income in 1960, amounting to 15.95% of that variable's mean.

Possible Endogeneity of Genetic Distance. Next, we attempt to control for the possible endogeneity of genetic distance with respect to income differences. While differences in allele frequencies between the populations of two countries obviously do not depend directly on income, migration could lead to a pattern of genetic distances today that is closely linked to current income differences. Consider for instance the pattern of colonization of the New World starting after 1500. Europeans tended to settle in larger numbers in the temperate climates of North America and Oceania. If geographic factors were to bear a direct effect on income levels, and geographic factors were not properly accounted for in the regressions through included control variables, then genetic distance today could be positively related to income distance not because cultural distance precluded the diffusion of development, but because similar populations settled in regions prone to generating similar incomes.

To assess this possibility, column (4) of Table 3 excludes from the sample any pairs involving one or more countries from the New World (defined as countries in North America, Latin America, the Caribbean and Oceania), where the problem identified above is likely to be most acute. The effect of genetic distance is now actually marginally *larger* than in column (6) of Table 2. Note also that the difference in latitudes becomes three times larger, an observation to which we will return below. Next, we use our data on F_{ST} genetic distance as of 1500 instead of current genetic distance in column (5) of Table 3, and as an instrument for current genetic distance in column (6) of Table 3. This variable reflects genetic distance as it stood before the great migrations of the modern era, and yet is highly correlated with current genetic distance, so it fulfills the conditions of a valid instrument. Again, the magnitude of the genetic distance effect is raised in both cases - in fact it is raised by 20% when 15th century genetic distance is used as an instrument for current genetic distance. This suggests that, if anything, our baseline results were understating the effect of genetic distance on income differences.

The Diamond Gap. Jared Diamond's (1997) influential book stressed that differences in latitude played an important role as barriers to the transfer of technological innovations in early human

history, an effect that could have persisted to this day. Our estimates of the effect of latitudinal distance provided only limited evidence that this effect was still at play: in our regressions we have found little evidence that differences in latitudes help explain much of the income differences across countries, although the effect was much larger when excluding the New World from our sample. However, Diamond took his argument one step further, and argued that Eurasia enjoyed major advantages in the development of agriculture and animal domestication because a) it had the largest number of potentially domesticable plants and animals, and b) had a predominantly East-West axis that allowed an easier and faster diffusion of domesticated species. By contrast, differences in latitudes in the Americas and Africa created major environmental barriers to the diffusion of species and innovations. More generally, Eurasia might have enjoyed additional benefits in the production and transfer of technological and institutional innovations because of its large size (as stressed by Michael Kremer, 1993).

To test and control for a Eurasian effect, we constructed a dummy variable that takes on a value of 1 if one and only one of the countries in each pair is in Eurasia, and 0 otherwise (the "Diamond gap").[37] In order to test Diamond's hypothesis, we added the Diamond gap to regressions explaining income differences in 1995 (column 7) and in 1500 (column 8). For the former regression, we restrict our sample to the Old World. It is appropriate to exclude the New World from the sample when using 1995 incomes because Diamond's theory is about the geographic advantages that allowed Eurasians to settle and dominate the New World. If we were to include the New World in a regression explaining income differences today, we would include the higher income per capita of non-aboriginal populations who are there because of guns, germs and steel, i.e. thanks to their ancestors' Eurasian advantage. As expected, in the regression for 1995 income differences, the Diamond gap enters with a positive and significant coefficient, and its inclusion reduces (but does come close to eliminating) the effect of genetic distance. In column (8), for 1500 income differences, the Diamond gap is again significant and large in magnitude, despite the paucity of observations. This provides suggestive quantitative evidence in favor of Diamond's observation that the diffusion of development was faster in Eurasia. We also conclude that genetic distance between populations plays an important role in explaining income differences even when controlling for the environmental advantages and disadvantages associated with Eurasia.

---

[37] For further tests providing statistical support for Diamond's observations, see Olsson and Hibbs (2005).

Genetic Distance across European Countries. Cavalli-Sforza et al (1994) provide data on genetic distances within certain regions of the world, in particular Europe. These data are more disaggregated (i.e. cover more distinct populations) than the matrix of distances for 42 worldwide populations. For Europe, they present a distance matrix for 26 populations that can be matched quite readily to 24 European countries.[38] Analyzing these data can be informative for several reasons. First, it constitutes a robustness check on the worldwide results. Second, matching populations to countries is much more straightforward for Europe than for the rest of the world, because the choice of sampled European populations happens to match nation state boundaries. Third, genetic distances are generally much smaller across countries of Europe, and genetic specificities within Europe have developed over the last couple of thousand years (and not tens of thousands of years). It is very unlikely that any genetic traits have risen to prominence as the result of strong natural selection over such a short period of time, so a finding that genetic distance based on neutral markers within Europe is associated with income differences would be evidence in favor of the "barriers" interpretation rather than a direct effect of genetic traits.

Table 4 presents the results. Much to our surprise, genetic distance is again positively and significantly associated with income differences. Moreover, the effects are large in magnitude. The baseline results of column (5), which includes several controls (we now exclude the colonial variables for obvious reasons) suggest that a one standard deviation change in the log of genetic distance is associated with a 0.331 change in log income differences, or 59.62% of that variable's sample mean. Thus, cultural barriers captured by genetic distance seem very strongly associated with income differences. Physical distance measures correspondingly bear small or insignificant coefficients, suggesting that geographic barriers are not a big hindrance to the diffusion of income across countries of Europe.

4.3 Effects on the Proximate Determinants of Income Levels

Our approach to quantifying the barriers to the diffusion of development has been of a resolutely reduced-form nature. That is, we did not specify what economic factors make incomes similar or different, and have instead focused on the effects of geographic and cultural barriers on income differences directly. However, differences in income result from more proximate causes. Several

---

[38] These countries are Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Greece, Hungary, Iceland, Ireland, Italy, Macedonia, Netherlands, Norway, Poland, Portugal, Russian Federation, Slovak Republic, Slovenia, Spain, Sweden, Switzerland and the United Kingdom.

prime candidates have been offered to explain differences in income per capita. These factors are summarized in the model of Section 2 by parameter A, and we now discuss them in greater detail.

In the tradition of the Solow model, steady-state income per capita is positively affected by *rates of factor accumulation* (in physical and human capital), and negatively affected by the depreciation of capital per worker, which is more rapid when *population growth* is faster. The level of total factor productivity, in growth accounting or income accounting exercises, has been found to account for much of the variation in growth and income levels.[39] What causes differences in the levels of total factor productivity, however, is largely unknown. On a general level, TFP is "technology", though the deeper determinants of the adoption of better technologies are generally left unspecified. A recent literature has stressed the importance of *institutions* as a determinant of productivity (the seminal contribution here is Acemoglu, Johnson and Robinson, 2001), and Glaeser et al. (2004) have recently reemphasized the importance of *human capital accumulation*, rather than institutions, as a central determinant of income levels. Finally, a large literature emphasized the role of *market size and openness* as a driver of growth and income levels. We consider these five proximate causes of income levels and examine the effects of our measures of distance on pairwise differences in the rate of physical capital accumulation, the rate of population growth, institutional quality, the stock of human capital and openness. These are meant to reflect the prime candidate explanations for differences in income levels. Our empirical model can be summarized as:

$$
\left.\begin{array}{c}
\text{Genetic distance} \\
\text{Geodesic distance} \\
\text{Latitudinal distance} \\
\text{Contiguity, colonial past...}
\end{array}\right\} \Rightarrow
\left.\begin{array}{c}
\text{Institutional quality differences} \\
\text{Human capital differences} \\
\text{Investment differences} \\
\text{Population growth differences} \\
\text{Openness differences}
\end{array}\right\} \Rightarrow \text{Income Differences}
$$
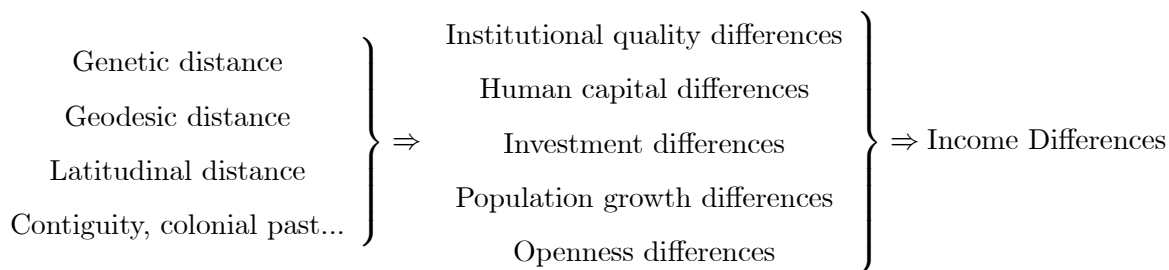
Our paper is primarily concerned with the first arrow in this diagram, and there are voluminous literatures and debates on the respective roles of the proximate determinants of income. Thus, we do not attempt to decompose income differences into differences into the underlying proximate cause, a task that is both a tall order and beyond the scope of this paper. We focus instead on investigating empirically the role of our distance measures as determinants of differences in the proximate determinants of income.

---

[39]For an excellent survey on this point, see Caselli (2005).

Institutions. In an influential series of papers, Acemoglu, Johnson and Robinson (2001) and Rodrik, Subramanian and Trebbi (2004) suggest that the prime cause of economic development is the quality of a country's institutions. If this is the case, countries that are distant in terms of institutional quality should also be distant in terms of income per capita. We use a commonly-used measure of institutional quality, the risk of expropriation variable (this variable, which ranges from 0 to 10, was used for instance in Acemoglu et al., 2001). Constructing the pairwise difference in the risk of expropriation for 1990, we regress this variable on our various measures of geographic and cultural distance. Table 5, column (1) presents the results. The three main measures of distance are significant when entered individually, but the log of geodesic distance becomes small in magnitude when genetic distance and latitudinal distance are entered alongside it.[40] This is interesting, as latitudinal distance had a small effect in the income differences regression. Our finding that latitudinal distance matters for institutional differences is consistent with the view that geographic and climatic factors have historically constituted hindrances to the diffusion of institutions conducive to higher incomes. Our estimates suggest that a one standard deviation change in genetic distance results in a 0.31 change in the difference in expropriation risk, which corresponds to 15.27% of this variable's mean.

Human Capital. In a recent paper taking issue with the literature on the primacy of institutions, Glaeser et al. (2004) suggest that variation in human capital is the fundamental cause of income differences across countries. To examine the role of geographic and cultural barriers in preventing countries from adopting high levels of human capital, we reran the specification of equation (27), replacing the left hand side variable with the absolute difference in the stock of human capital, measured by the average number of years of primary, secondary and tertiary schooling in the population aged 25 and above in 1990.[41] Again, our three categories of distance measures bear positive and significant coefficients when entered individually, and remain so when entered jointly (Table 4, column 2). The pattern of coefficients is similar to that obtained for institutions, with a larger effect of geodesic distance. A one standard deviation change in genetic distance is associated with a 0.55 change in the absolute difference in human capital, or 16.91% of this variable's mean.

---

[40]Results for the specifications where distance measures are entered individually are available upon request.

[41]The human capital data is from Barro and Lee (2000). Again, this is a commonly used, if imperfect, measure of the stock of human capital.

Population Growth.   In column (3) of Table 4, we examine the determinants of differences in rates of population growth. In neoclassical growth models, rapid population growth reduces the steady-state level of income per worker. Thus, differences in population growth (in turn resulting mainly from differences in mortality and fertility) are thought to be associated with differences in income. Again, measures of cultural and geographic distance may affect how differently countries' populations grow. We particularly expect geographic distance measures to be correlated with differences in population growth, as countries located closer to the equator tend to have higher rates of population growth. Indeed, Table 4, column (3) shows that latitudinal distance is positively related to population growth differences (where population growth is defined over the 1960-1990 period): a one standard deviation change in latitudinal distance is associated with a 0.049 change in population growth differences, or 15.41% of this variable's mean. Genetic distance again appears to be significantly related to the dependent variable: a one standard deviation change in genetic distance is associated with a 0.049 change in differences in population growth, or 16.62% of this variable's mean. One interpretation of this finding is that vertically transmitted characteristics are associated with persistent differences in norms of behavior affecting population growth, particularly fertility which dominates the cross-country variation in population growth rates.[42]

Physical Capital Investment.   The rate of investment in physical capital is also a determinant of steady-state income levels in the neoclassical model. How do the geographic and cultural barriers to the diffusion of development relate to differences in this proximate cause of income levels? Table 4, column (4) provides the corresponding estimates, using the Penn World Tables version 6.1 series on the investment share of GDP for 1990. Distance measures bear positive signs when they are entered separately as well as jointly. Genetic distance is positively related to investment differences: a one-standard deviation change in genetic distance is associated with a 1.28 difference in investment rates, which corresponds to 13.79% of this variable's sample mean. We can interpret this finding as indicating that cultural barriers hinder the adoption of norms of investment behavior that are possibly conducive to superior economic outcomes.

Extent of the Market.   Finally, there exists a very large literature on the causal links between international openness, the extent of the market, and economic development. Recent contributions

---

[42] A regression relating differences in fertility rates to our distance measures reveals a very significant and large effect of genetic distance on fertility differences. These results are available upon request.

that have emphasized the extent of the market as an important determinant of economic performance include Ades and Glaeser (1999), Alesina, Spolaore and Wacziarg (2000, 2005) and Alcalá and Ciccone (2004). We examine here whether differences in market size are related to our various measures of distance. Our measure of market size is the conventionally used ratio of imports plus exports to GDP (in 1990), which proxies for access to world markets. Table 4, column (5) displays the results for the openness variable. Differences in openness seem unrelated to most of the variables included in our model, with the exception of the post-war common colonizer variable (this could capture an Africa effect) and the common language variable. Genetic distance appears unrelated to differences in trade openness.

To summarize, genetic distance is significantly positively associated with differences in four of the five proximate determinants of development that we considered, and these effects are generally large in magnitude. This confirms the results found using income differences directly, and provides suggestive evidence that genetic distance acts through several channels. Latitudinal differences were also significantly related to differences in some of these variables (notably institutional quality, human capital and investment), providing some evidence that climactic factors stressed in Sachs (2001) or the geographic barrier effects stressed by Diamond (1997) might be at play for some of the proximate determinants of development - despite being only weakly related to income differences directly.

5 Conclusion

In this paper we have established the following facts: First, differences in income per capita across countries are positively correlated with measures of genetic distance between populations. Second, the effect of genetic distance - which we interpret as an overall measure of differences in vertically transmitted characteristics across generations - hold even when a large set of geographical and other variables are controlled for. Third, the patterns of relationships between income differences and measures of genetic and geographical distances hold not only for current worldwide data but also for estimates of income per capita and genetic distance in the years 1500 and 1700, as well as in a sample of European countries. Finally, similar patterns hold when the dependent variable is differences in human capital, institutional quality, population growth and investment rates.

These results strongly suggest that characteristics transmitted from parents to children over long historical spans play a key role in the process of development. In particular, the results are consistent

with the view that the diffusion of technology, institutions and norms of behavior conducive to higher incomes, is affected by differences in vertically transmitted cultural characteristics associated with genealogical relatedness: populations that are genealogically far are more likely to differ in those characteristics, and hence less likely to adopt each other's innovations over time. Future research should seek to identify the precise mechanisms through which the vertical transmission of characteristics across generations affect the diffusion of development across time and space.

References

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

Acemoglu, Daron, Simon Johnson and James A. Robinson (2002), "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution", *Quarterly Journal of Economics*, volume 117, pp. 1231-1294, November.

Ades, Alberto and Edward Glaeser (1999), "Evidence on Growth, Increasing Returns and the Extent of the Market", *Quarterly Journal of Economics*, vol. 114, no. 3, August, pp. 1025-1045.

Alcalá, Francisco and Antonio Ciccone (2004), "Trade and Productivity", *Quarterly Journal of Economics*, vol. 119, no. 2, May.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat and Romain Wacziarg (2003), "Fractionalization", *Journal of Economic Growth*, vol. 8, no. 2, June 2003, pp. 155-194.

Alesina, Alberto, Enrico Spolaore and Romain Wacziarg (2000), "Economic Integration and Political Disintegration", *American Economic Review*, vol. 90, no. 5, pp. 1276-1296, December.

Alesina, Alberto, Enrico Spolaore and Romain Wacziarg (2004), "Trade, Growth, and the Size of Countries", in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Amsterdam: North Holland.

Barro, Robert J. and Jong-Wha Lee (2000), "International Data on Educational Attainment: Updates and Implications", *Harvard CID Working Paper* no. 42.

Barro, Robert J. and Xavier Sala-i-Martin (1996), *Economic Growth*, Cambridge: MIT Press.

Barro, Robert J. and Xavier Sala-i-Martin (1997), "Technological Diffusion, Convergence and Growth", *Journal of Economic Growth*, vol. 2, no.1, pp. 1-26.

Baumol, William (1986), "Productivity Growth, Convergence and Welfare: What the Long Run Data Show", *American Economic Review*, vol. 76, pp. 1072-1085.

Bisin, Alberto and Thierry Verdier (2000), "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits," *Quarterly Journal of Economics*, vol. 105, no. 3, pp. 955-988.

Bisin, Alberto and Thierry Verdier (2001), "The Economics of Cultural Transmission and the Evolution of Preferences," *Journal of Economic Theory*, vol. 97, no. 1, pp. 298-319.

Boyd, Robert and Peter J. Richerson (1985), *Culture and the Evolutionary Process*, Chicago: University of Chicago Press.

Boyd, Robert and Silk, Joan B. (1997), *How Humans Evolved.* New York: W. W. Norton..

Case, Anne (1991), "Spatial Patterns in Household Demand", *Econometrica*, vol. 59, no. 4, July, pp. 953-965.

Caselli, Francesco (2005), "Accounting for Cross-Country Income Differences", in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Amsterdam: North Holland.

Caselli, Francesco and John Coleman (2002), "On the Theory of Ethnic Conflict", *working paper*, Harvard University.

Cavalli-Sforza, Luigi L., Paolo Menozzi and Alberto Piazza (1994), *The History and Geography of Human Genes*, Princeton: Princeton University Press.

Cavalli-Sforza, Luigi L., and Francesco Cavalli-Sforza (1995), *The Great Human Diasporas*, Reading, MA: Addison Wesley Publishing Company.

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

Diamond, Jared (1997), *Guns, Germs, and Steel: The Fates of Human Societies.* New York: W. W. Norton & Company.

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

Easterly, William and Ross Levine (2003), "Tropics, Germs, and Crops: the Role of Endowments in Economic Development", *Journal of Monetary Economics*, vol. 50, no. 1, January.

Engerman, Stanley L., and Kenneth L. Sokoloff (1997), "Factor Endowments, Institutions, and Differential Paths of Growth among New World Economies: A View from Economic Historians

of the United States", in Stephen Haber, ed., *How Latin America Fell Behind: Essays on the Economic Histories of Brazil and Mexico, 1800-1914*, Stanford: Stanford University Press.

Galor, Oded and Omer Moav (2002), "Natural Selection and the Origin of Economic Growth", *Quarterly Journal of Economics*, vol. 117, pp. 1133-1191, November.

Galor, Oded (2005), "From Stagnation to Growth: Unified Growth Theory", in Philippe Aghion and Steven Durlauf, eds, *Handbook of Economic Growth,*, Amsterdam: North-Holland.

Glaeser, Edward, Rafael LaPorta, Florencio Lopez-de-Silanes and Andrei Shleifer (2004), "Do Institutions Cause Growth?", *Journal of Economic Growth*, vol. 9, pp. 271-303, September.

Guiso, Luigi, Paola Sapienza and Luigi Zingales (2004), "Cultural Biases in Economic Exchange", *NBER Working Paper* No. 11005, December.

Hall, Robert and Jones, Charles I. (1999), "Why Do Some Countries Produce so Much More Output Per Worker than Others?", *Quarterly Journal of Economics*, vol. 114, no. 1, pp. 83-116, February.

Harrison, Lawrence E. and Samuel P. Huntington, eds. (2000), *Culture Matters: How Values Shape Human Progress*, New York: Basic Books.

Heston, Alan, Robert Summers and Bettina Aten (2002), "Penn World Table Version 6.1", *Center for International Comparisons at the University of Pennsylvania* (CICUP), October.

Jobling, Mark A., Matthew E. Hurles and Chris Tyler-Smith (2004), *Human Evolutionary Genetics. Origins, Peoples & Diseases*, New York: Garland Science.

Kimura, Motoo (1968), "Evolutionary Rate at the Molecular Level", *Nature*, vol. 217, pp. 624-626.

Kremer, Michael (1993), "Population Growth and Technological Change: 1,000,000 B.C. to 1990", *Quarterly Journal of Economics*, August, pp. 681-716.

Kremer, Michael, Alexei Onatski and James H. Stock (2001), "Searching for Prosperity," *Carnegie-Rochester Conference Series on Public Policy*, vol. 55, pp. 275-303.

Masters, William A and Margaret S. McMillan (2001), "Climate and Scale in Economic Growth," *Journal of Economic Growth*, vol. 6, no. 3, September.

McNeill, John R. and William H. McNeill (2003), *The Human Web: A Bird's-Eye View of World History.* New York: W. W. Norton & Co, Inc.

Maddison, Angus (2003), *The World Economy: Historical Statistics*, Paris: OECD Development Center.

Mukand, Sharun and Dani Rodrik (2005), "In Search of the Holy Grail: Policy Convergence, Experimentation, and Economic Performance", *American Economic Review*, vol. 95, no. 1, pp. 374-383, March.

Olsson, Ola and Douglas A. Hibbs Jr. (2005), "Biogeography and Long-Run Economic Development", *European Economic Review*, vol. 49, pp. 909-938.

Parente, Steven L. and Edward C. Prescott (2002), *Barriers to Riches*, Cambridge: MIT Press.

Rodrik, Dani, Arvind Subramanian and Francesco Trebbi (2004), "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development", *Journal of Economic Growth*, vol. 9, no.2, June.

Sachs, Jeffrey (2000), "Notes on a New Sociology of Economic Development," in L. E. Harrison and S. P. Huntington, eds, *Culture Matters: How Values Shape Human Progress*, New York: Basic Books.

Sachs, Jeffrey (2001), "Tropical Underdevelopment", *NBER Working Paper* #8119.

Sachs, Jeffrey (2005), *The End of Poverty: Economic Possibilities for Our Time*, New York: Penguin Books.

Spolaore, Enrico and Romain Wacziarg (2005), "Borders and Growth", *Journal of Economic Growth*, vol. 10, no 4, December.

Tabellini, Guido (2005), "Culture and Institutions: Economic Development in the Regions of Europe", *IGIER Working Paper* #292, June.

Weitzman, Martin L. (1992), "On Diversity," *Quarterly Journal of Economics*, vol. 107, no. 2, pp. 363-405, May.

Table 1 – Summary Statistics for the Main Variables

Panel a. Simple Correlations among Distance Measures

	Log of distance	Log Diff Latitudes	Log F_{ST} GD	Log F_{ST} GD, 1500	Log Nei GD	Log Nei GD 1500
Log difference in absolute latitudes	0.426 (13399)	1 (13558)				
Log F_{ST} Genetic Distance	0.349 (12522)	0.100 (12601)	1 (12662)			
Log F_{ST} Genetic Distance, as of 1500	0.546 (12372)	0.185 (12461)	0.749 (12113)	1 (12517)		
Log Nei Genetic Distance	0.355 (12522)	0.147 (12601)	0.943 (12662)	0.713 (12113)	1 (12662)	
Log Nei Genetic Distance, as of 1500	0.543 (12372)	0.227 (12461)	0.711 (12113)	0.948 (12517)	0.753 (12113)	1 (12517)
Abs. log income difference, 1995	0.106 (10939)	0.147 (10875)	0.121 (10102)	0.159 (9981)	0.162 (10102)	0.202 (9981)

(number of observations in parentheses)

Panel b. Means and Standard Deviations

Variable	# Obs.	Mean	Std. Dev.	Min	Max
Log Geodesic distance	13470	8.711	0.797	2.349	9.901
Log Latitudinal distance	13558	2.951	1.060	-3.507	4.663
Log F_{ST} Genetic distance	12662	6.836	0.989	3.045	8.428
Log F_{ST} Genetic distance, 1500	12517	6.979	0.879	3.045	8.428
Log Nei Genetic distance	12662	4.990	0.969	1.099	6.436
Log Nei Genetic distance, 1500	12517	5.072	0.865	1.099	6.479
Abs. log income difference, 1995	10939	1.269	0.902	0.000	4.294

Table 2 - Baseline regressions
(Common-country fixed effects estimates, dependent variable: absolute value of income differences, 1995)

	(1) Distance	(2) Latitude Difference	(3) FST Gen Dist	(4) Nei Gen Dist	(5) All Distances (Nei)	(6) All distances (FST)	(7) Full regression	(8) Full regression
Log FST Genetic Distance			0.259 (0.011)**			0.236 (0.011)**	0.234 (0.011)**	1.048 (0.113)**
Log Nei Genetic Distance				0.260 (0.011)**	0.238 (0.011)**			
Log of Distance (CEPII)	0.271 (0.011)**				0.185 (0.015)**	0.179 (0.014)**	0.168 (0.015)**	0.786 (0.090)**
Log difference in absolute latitudes		0.128 (0.008)**			0.019 (0.009)**	0.024 (0.009)**	0.019 (0.008)**	0.002 (0.009)
1 for contiguity							-0.222 (0.094)**	-0.225 (0.094)**
1 if countries were or are the same country							-0.172 (0.110)	-0.185 (0.106)*
1 if common language (9% threshold)							0.073 (0.027)**	0.072 (0.027)**
1 for pairs ever in colonial relationship							0.147 (0.073)**	0.136 (0.071)*
1 for common colonizer post 1945							-0.096 (0.031)**	-0.105 (0.030)**
1 for pairs currently in colonial relationship							-0.167 (0.112)	-0.209 (0.112)*
Log Distance * Log FST Genetic Distance								-0.095 (0.013)**
Observations	10939	10875	10102	10102	10048	10048	10048	10048
Total # of countries	156	156	155	155	155	155	155	155
Adjusted R-squared	0.32	0.30	0.41	0.41	0.42	0.42	0.42	0.43

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

Table 3 - Extensions and Robustness Checks
(Common-country fixed effects estimates, dependent variable: absolute value of income differences)

	(1) Income 1500	(2) Income 1700	(3) Income 1960	(4) Without New World	(5) 1500 GDs	(6) IV with 1500 GDs	(7) Diamond Gap, w/o New World	(8) Income 1500, Diamond Gap
Log FST Genetic Distance			**0.180** (0.011)**	**0.266** (0.016)**		**0.323** (0.016)**	**0.214** (0.019)**	
Log FST Genetic Distance, as of 1500	**0.139** (0.015)**	**0.184** (0.015)**			**0.266** (0.013)**			**0.138** (0.013)**
Log difference in absolute latitudes	-0.012 (0.017)	0.027 (0.017)	0.049 (0.010)**	0.089 (0.015)**	0.027 (0.009)**	0.031 (0.009)**	0.070 (0.015)**	0.001 (0.016)
Log of Distance (CEPII)	0.009 (0.026)	0.017 (0.025)	0.098 (0.016)**	0.144 (0.026)**	0.093 (0.020)**	0.092 (0.020)**	0.112 (0.026)**	-0.036 (0.023)
Diamond Gap							0.222 (0.035)**	0.203 (0.030)**
Observations	275	354	4955	5977	9931	9582	5977	275
Total # of countries	25	28	107	115	155	155	115	25
Adjusted R-squared	0.50	0.71	0.53	0.36	0.38	0.42	0.36	0.54

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

Table 4 - Regressions for European Countries

(Common-country fixed effects estimates, dependent variable: absolute value of income differences, 1995)

	(1) Distance	(2) Latitude Difference	(3) Genetic Distance	(4) All Distances	(5) Full regression
Log of FST genetic distance in Europe			**0.426** **(0.051)****	**0.460** **(0.063)****	**0.462** **(0.061)****
Log of Distance (CEPII)	0.181 (0.047)**			0.091 (0.062)	0.023 (0.068)
Log difference in absolute latitudes		-0.002 (0.024)		-0.097 (0.027)**	-0.102 (0.027)**
1 for contiguity					-0.071 (0.099)
1 if countries were or are the same country					-0.107 (0.126)
1 if common language (9% threshold)					-0.188 (0.116)
# observations	272	269	268	265	265
# countries	24	24	242	24	24
Adjusted R-squared	0.35	0.31	0.46	0.49	0.49

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

Table 5 – Regressions for the proximate determinants of income differences
(Common-country fixed effects estimates, dependent variables for 1990 specified in column headings)

	(1) Institutions	(2) Human Capital	(3) Population growth	(4) Investment	(5) Openness
Log FST Genetic Distance	**0.318**	**0.624**	**0.050**	**1.293**	**-0.003**
	(0.021)**	**(0.037)****	**(0.003)****	**(0.074)****	**(0.003)**
Log difference in absolute Latitudes	0.216	0.274	0.046	1.066	-0.003
	(0.021)**	(0.035)**	(0.003)**	(0.088)**	(0.004)
Log of Distance (CEPII)	0.082	0.174	0.021	1.006	-0.002
	(0.035)**	(0.059)**	(0.006)**	(0.148)**	(0.006)
1 for contiguity	-0.295	-0.498	-0.004	0.255	-0.091
	(0.234)	(0.264)*	(0.034)	(0.768)	(0.036)**
1 if countries were or are the same country	-0.172	-0.636	0.016	-0.248	0.055
	(0.264)	(0.388)	(0.030)	(0.752)	(0.035)
1 if common language (9% threshold)	0.025	0.199	0.026	0.038	-0.022
	(0.058)	(0.094)**	(0.008)**	(0.232)	(0.011)**
1 for pairs ever in colonial relationship	0.296	-0.065	0.058	1.716	0.037
	(0.117)**	(0.224)	(0.024)**	(0.686)**	(0.023)
1 for common colonizer post 1945	-0.096	-0.423	-0.033	-0.754	-0.117
	(0.079)	(0.113)**	(0.010)**	(0.289)**	(0.017)**
1 for pairs currently in colonial relationship	-0.145	1.842	0.169	-5.091	0.259
	(0.189)	(0.376)**	(0.043)**	(1.009)**	(0.043)**
Observations	4957	4645	5002	6705	6705
Total # of countries	107	103	108	127	127
Adjusted R-squared	0.62	0.45	0.51	0.51	0.83

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