The Roots of Ethnic Diversity

Pelle Ahlerup* Ola Olsson†

May 30, 2007

Abstract

The level of ethnic diversity is believed to have major consequences for economic and political development within countries. In this article, we provide the first systematic empirical analysis of the determinants of ethnic diversity in the world. Our main finding is that the duration of human settlements has a strong positive association with ethnic diversity. Ethnic diversity decreases with the length of state experience and with distance from the equator. Both 'primordial' and 'constructivist' hypotheses of ethnic fractionalization thus receive some support by our analysis.

Keywords: ethnicity, ethnic diversity, human origins
JEL Codes: N40, N50, P33.

1 Introduction

It is widely agreed that ethnic cleavages within countries can have far-reaching consequences for political processes as well as for economic development. Accepting this observation naturally leads to the question: Why are some countries more ethnically fractionalized than others? For instance, why is it the case that the probability that two randomly chosen individuals in South Korea are from different ethnic groups is 0.2 percent, whereas the same probability is roughly 93 percent in Uganda?[1]

The broad aim of this article is to offer an empirical analysis of the determinants of cross-country ethnic diversity. We explore the explanatory power of two main hypotheses regarding the formation of ethnic identities; the *constructivist* view, arguing that ethnic divisions are primarily a product of recent state formation processes during modernity, and the *primordial* view, contending that ethnic divisions have deep roots in history and can be analyzed in an evolutionary framework. In order to be able to test the latter hypothesis, we use recent research on the human genome to develop a new variable capturing the duration of settlement by modern humans for all countries in the world. A

*Dept. of Economics, Göteborg University. Email: pelle.ahlerup@economics.gu.se.
†Dept. of Economics, Göteborg University. Email: ola.olsson@economics.gu.se.
[1]The estimates are taken from Alesina et al. (2003).

prediction from our primordial 'kinship model' of ethnicity is that the antiquity of human settlement should be positively correlated with current levels of ethnic fractionalization.

In our empirical analysis, we demonstrate that our hypothesis regarding the duration of human settlements receives surprisingly strong support, even when controlling for numerous other variables, and that our settlement duration-variable alone explains more than a fourth of the total cross-country variation in ethnic diversity. We also find, in line with the constructivist view, that various indicators of state history tend to increase ethnic homogeneity, and that ethnic diversity decreases with distance from the equator. Our results have particular relevance for the highly fractionalized African countries. Although factors such as the very long settlement of humans, the nearness to the equator, and the lack of historical state experience all serve to explain the current high level of ethnic diversity in Africa, our results also suggest that fractionalization should decrease with time as states mature.

Our work is motivated by a very large literature in social science on the political and economic impacts of ethnic diversity. In economics, an early influential study was Easterly and Levine (1997) who showed that the high degree of ethnic fractionalization in Africa could explain a large part of the continent's dismal growth performance. The proposed link from ethnic diversity to poor economic performance, a lower provision of public goods, was later disputed by Arcand et al. (2000) who also found that the effect of ethnic diversity on growth was even larger in sub-Saharan Africa than in other parts of the world.

There is a widespread notion that the negative association between ethnic fractionalization and growth is not found in all countries. Collier (2000) argues that this association is only found in nondemocratic countries and Easterly (2001) holds that the effect is weaker where institutional quality is higher. Alesina and La Ferrara (2005) basically confirms Easterly and Levine's (1997) results and find that the negative association is less pronounced in rich countries, and that after controlling for this effect the effect of democracy is nonsignificant. Ethnically fractionalized countries tend to be more corrupt and have longer bureaucratic delays, as well as a weaker provision of public goods such as infrastructure, school attainment, and health (La Porta et al., 1999; Alesina et al., 2003).[2] If we allow for a distinction between public goods that are more or less excludable, ethnically diverse societies tend to use relatively more resources on excludable goods and less on non-excludable goods, such as education and

[2]The results in Alesina et al. (2003) are not robust to the inclusion of latitude as a control variables. Since we show that latitude is an underlying factor causing ethnic diversity their approach, to exclude latitude, is not without problems. If latitude is directly associated with, say, productivity omitting latitude from the regressions may produce an entirely spurious association between ethnic diversity and economic growth.

defence (Alesina and Wacziarg, 1998; Kimenyi 2006). However, at the same time as public productivity is lower in ethnically fractionalized countries they have a higher private productivity (Collier 2001). Thus, the negative association between ethnic fractionalization and economic growth feeds from other social failures, be they poverty, poor institutions or lack of democracy.

Dimensions of ethnic diversity have also been discussed in conjunction with civil wars and political instability (Collier and Hoeffler, 2004).[3] The increased scholarly interest in the effects of ethnic diversity stimulated the emergence of two new indices of ethnic fractionalization: Alesina et al. (2003) and Fearon (2003). In our empirical section, we will use the index created by Alesina et al. (2003).

Despite this widespread recognition of the great importance of ethnicity in politics and economic development, we are aware of no systematic attempt that tries to account for the international variation in ethnic diversity. In a rare attempt within economics, Leeson (2005) makes a case that the colonial institutions in Africa accentuated ethnic heterogeneity and therefore contributed to the significant ethnic divisions observed today. Bad institutions thus basically caused ethnic divisions, rather than the other way around. However, Leeson does not have a general theory of why certain countries are more ethnically divided than others. In a theoretical model of ethnic conflict, Caselli and Coleman (2006) propose that people on the losing side of the conflict might switch ethnicity endogenously if the costs of switching are not too high. Their model is in the spirit of the constructivist literature cited below.

In political science and sociology, a rich tradition has studied the sources and impacts of ethnicity on state formations and other historical developments. In the constructivist literature, ethnic identification is basically regarded as a socially constructed phenomenon appearing during modernity for the purpose of uniting nations into states (Gellner, 1983; Anderson, 1983). The primordial tradition, on the other hand - with somewhat less influential proponents like Smith (1986) and van den Berghe (1981) - argues that ethnic identification was a natural and indeed rational behavior that has existed throughout history. Section two below surveys this literature and our empirical analysis is basically structured around the constructivist-primordial hypotheses.

The socio-biological model of van den Berghe (1981) has clear linkages to current advances in genetic research, which is another building bloc of our analysis. The rapid progress in this area has drastically changed our views on how the world was populated in prehistorical times and also sheds light on how closely related different ethnic groups are (Oppenheimer, 2003). This path-breaking re-

[3] Alesina and La Ferrara (2005) provides a comprehensive overview of much of this literature.

search on the human genome has so far had almost no impact in social science.

This brief literature overview suggests that our article makes at least two contributions to the existing literature: Firstly, our article provides the first attempt at measuring the duration of uninterrupted human settlement in a cross-country setting. Secondly, we provide the first systematic empirical assessment of the determinants of ethnic diversity across the world. Our main finding in this regard is that the timing of initial settlement by modern humans still can explain a large fraction of existing differences in ethnic fractionalization. This result is particularly relevant for our understanding of the reasons behind the dismal performance of economic growth in Africa.

The article is structured as follows. In section two, we survey the literature on the constructivist and primordial view on ethnicity. In section three, we outline our own theoretical framework for understanding the reasons behind the differences in ethnic diversity across countries. In section four, we briefly discuss the construction of our important proxy variable for the duration of human settlements. In section five, we present the main empirical results. In section six we allow ourselves to speculate about the implications of our findings and section seven concludes.

2 Theories of ethnic diversity

Before we discuss theories of ethnic diversity, we need to define what we mean by an ethnic group. We define an ethnic group as a social entity with two basic features: 1) The group members have a shared belief of a common history or ancestry, often associated with a 'homeland', a 'founding migration', or a settlement of new territory. 2) The group currently forms a cultural community, manifested for instance in common language, religion, and customs (Fearon, 2003; Bates, 2005). As we shall see, ethnic groups might fulfill these two criteria to a different degree.

Ethnicity should be distinguished from related concepts such as 'race' and 'nation'. Although both terms are generally poorly defined in the literature, race usually refers to physical distinguishing features such as skin color, hair texture, or stature. The concepts of nation and nationalism, on the other hand, are also based on notions of a shared ancestry and a cultural community, but most authors consider nationalism to be a concept that is primarily used in conjunction with discussions about state formation during modernity (Gellner, 1983).

The basic issue that we address in this article is why some countries are more ethnically heterogeneous than others and the remainder of this section discuss this from the two main viewpoints, the *constructivist* view and the *primordial*

view.

2.1 The constructivist view

In essence, the constructivist (or modernist) view maintains that there is nothing profound or deterministic about ethnic divisions. Scholars in this tradition tend to emphasize the cultural community aspect in the definition above and stress that notions of ancestry are usually myths that might not be based on historical facts. One of its most famous proponents, Eric Hobsbawm, even argues that national traditions are sometimes invented to create a stronger sense of national unity (Hobsbawm and Ranger, 1983). Ethnicity is seen as a malleable concept and ethnic self-indentification can change over time depending on the historical circumstances. In economic terms, ethnic identity often serves as a commitment technology for collective action problems in the sense that it provides selective incentives for cooperation. On a macro level, ethnicity and nationalism are regarded as intertwined phenomena that typically play important roles in the formation of nation-states from larger political units.

The constructivist view of ethnicity developed during the 1960s and 1970s, partly as a reaction to the ethnic nationalism that had prevailed previously in both politics and social science. This older view of nations as a universal and natural phenomenon had its intellectual roots in the writings of Rousseau and the nineteenth century German Romanticism. In the seminal works by Anderson (1983) and Gellner (1983), nations (and ethnic identities) are seen as 'imagined communities' that do not become a relevant historical force until the industrial era with print-capitalism, universal literacy, and occupational mobility. Nationalism and ethnic identification are seen as an inevitable response to the gradual decline of dynastic empires and religious communities in the modern era. Usually, this era is thought to have started with the French Revolution in 1789.

In a critique of the Hegelian romanticist view of nationalism, Gellner (1983:48) writes:

> "Nations as a natural, God-given way of classifying men, as an inherent though long-delayed political destiny, are a myth; nationalism, which sometimes takes pre-existing cultures and turns them into nations, sometimes invents them, and often obliterates pre-existing cultures, *that* is reality, for better or worse, and in general an inescapable one. Those who are its historic agents know not what they do..."

Modernist scholars in this tradition thus regards nationalism and ethnic identification as instrumental factors in the shaping of nation-states. In this

process of state consolidation, some existing cultures or ethnicities are inevitably suppressed until they disappear from history whereas the leading group in the process often emerges with a deepened sense of nationalist and ethnic sentiment. Nonetheless, some ethnic communities manage to survive despite a failure to create nation-states like the Kurds in the Middle East and the Basques in Spain and continue to pose challenges to existing nation-states.

However, the social construction of ethnicity was not only a product of nationalist ambitions in the Western world, it was also practiced by the colonial powers in their acquired possessions. A 'divide-and-rule' policy was for instance practiced by the Belgians in Rwanda where the small minority of Tutsis was favored and used to keep the majority of Hutus in check. The English were perhaps even more skilled in the art of creating alliances and playing out groups against each other which also often had the result of strengthening ethnic identification. In a study of a single ethnic group living in both Kenya and Tanzania, Miguel (2004) shows that even among former colonies, governments might have an important role to play in fostering a national identity over tribal identification.

In economics, not much attention has been put to understanding the sources of ethnic diversity. Most economists appear to favor the instrumental view of ethnicity. In a rare attempt at addressing these issues, Caselli and Coleman (2006) present a model of ethnic conflict where two groups compete over a common resource. If the dominant group tries to appropriate the resources of the weaker group, people in the latter group might choose to switch ethnic identity, at least if the 'costs' of doing so are not too high.[4]

In Alesina and Spolaore (1997), it is instead a country's borders that are endogenous. A potential enlargement of a country entails a trade-off between the negative effect of increased preference heterogeneity among the population and the positive effect of decreasing costs per capita of public goods. In optimum, country borders should thus be adjusted to take into account for instance the degree of ethnic fractionalization. All else equal, regions with a larger number of ethnic groups (and therefore a greater degree of preference heterogeneity) should therefore optimally contain relatively many but small nation-states. Such a configuration appear to be relevant for Europe after 1800 but less obviously so for many former colonies where borders in a given territory often do not take into account the ethnic composition (Herbst, 2000; Olsson and Hansson, 2007).

[4] The costs are higher the greater the 'ethnic distance' between groups. Skin color is an ethnic marker that is more or less impossible to change, whereas it is easier to convert to another religion for the sake of joining the victorious ethnic group.

2.2 The primordial view

Although the modernist (or social constructivist) view is currently the dominant paradigm on ethnicity and nationalism, there are a number of influential proponents of a more long-run perspective of ethnic diversity.

At the core of the modern primordial view lies an emphasis of the history and origins of ethnic groups. Its most famous scholar, Anthony D. Smith, criticizes the work of modernists like Gellner (1983) and Anderson (1983) for having a too short-sighted and instrumental view of ethnicity. Rather than being a product of specific European conditions after 1800, Smith (1986) contends that nations and ethnic identification have been in place at least since the emergence of the first civilizations. Already in the late third millennium BC, there was a system of states in the Near East based on ethnic core populations, including the Egyptian and Sumerian civilizations. The written records from this time also mention other ethnic groups and state formations including those of the Amorites and the Canaanites. The growing states of sedentary farmers gradually absorbed smaller ethnic groups into their cultural community, a process that continued over several millennia in China (Diamond, 1997). Smith (1986) therefore argues that the modernist conception of ethnic identity and nationalism as a post-1800 phenomenon, is not correct.

Since ethnic identification appears to have been around throughout recorded time, primordialists consider ethnic identification as a natural organization of human existence, i.e. natural in the positivist sense of being a stylized empirical regularity. But if fully developed nation-states with distinct written languages, religions, customs, and traditions were in place in the Near East as early as 3000 BC, where did they originate from? Smith does not attempt to explain this question.

One potential explanation is provided by the socio-biological theory of ethnic origins, associated mainly with Edward Shils (1957) and Pierre van den Berghe (1981, 1995). Firmly rooted in evolutionary biology, van den Berghe develops a model of ethnicity as 'extended kinship'. The basis for the argument is that humans - like other mammals - are by nature nepotistic, favoring kin in the daily struggle for survival. By the evolutionary logic, given a lifetime budget constraint of time and energy, an individual has a greater chance of passing on her genes to future generations if she invests all her resources into her off-spring and family, rather than if she spends her time and effort on unrelated people. This means that nepotistic genotypes will generally have a greater reproductive success and tend to dominate all populations.

The nepotism argument applies also to members of the extended family since they also carry one's genes, though not to the same extent as direct offspring.

Individuals therefore naturally develop a sense of loyalty with their nearest family, their extended family, their clan, and so on. Since this extended kinship eventually becomes very large and since it is usually hard to distinguish kin just by physical appearance (as neighboring people tend to look alike), particular cultural markers evolve such as dialects, customs, and traditions in order to differentiate from 'the others'. In this way, extended families evolve over generations to become ethnic groups.[5]

3 A joint framework

In this section, we use the literature cited above and extend it to provide a joint theoretical framework for the empirical analysis. A key feature in this framework and in the empirical tests that are presented below, is the primordial view of ethnicity as extended kinship.

3.1 The kinship model of ethnic diversity

A straightforward implication of van den Berghe's theory is that we have every reason to believe that distinguishable extended kinships of the type described above have existed throughout the whole of human history. It is by now generally agreed that the history of modern homo sapiens goes back roughly 200,000 years. The oldest fossil of an anatomically modern human (AMH) is the so-called Omo I skeleton retrieved from a site in Kibish in southwestern Ethiopia. It was recently dated to be approximately 195,000 years old (McDougall et al., 2005). Genetic research on human origins have suggested that all human beings in the world today - and therefore also all ethnic groups - have originated from a founding population of a few thousand individuals who lived in East Africa (Oppenheimer, 2003). By roughly 12,000 years ago, all the major continents had been populated by AMH and the world population size had increased to approximately 4 million (Kremer, 1993). We argue that ethnic diversity had its primary origin in this long prehistorical era.

The model of ethnicity as extended kinship that we propose has essentially the same logic as genetic studies on human genealogy. Over the thousands of years of prehistorical time when AMH roamed the earth as hunter-gatherers, evolution appears to have favored individuals who practiced kinship nepotism (Jones, 2000). Nepotistic individuals that organized in extended family groups had an advantage in having an efficient mechanism for sustaining various forms of collective action. Family ties restricted free riding behavior and provided

[5] Myths of national origins indeed often have this feature. For instance, the Bible describes how the Israeli nation emerged from Jacob's twelve sons who formed the tribes of Israel.

an informal rule-based system in the absence of codified law or a ruling elite. More or less complicated family networks supplied selective disincentives against cheating on delivering collective goods.

As AMH migrated from their East African home to other parts of sub-Saharan Africa, an inevitable process of ethnic (and genetic) fractionalization started. This process would continue throughout prehistory as AMH then settled continent after continent. Adapting to local climates and environments gave rise to new customs as well as biological changes. When all the major continents were settled by 12,000 BP (before present time), the formation of new ethnicities due to colonization ceased.[6]

However, since the level of the population slowly but steadily increased even after 12,000 BP (Cohen, 1977), we argue that the process of ethnic fractionalization should have continued among the hunter-gatherer societies. A growing population within a kinship-based ethnic group will eventually imply that it becomes impossible to keep track of all family relationships. In the absence of written language, sedentism, and supra-tribal authority, we postulate that such groups will split to remain at a balanced level of social control through extended kinship.

Our kinship model of ethnic fractionalization thus implies that there should be strong similarities between the formation of genetic and ethnic groups. Attempts at linking genetic and ethnic diversity into one framework have previously been made outside economics. It has been shown that linguistic groups to some degree follow genetic patterns, suggesting "parallelism between genetic and linguistic evolution" (Cavalli-Sforza et al., 1988:6002). The development of mutually unintelligible languages will further take a mere 1000 to 1500 years if a population with a common language becomes separate into two groups (Cavalli-Sforza and Cavalli-Sforza 1995). Recent research by Dunn et al. (2005) on indigenous peoples in South Asia further suggest that there appears to be close similarities between the genetic relatedness among groups and differences traced by studying differences in language structure.

In Papua New Guinea (PNG) - where isolated primitive peoples have populated the greater part of the country to this day - it is estimated that about 820 different languages are spoken among its 5.6 million inhabitants (CIA, 2007). Fearon's (2003) score for ethnic fractionalization in the country is 1, i.e. there is a 100 percent probability that two randomly drawn individuals in the country belong to different ethnic groups. One might argue that PNG is special because of its extreme geography with mountains and impenetrable rain forests. However, PNG is believed to have been populated for some 65,000 years and is

[6] The peopling of islands - in the Pacific, in the Atlantic, and in the Caribbean - continued for several millennia after 12,000 BP.

therefore one of the countries with the oldest presence of AMH outside Africa. We think it is highly likely that a similar or even higher level of ethnic diversity was in place among hunter-gatherers in tropical Africa before the rise of agriculture on that continent.

More specifically, our kinship model of ethnicity proposes that assuming otherwise identical conditions and significant barriers to migration between continents, ethnic diversity in a specific area is essentially a function of historical time elapsed since the originating settlement of AMH. Figure 1 shows the main intuition in an extremely simplified sense. Let us consider three continents A, B, and C. Colonization can only occur from a more populous continent to a less populous one. The first group of modern humans A_0 emerges in period 0 in continent A. In period 1, this initial ethnic group splits due to population growth and migration into the equal-sized new groups A_1 and A_2- who remain in continent A- and B_0 who migrate to continent B. Ethnic fractionalization then continues at the same rate in period 2 so that six new equal-sized ethnic groups form out of A_1 and A_2 whereas founding 'out-of-A'-group B_0 splits into B_1, B_2, and C_0, the latter being the founding population on continent C.[7]

The key point with this simple diagram is to show that when we assume otherwise identical population growth, geographical conditions, and fractionalization rates, ethnic diversity should be largest on continent A, with a long history of human settlement, and smallest on continent C with a short history. In fact, using the standard Herfindahl index of ethnic diversity, our simple example implies that the A-continent should have an ethnic fractionalization of 5/6 in period 2, the B-continent a score of 1/2, and the C-continent an index number of 0, indicating complete homogeneity. The prediction for our cross-country study is therefore that ethnic diversity should increase with the duration of continuous human settlement in that particular country.[8] We will return to the issue of measuring duration of human settlement in section three below.

3.2 Neolithic agriculture, statehood, and colonialism

The primordial hypothesis, based on the kinship model of ethnicity, might be contrasted with constructivist hypotheses. Although we argue that we should be able to detect a signal from the timing of settlement to current levels of ethnic

[7] The emergence of new ethnic groups is thus similar to a 'mutation'. In fact, studies on the human genome have produced genealogical trees showing how related or ethnically distant populations around the world are. As expected, all studies show that the genetic diversity is greatest within Africa, consistent with the hypothesis of an African human origin (Ingman et a, 2000; Oppenheimer, 2003).

[8] Note also that our example implies that the 'salience' of ethnic divisions - sometimes referred to as ethnic distance - is on average greatest in the A-region. For instance, A_{21} and A_{13} are 'cousins' in the ethnic family tree whereas B_1 and B_2 are 'sibling' groups.

diversity, there are numerous more recent factors that could potentially weaken this relationship. Figure 2 outlines some of these other factors. A turning point in human history was for instance the rise of Neolithic agriculture. The transition was first initiated in the Fertile Crescent in the Near East around 10,500 BP, from which it spread westward to Europe and eastward towards the Indus Valley. Independent transitions also occurred in China (9000 BP), in South America (4300 BP), and in some other places (Smith, 1998; Putterman, 2007a). From having been nomadic hunter-gatherers, people became sedentary farmers relying on domesticated crops and animals. Sedentism and farming revolutionized human lives in several aspects. Two of the most important changes were firstly a large increase in population growth, and secondly, the introduction of a new class of specialists including warriors, craftsmen, priests, and rulers (Diamond, 1997; Olsson and Hibbs, 2005).

On all continents, the rise of sedentary agriculture and a more stratified society was relatively soon followed by the emergence of states (supratribal authority), writing, and monumental collective works such as the pyramids in Egypt, Sumer, and Mexico, i.e. what we usually refer to as 'civilization'. What consequences did this development have for ethnic diversity? Gellner (1983) argues that since the masses of farmers were relatively immobile and since literacy was only reserved for a small elite, the type of cultural homogenization that took place in Europe after 1800 was not possible. On the other hand, Smith (1986) shows that for instance the ancient Sumerians - scattered around cities in the densely populated Iraqi river plains - had a strong sense of a distinct ethnic identity with a common language and religion, as had the Egyptians, and many other peoples during the same time in the Near East. In China, it is well-known that the state gradually incorporated surrounding ethnic groups into their Han culture (Diamond, 1997). Numerous other more recent historical accounts of medieval and modern state formations in for instance France, Germany, and Spain also suggest that statehood experience in general has a homogenizing influence on culture and on ethnic identity.

We believe that the main reason for this tendency is that within states, extended kinships partly lose their role as the most efficient mode of organizing collective action. State institutions like codified law, courts, taxation, and military protection thus substitute the services provided by extended kinships which is the reason why many small ethnic groups disappear in such an environment.[9] In our empirical analysis, we will therefore pose the hypothesis that the length of statehood experience (and the associated time since the transition to agriculture) should have a negative influence on ethnic diversity.

[9] In the short run, however, an increasing state influence can might actually mobilize ethnic sentiment and lead to ethnic confrontations, as in many African countries today.

States have not only created institutions that passively reduced heterogeneity, but have also worked actively to obtain more homogenous populations. This process gained momentum as the modern industrial European states at the turn of the eighteenth century acquired both the means and the motivation for nation-building. The modern industrialized society saw an ever increasing division of labor which, in combination with rapid changes, created a situation best dealt with by creating a dynamic and mobile workforce. The fact that education became generic and standardized meant that an increased economic specialization went hand in hand with a culturally more homogenous population. The industrial society required that one could easily communicate with people one had never met before and this meant not only sufficient linguistic homogeneity but also cultural homogeneity in order to reduce costly misunderstands and confusion to a minimum (Gellner 1983).[10] Tilly (1992) has a slightly different perspective, where the driving force behind the attempts to create a homogenous population was the desire to gather enough resources and enthusiasm from a population to enable the waging and winning of wars. The process started well before the era of industrialization, when direct rule replaced indirect rule by intermediaries, and the process was both influenced by and a cause of the extent of homogeneity in the population. Tilly argues that "rulers frequently sought to homogenize their populations in the course of installing direct rule" and the reason was that "within a homogenous population, ordinary people were more likely to identify with their rulers, communication could run more efficiently, and an administrative innovation that worked in one segment was likely to work elsewhere as well. People who sensed a common origin, furthermore, were more likely to unite against external threats." (1992:106)

The process whereby the European states enforced national rather than local loyalty began in the eighteenth century yet it was first after the middle of the nineteenth century that states forcefully began expand into nonmilitary activities and populations increasingly came to view the state as the natural provider of services previously provided at the local level. Subnational loyalties and identification withered as "states undertook to impose national languages, national education systems, national military service, and much more. [...] National symbols crystallized, national languages standardized, national market organized." (Tilly 1992:115-6). Though genuine measures for the ability of the state to accomplish these goals are hard to come by we expect the available measures to indicate that states that were more powerful during the modern

[10] Culture can be understood as a set of solutions to a large number of coordination problems. Given that different solutions to coordination problems are equivalent from an economic efficiency perspective different societies can obtain different sets of solutions - i.e. different cultures. The power to impose one solution to a coordination problem on a population increased as the centralized industrial state grew in strength.

era have a lower degree of ethnic diversity today.

Another major historical event with substantial repercussions was Western colonialism from the fifteenth century onwards. This extremely heterogeneous process, coherently analyzed in for instance Osterhammel (2005) and Olsson (2007), had very diverse effects depending on the time and duration of colonial dominance, the identity of the colonizer, the geography of the area, and the initial wealth of the population. However, since colonial states were primarily created to suit the needs of the colonial power, there were no strong incentives among the ruling groups of Westerners towards ethnic homogenization. On the contrary, 'divide-and-rule' appeared to be the most often used principle for keeping colonies under control, all since the days of Cortes' exploitation of ethnic conflicts during his conquest of the Aztec empire, to the cynical differential treatment of Hutus and Tutsis by the Belgians in twentieth century Rwanda. One may therefore be lead to believe that ethnic diversity should increase with colonial experience. But, during the colonization of the Americas large segments of the population was killed by the introduction of for them lethal diseases (Diamond 1997) and new population groups was introduced in the form of slaves with African descent. If there is an overall effect of colonization on ethnic diversity remains an empirical issue and, as we show below, we do not find strong evidence in favor of such an effect.

3.3 Geographical factors

Apart from these influences, there are factors at the micro level that could potentially affect the level of ethnic diversity in a country. Generally speaking species richness, or diversity, is a product of isolation and adaption. Beginning with the latter we know from ecology that species richness increases as we get closer to the equator[11] and from paleoanthropology and medical science that variation in human skin color comes partly from differences in UV radiation, which in turn is affected by latitude, altitude and humidity. In fact, natural variations in UV radiation, by latitude and altitude, and precipitation can explain most of skin color variation (Chaplin 2004).[12] The residual variation can to some degree be explained by quite recent migrations, where populations have not had enough time to adjust yet (Diamond 2005). This implies that similarity of skin color is a weak predictor of close genetic connections (Jablonski 2004). Differences in skin color do not create ethnic groups by itself, but classification of people into groups may be easier where there are notable differences in skin color, making the formation and identification of ethnic groups more rapid and

[11] In fact, the highest diversity is generally found around latitude 20-30°.

[12] Skin color affects production of vitamin D, protection againts skin cancer, and can also work as a camouflage! (Diamond 2005)

detailed (Caselli and Coleman, 2006). This implies that diversity within a country can be related to latitude, as well as within-country differences in latitude, humidity, and altitude. Studying pre-colonial North America Mace and Pagel (1995) find that language diversity follows the same latitudinal pattern as have been found for other mammals and birds. They also find that in this pre-colonial environment linguistic diversity was higher in areas with more habitat diversity.

The impact of latitude on ethnic diversity is however complex. Cashdan (2001) has shown that the correlation between latitude and ethnic diversity is largely due to climatic variability, habitat diversity, and pathogen loads. Where climate is variable and unpredictable populations are forced to become generalists and use wider ecological niches, and the presence of high pathogen loads can, when local populations have adapted to them, be an isolating force by working both as barriers to their own movement outside their territory and other populations' movement or conquest into the territory. Collard and Foley (2002) find that the number of 'cultures' within a certain area follows geographical patterns, and falls with latitude, temperature, and rainfall, and that this pattern holds both in 'new continents' such as the Americas, and 'old continents' such as Africa.

Differences between habitats could have worked to create ethnic groups in premodern societies, which suggests that we should investigate the influence of variation in soil and vegetation types. Also, some factors could have a positive association with diversity under some circumstances but a negative association under other circumstances. In harsher environments variation in climate may force people to move over the seasons, e.g. in order to follow their cattle as it graces vast areas, while people living in naturally more affluent areas may become isolated in their smaller, and yet sufficiently productive habitat.

A problem with geographical factors in our empirical study is that they are likely to both affect our dependent variable directly in a 'biological' sense, but also indirectly through their influence on society in general, as indicated in Figure 2. For instance, as emphasized by Diamond (1997) and Olsson and Hibbs (2005), the populations living in areas with a biogeography favorable for agriculture - e.g. riverine habitats with irrigation potential and many suitable plants and animals for domestication - would be the first to make the transition and develop dense sedentary farming populations.[13] A high population density means less isolation, ceteris paribus, and should thus be associated with less diversity. As mentioned above, the transition to agriculture was usually soon followed by the formation of states (Chanda and Putterman, 2006). Hence, a

[13] In the Near East, this biogeographic potential included the unusual abundance of grasses with a heavy kernel (such as the wild variants of wheat and barley) as well as many large and easily domesticated mammals (such as the wild ancestors of sheep, goat, and cattle).

high population density would tend to decrease ethnic diversity both by decreasing isolation and by fostering statehood. Population densities today may have little to do with population densities hundreds or thousands of years ago. We will include some proxies for both historical and current densities in our empirical assessments.

As any species spread out from its origin genetic diversity declines naturally due to mechanisms such as *genetic drift* and the *founder's effect*.[14] These types of effects are very similar to the scenario for ethnic groups outlined in Figure 1. One hypothesis maintains that the first humans initially followed the coastlines as they spread out from Africa, with a beachcombing lifestyle, which means that areas closer to the coast, and maybe waterways connected to it, were settled quite long time before the inland (Oppenheimer, 2003; Macauley et al., 2005). This suggest that coastal areas on the one hand could harbour more diverse populations due to their longer time as settlements, but on the other hand populations in these areas are less isolated. Over time, we believe that the latter effect should dominate so that diversity is lower if larger parts of the country is relatively close to the sea. We will attempt to test this hypothesis in the empirical section.

4 Coding initial human settlement

The main variable in our empirical analysis is our proxy for the historical duration of human settlements, *Origtime*. Since we have coded this variable ourselves and since our major results concern this variable, we will briefly discuss in this section how it has been constructed.

What we have tried to establish is the date of the first uninterrupted settlement by AMH for a sample of 191 countries. Our main sources for this data collection has been Oppenheimer (2003) and Bradshaw Foundation (2007), complemented for islands with Encyclopedia Britannica (2007).[15] Oppenheimer (2003) provides a synthesis of genetic, archeological, and climatological evidence for constructing the likely paths of how AMH settled the world. It should be recognized from the start that the data has numerous sources of potential measurement error. The most definitive evidence of human presence in a country -

[14] Genetic drift: over time a certain feature or genetic marker will either die out or completely dominate a population. Note that this is not the same as natural selection which is a process where over time the beneficial features or genetic markers will come to dominate. In smaller populations genetic drift will dominate natural selection.

Founder's effect: when only a small fraction of the whole population moves on to establish a new population it will have a lower degree of genetic variation. Events such as population bottlenecks, where for some reason the population size is dramatically reduced, can have the same effect.

[15] Bradshaw Foundation (2007) builds to a large extent on Oppenheimer (2003).

fossils of accurately dated human skeletons or artefacts - are only rarely available for individual countries. What researchers need to rely on is instead deductive reasoning on the basis of mainly genetic evidence.

Genetic research on human origins has developed very fast since the initiation of the Human Genome Project at the end of the 1980s. Every cell nucleus of the human body contains DNA that is inherited from parents to children. This genetic material in turn hosts up to 100,000 genetic sites, or 'loci', that can be mapped by geneticists. Only very few of these loci provide any useful information on human origins since the rate of genetic recombination is often too large from generation to generation. The most often used genetic marker is mitochondrial DNA (mtDNA) which is only inherited down the female line.[16] This genetic marker is very rarely subject to mutation and at a rate that is random but with an estimated expected value. Thus, by observing two persons' mtDNA, we can make a rough estimate of how far back these persons had a common ancestor (down the female line). By also taking into account their current geographical residence, researchers are able to construct *phylogeographic* trees, mapping the likely paths of migration of AMH from their East African origins, as well as the proximate dates for these migrations.

There is still not full consensus among researchers regarding the contours of the peopling of the world. Like most other researchers from Stringer et al. (1988) onwards, Oppenheimer (2003) sides with the 'Recent African Origin'-hypothesis (RAO) proposing that all modern human beings n the world today are the descendants of a small population that migrated from Africa and then over several millennia settled the whole world. The competing hypothesis - the 'Multiregional'-hypothesis, suggesting that modern man originated independently in several regions from existing branches of the *homo*-family - is nowadays believed to be false by most scholars (Tishkoff and Verelli, 2003).

A more controversial assumption that Oppenheimer (2003) and Bradshaw Foundation (2007) make is that the first migrants out of Africa did not move through the Levant into the Near East and Europe, but rather through a southern 'beachcombing' route. This route started by crossing the Red Sea at the Gate of Grief between Eritrea and Yemen about 85,000 BP during an ice age with low sea levels. The descendants of this first group outside Africa then followed the beaches of the Indian Ocean towards India, South East Asia, and Australia in a relatively short time. The previous standard hypothesis - still endorsed by many researchers - is that AMH walked out of Africa through the Levant during an earlier warm interglacial period. Recent genetic evidence

[16] Another genetic marker is the non-recombining part of the Y chromosome, only passed down the male line. Research is currently being conducted on the usefulness of other loci for understanding human origins (Tishkoff and Verelli, 2003).

(Macauley et al., 2005), as well as very early archeological findings of AMH in Australia, appear to support a beachcombing route.[17]

Let us then briefly present the broad outlines of the peopling of the world as it is represented in our data. The journey starts 160,000 BP in the Rift Valley area of Ethiopia and Kenya. From here, the rest of continental Sub-Saharan Africa was populated around 135,000 BP. From Eritrea, modern humans crossed the Red Sea to Arabia, as referred to above, and had then spread to most of South Asia including China by 75,000 BP. By 74,000 BP, a gigantic volcanic eruption at Toba in Sumatra left the Indian and South East Asian peninsulas in desolation and presumably extinguished a large part of all humans alive outside Africa. South East Asia was not repopulated until 65,000 BP and India not until 52,000 BP. Meanwhile, AMH presumably settled Australia already 65,000 BP.

From South Asian and Near Eastern origins, Eastern and Southern Europe were finally settled around 45,000 BP, followed by North Africa and Central Asia. By 22,000 BP, human crossed the Bering Strait into North America. Only about 10,000 years later, the whole American continent was settled. Following the retreat of the ice caps from the last ice age, Northern Europe and Scandinavia was populated around 8,000 BP. Islands in the Caribbean and in the Pacific were then gradually reached in the preceding millennia, until French colonists settled the previously uninhabited Seychelles by 1756, which thus is the most recently settled country in our sample. Table 7 contains the estimated Origtime for the 191 countries in our sample.

5 Empirical results

5.1 Measuring ethnic diversity

So far we have discussed ethnic diversity in general terms and have thus avoided being specific on exactly how one should measure diversity. Reducing the multiplicity of ethnic diversity to a one-dimensional measure necessarily means missing some of the political nuances, but that would be less of a problem here since the focus is not on what ethnic diversity does but where it comes from. In the years following Easterly and Levine (1997) the most common measure was ethnolinguistic fractionalization (ELF), where fractionalization is defined as the probability that two randomly selected individuals from a population come from different groups. The larger the number of groups above the chosen threshold value for inclusion, the larger the value on this index. The original ELF-index

[17] See Oppenheimer (2003) for an exhaustive discussion of this issue. A recent attempt to provide a timetable for the peopling of the world based on the northern route is Liu et al (2006).

was constructed from data collected by Soviet ethnographers in the 1960s, and has long stood ground as *the* index to be used for assessing the effects of ethnicity. More recent indices of ethnic fractionalization include those created by Fearon (2003) and Alesina et al. (2003), and the statistical investigation in this section will use the latter measure.[18] A full list of the variables included in this section as well as sources and detailed descriptions are presented in table 6.

The broad picture painted by the bivariate correlations in table 2 is that ethnic fractionalization (AEF) is higher in countries with a longer duration of human settlement (ORIGTIME) and lower in countries with a higher population density in year 1 (POPDEN), a more potent state apparatus after 1800 (MODERN), that are situated farther from the equator (ABSLAT).[19]

<Table 1: Descriptive statistics - about here>

<Table 2: Correlations - about here >

In figures 3 and 4 the degree of ethnic fractionalization is plotted over duration of human settlement, measured as ORIGTIME. A brief inspection tells us that there is a clear positive association between ethnic diversity and time since humans settled, despite the eons that have passed since most continents were populated. The pattern appears both if we look at the world as a whole, as in figure 3, and if we exclude the what could be seen as special cases of sub-Saharan Africa and America, as in figure 4. These two may be seen as special cases in the sense that sub-Saharan Africa has had a considerably longer history of human settlements and also displays the greatest ethnic diversity of the continents, and following the conquest of the Americas large segments of the indigenous populations died off and new segments were introduced.

<Figure 3: Ethnic fractionalization and time since first human settlement - about here>

<Figure 4: Ethnic fractionalization and time since first human settlement. America and SSA excluded - about here>

Though illuminating, bivariate correlations hold limited persuasive power so in the remainder of this section we investigate the empirical strength of our theoretical and historical discussion in OLS regressions.

5.2 Historical factors

In this section we investigate the effects of the historical factors informally introduced in the upper part of figure 2. Needless to say these factors, the in-

[18] These two are available for larger set of countries and our result is not sensitive to which of these we use. Using ELF is not to be recommended for at least two reasons. First, there are doubts regarding the correctness of some of the codings, and secondly the sample is considerably smaller.

[19] In tables 1-5 ORIGTIME is transformed into 10000 years units.

troduction of sedentary agriculture, statehood, and modernity, differ greatly in their time dimension.

< Table 3. Historical factors - about here >

In specification (3.1) of table 3 we find that ORIGTIME alone can explain 27.7% of the observed variation in ethnic diversity. Given that this measure is new to the literature this is a strong result. The size of the coefficient implies that 10000 years earlier human settlement is associated with on average 0.028 points lower ethnic fractionalization.

5.2.1 Agriculture and the rise of human civilization

The influence of the introduction of sedentary agriculture or the rise of civilization are captured by two variables; the timing of the transition from hunter-gatherer to agricultural production (AGRITIME) from Putterman (2007a) and population density around year 1 (POPDEN). Since there are no a priori reasons to expect that areas where the transition occurred first are the same as the ones where this new form of production proved to be most successful the latter measure should better capture the impact of sedentism and civilization. An earlier transition is associated with lower ethnic diversity, see specification (3.2), yet when we add POPDEN the coefficient for AGRITIME shrinks and becomes nonsignificant. A reasonable interpretation is that the actual date of the introduction of sedentary agriculture is not what matters for how the increased societal complexion reduced ethnic diversity. Rather, it is the combined effect on total production of the timing of the transition and geographical factor determining the productivity of agriculture. Gellner (1983) argues that premodern states could not effectively homogenize their populations, while Smith (1986) argues that ethnic identification with a common language and religion was found in early agricultural societies. That a higher population density, which can be supported by relatively more developed societies, is associated with lower ethnic diversity today, is tentative support for the latter perspective. Since homogenizing the population was not a direct policy objective before the modern era it should turn out to be a slow process. In fact, population density in year 1 has a stronger association with lower diversity than population density at other more recent dates, such as in year 1500 or 1900 (not reported).

5.2.2 Statehood and the strength of the modern state

The measure for state antiquity developed by Bockstette et al. (2002) captures the extent to which states have controlled their present territory, and we include this measure (STATEHIST) in specification (3.4).[20] The effect is the expected

[20] We use the updated version of this index, from Putterman (2007b)

- while the positive effects of ORIGTIME and the negative effect of POPDEN remain strong having had more control of the present territory for a considerable time is associated with less diversity, Deliberate homogenizing efforts of the kind that Gellner (1983) and Tilly (1992) discuss requires means and motivation only available rather late in history. In order to investigate whether the estimated effect reflects millennia-long unintended effects of state activities or the more recent phenomena of the rise of the nation-state we disaggregate STATEHIST into two measures, one representing statehood before the modern era (PREMODERN) and one statehood in the modern era (MODERN), with year 1800 as the natural break point. In fact, when both are included simultaneously MODERN is the stronger one - it *was* in the modern era the reduction of ethnic diversity took place. Since replacing STATEHIST with MODERN means that we loose the information on statehood from year 1 to year 1800 the fact that the explanatory power of the model is only weakly affected by dropping PREMODERN is an indication of just how strong homogenizing effect the policies of the modern state have had.

A variable related to STATEHIST and MODERN is for how long the state has been sovereign and independent (INDTIME). Specification (3.8) suggests that the latter variable is the more relevant one yet when we include also a dummy for being colonized by Europeans (EXCOL) only this variable enter as significant. It turns out that if we want to control for past colonial status we cannot include both INDTIME and MODERN. Since MODERN is less correlated with colonial status and has a clearer link to the theoretical literature we prefer this measure.

Before we proceed to a more detailed analysis of the impact of colonialism and geographical variables we hint at the effect of including two of our preferred geographical variables, the absolute latitude of the country (ABSLAT) and a proxy for geography-induced isolation (VEGDIV), see specification (3.12). The most important result, as compared to specification (3.11), is that models without geographical control variables will exaggerate the effect of ORIGTIME and erroneously suggest that colonies still have higher ethnic diversity when other determinants are controlled for.[21] To conclude this subsection we note that even when we control for long term effect of the rise of civilization, the strength of the state during the modern era, colonial status, and relevant geographical factors, the time since humans first settled continues to have a significant effect.

[21] When these factors are controlled for colonial status may actually be associated with *less* ethnic diversity, but the estimate is not statistically significant.

5.3 Colonialism

Former European colonies have higher ethnic fractionalization than other countries. Among the 143 countries included in specification (3.7) the 84 countries that are coded as colonies in Olsson (2007) have an average AEF value of 0.53 while the average for the 49 other countries is 0.33. However, the former colonies outside sub-Saharan Africa have an average of 0.41 which is slightly *lower* than the average of 0.43 for the countries that were not colonized and are not European. Thus, the apparent relationship between colonial status and ethnic diversity is entirely driven by the difference in ethnic diversity between Europe and sub-Saharan Africa. Instead of including a dummy for colonial status we allow all coefficients in specification (3.7) to differ due to colonial status in specification (4.1) and (4.2) of Table 4.

< Table 4. Colonialism - about here>

This is a rather hard test, yet even if some of the estimates do not survive as significant on standard levels of significance the coefficients stay in roughly the same region. The two continents that were arguably most affected by European colonial policies are sub-Saharan Africa and (North and South) America. We split the sample into one for these two continents and one for all other countries and run specification (3.7) on them separately in (4.3) and (4.4). Again the coefficients are not entirely stable but they remain largely significant. More interesting is perhaps that the dummy for colonial status again becomes nonsignificant when countries on the American continent are excluded, and so does the otherwise positive coefficient for time spent under colonial rule (DURATION). Importantly, while the effect of colonial past is sensitive to these changes, so is not the case for the effect ORIGTIME or POPDEN. Furthermore, as seen in specification (4.8) the identity of the colonizer is not of direct importance. Controlling if there is an effect of identity of the colonizing power, we include dummies for being subject to any longer colonization by Great Britain, France, or Spain, and none of these enter as significant. Of the regional dummies for sub-Saharan Africa, America, Asia and Europe (with the Pacific and North Africa as excluded), the first three are positive and significant while the coefficients of our basic variables are fairly robust. The significant region dummies indicate that ethnic diversity differ systematically between countries on different continents, but since the other explanatory variables do so too the interpretation is far from not clear.

When our preferred geographical variables are added to specification (4.1) and (4.2) we find that even if the coefficients for ORIGTIME are the same in both samples it is completely nonsignificant in the smaller sample of countries not colonized by Europeans (4.12). At the same time the coefficients for POP-

DEN (with a p-value of 0.11) and MODERN are even larger (in absolute) for non-colonies. A reasonable interpretation of why the effect of ORIGTIME on AEF is found in the full sample and among former colonies but not among countries with no previous colonial status is that countries in this latter group generally have higher values on MODERN and POPDEN and that the effect of these two factors has been strong enough to eliminate the effect of ORIGTIME. Thus, the effect of duration of human settlement still shows in the full sample since sufficiently many countries have sufficiently short history as densely populated societies and less experience with an own strong modern state apparatus.

So far we have treated colonial status as completely exogenous but dealing with time spans like the ones we have here, tens or even hundreds of thousands of years, geographical variables become inseparable from historical. Geography is a prime determinant of agricultural productivity, access to trade routes, natural resources, etc. and therefore related to both the timing of the Neolithic revolution, as well as population density, strength of the modern state, and even of colonization patterns. More diverse societies could find it harder to unite in order to muster a defense against occupants, looters, or colonizers, and more ethnically homogenous societies may find it easier to gather forces for conquest of new territories. We would expect to observe a higher level of diversity in former colonies, and accordingly in those regions that were colonized, even if colonial policies had been neutral in terms of their effects on ethnic diversity. Thus, since we have no good measure of ethnic diversity in, say, pre-colonial sub-Saharan Africa we cannot tell for sure which way the flow of causality between ethnic diversity and colonial experience runs, if there is any.

5.4 Geography

That human diversity, as the richness of other species, follows geographical patterns was discussed in section 3. While we have no direct measures of isolation and adaption we proxy for their effects by habitat diversity and within-country differences and variation in aspects such as temperature, precipitation, altitude, and vegetation types, i.e. geographical living conditions. What are the expected results on the estimated coefficient for ORIGTIME of including geographical variables? First of all, the humans that settled the world did not do so randomly and ought to have preferred to settle where living conditions were such that they could use their acquired knowledge and skills in order to survive. Thus, the first waves of human settlements must have been directed by factors such as climate and vegetation. Including these in our regressions could naturally reduce the coefficient for ORIGTIME while the variables measuring how much the geography of a country encourages or enables isolation should

enter with positive coefficients in our regressions.

The general pattern found in ecology, that species richness is higher in areas closer to the equator, holds also for diversity among humans - ethnic diversity follows a latitudinal gradient, see specification (5.1) of table 5. This finding is, as is shown by the stability of the coefficient to the inclusion of other geographical controls, a robust finding. Adding distance from the equator (ABSLAT) considerably reduces the coefficient for ORIGTIME, but it is still highly significantly estimated.

One may be lead to believe that countries spanning over a larger territory would tend to have more diversity, but this is not the case. With the proper controls the physical area of the country (AREA) is not related to the level of ethnic diversity.

If population density in the modern era affects level of diversity we should find an effect from the population density in 1900 (POPDEN1900), yet we do not. The fact that population density in 2005 (POPDEN2005) is significantly negatively associated with present ethnic diversity, see specification (5.4), indicates either reversed causality or a spurious relationship since the present population density cannot affect diversity to any measurable extent.

Controlling for distance from equator is sufficient to make the dummy for being colonized by Europeans nonsignificant. Noting that the density of colonization follows a latitudinal gradient this finding suggests that it is the latter factor (ABSLAT) rather than the former (EXCOL) that has a causal relationship with the degree of ethnic diversity. That latitudinal stretch (LATDIFF) is nonsignificant indicates that *if* within-country latitudinal differences affect group formation they do not do so to the extent that it becomes visible in our data.

< Table 5. Geographical variables - about here>

Turning to factors more directly associated with habitat diversity we first note that within-country difference in temperature and altitude are highly correlated (0.88) and including them both also makes both nonsignificant (not reported). Since the exact source of the diversity in geographical living conditions is not in focus here we are satisfied with including one of them, and leave the question of their relative importance open.[22] The richness of different vegetation types (VEGDIV), and within-country differences in temperature (TEMPDIFF) and altitude (ALTDIFFF) are positively associated with diversity, see specifications (5.8) and (5.9). For precipitation both the mean (PRECMEAN) and within-country differences (PRECDIFF) affect the level of ethnic diversity, but

[22] In this article we have no ambition to point out which of these variables that is the most important one. Rather, being able to show that several variables that are associated with isolation are, when taken one by one, associated with more diversity is stronger support.

the magnitude of the effects are small. Also, mean precipitation is negatively associated with diversity. The exact values for the marginal effects of a one standard deviation naturally differs as coefficients differ slightly between the specification. A standard deviation higher ORIGTIME or VEGDIV (see Table 1 for the size of these standard deviations) is in specification (5.8) associated with approximately 0.05 higher AEF, while the corresponding figure for POPDEN or MODERN is 0.05 lower AEF, and for ABSLAT 0.10 lower AEF.

We predicted earlier that of the two effects associated with a larger part of a country's area being close to coast or river, a positive effect on diversity from higher values on ORIGTIME following the beachcombing hypothesis and a negative effect on diversity from the population in these areas being less isolated, the latter (negative) effect is likely to dominate. Testing this argument requires that we include both share of area and share of population in coastal areas. Since the share of land and the share of population are highly correlated, a bivariate correlation of 0.92, including only share of land and not controlling for share of population means that the coefficient on share of land is likely to catch some of the effect of share of population.[23] Specification (5.11) includes both share of land and share of population within 100km from coast or river and while the former is positively (though not significantly so) the latter is negatively associated with diversity - a larger fraction of the population living in less isolated areas is associated with lower levels of ethnic diversity. Since we control for ABSLAT as well as TEMPMEAN and TEMPDIFF this does not simply capture the fact that populations in many tropical countries choose to live in the cooler inland areas. The effect of ORIGTIME, POPDEN, and MODERN remain significant when we control for these geographical variables.

Finally, by including region dummies in (5.12) we effectively kill the coefficient for ORIGTIME. While Europe is the omitted region dummies for Asia, North Africa, and the countries in the Pacific have significantly lower levels of ethnic diversity. This specification is included for the sake of completeness only. If the ambition is to explain a causal relationship including variables like region dummies makes little sense, and if one wants to test the strength of ORIGTIME it is a particularly ill-suited specification. The reason is that two factors make ORIGTIME discontinuously distributed over the continents; first there is the fact that AMH did not settle the Earth in a smooth and continuous fashion, and second there is the problem that there is very limited data on the exact dates for each country. These facts imply that in order to test the strength or ORIGTIME one should both use the full variation in ORIGTIME (omitting a region is

[23] Including present day population is to invite causality problems, but only quite recent population figures with the proper disaggregation are available and we think that it gives a nice illustration.

an unwise strategy) and refrain from including variables that too closely tracks the variation in ORIGTIME (including region dummies effectively cripples the model). The variables POPDEN, MODERN, and ABSLAT do not share this problem, and in fact these are all still significant in specification (5.12).

6 Implications

The empirical analysis does not prove neither primordialists nor social constructivists entirely wrong or entirely right. Yes, there are long–run factors behind differences in ethnic diversity, and yes, a stronger state apparatus in the modern era is associated with increased ethnic homogeneity. Also, on the one hand we find that social indicators such as the rise of civilization and the modern state ultimately have an effect on the ethnic richness of human societies and we argue that these factors affect human diversity both in largely unintended ways, as is the case for the effect of the rise of civilization or sedentism, and in deliberately ways, as was the case for the forceful homogenizing efforts of the modern state. On the other hand, the same geographical factors that affect the richness of other species apparently affect also richness among humans. An important difference is of course that diversity among humans have largely taken the form of ethnic, linguistic, or cultural diversity and that as a worldwide species humans display unusually small genetic variation (Pagel and Mace 2004). Our result shows that a serious understanding of ethnic diversity requires a synthesis of primordial and constructivist arguments.

A reasonable projection of the results implies that the effect of ORIGTIME eventually will vanish in all regions and in all countries (recall the discussion in section 5.3). However, we have deliberately used the term *modern* state since the evidence we present here potentially have a weaker bearing on the effects of how the developed (postmodern) state works in a global economy - it is fully possible that in the really long run lower diversity on the world level comes hand in hand with more diversity at the local level. Since ethnicity represent a mixture of linguistic, cultural, and historical aspects we may be looking forward to an ethnically, culturally, and linguistically poorer world. Traditions and stories will be forgotten, and with them the aggregated experience of billions of humans.

Furthermore, our results have important implications for how social scientists investigate the effect of ethnic diversity on economic and political outcomes, and we will briefly discuss one of these. A commonly used method for assessing the effect of ethnic diversity on economic growth has been to include a measure of ethnic fractionalization as one of many potential regressors. Since we have shown that stronger modern states are associated with lower degrees of ethnic

diversity today, and since there is a positive correlation between MODERN and growth performance since 1960, a negative coefficient on ethnicity in a growth regression could be an effect of the omission of state strength since 1800 from the growth regression.

7 Concluding remarks

Ethnic diversity has caught the attention of many a social scientist struggling to understand problems such as low provision of public goods, low quality governance, persistent economic backwardness, and civil wars. The general approach in much of this research has been to treat ethnic diversity as an exogenous factor and few have explicitly referred to the discussion among primordialists and social constructivists about where ethnicity comes from and why some countries are so much more ethnically diverse than others. In this article we have briefly portrayed this discussion and synthesized it with findings from ecology, anthropology, and medical science showing how geographical and ecological factors matters for human diversity just as for non-human species diversity. We have also constructed a new measure for how long an area has been continuously inhabited by anatomically modern humans and a theoretical model showing how this measure is related to ethnic diversity.

The empirical analysis clearly indicates that ethnic diversity is higher in countries where humans settled earlier, where geographical conditions enable and encourage isolation, and lower in countries where early civilization proved more successful and where the state was stronger during the modern era.

References

[1] Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat and Romain Wacziarg (2003) "Fractionalization", Journal of Economic Growth, 8:155-94.

[2] Alesina, Alberto and Eliana La Ferrara (2005) "Ethnic Diversity and Economic Performance", Journal of Economic Literature, 153(sept), 762-800.

[3] Alesina, Alberto and Enrico Spolaore (1997) "On the Number and Size of Nations", Quarterly Journal of Economics, November 1997.

[4] Alesina, Alberto and Romain Wacziarg (1998) "Openness, Country Size, and Government", Journal of Public Economics, 69:305-321.

[5] Anderson, Benedict (1983) "Imagined Communities", London: Verso.

[6] Arcand, Jean-Louis, Patrick Guillaumont, and Sylvaine Guillaumont Jeanneney (2000) "How to make a tragedy: on the alleged effects of ethnicity on growth", Journal of International Development, 12:925-938.

[7] Bates, R. (2005) "Ethnicity", forthcoming in the Elgar Companion to Development Studies.

[8] van den Berghe, Pierre L. (1981) The Ethnic Phenomenon, Westport: Praeger.

[9] van den Berghe, Pierre L. (1995) "Does Race Matter?", Nations and Nationalism, 1(3):357-68.

[10] Bockstette, Chanda and Putterman (2002) "States and Markets: The Advantage of and Early Start" Journal of Economic Growth, 7:347-69.

[11] Bradshaw Foundation (2007) `http://www.bradshawfoundation.com/stephenoppenheimer/`

[12] Caselli, Francesco and Wilbur John Coleman II. (2006) "On the Theory of Ethnic Conflict" London School of Economics, mimeo.

[13] Cashdan, Elizabeth (2001a) "Ethnic diversity and its environmental determinants: Effects of climate, pathogens, and habitat diversity." American Anthropologist 103:968-991.

[14] Cavalli-Sforza, Luigi Luca, Alberto Piazza, Paolo Menozzi, and Joanna Mountain (1988) "Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data" Proceeding of the National Academy of Science, 85:6002-6006.

[15] Cavalli-Sforza, Luigi Luca, and Francesco Cavalli-Sforza (1995) The Great Human Diasporas: The History of Diversity and Evolution, Cambridge, MA: Perseus Books.

[16] Chanda, Areendam and Louis Putterman (2006) "Early Starts, Reversals and Catchup in The Process of Economic Development", mimeo.

[17] Chaplin, George (2004) "Geographic Distribution of Environmental Factors Influencing Human Skin Coloration", American Journal of Physical Anthropology, 125:292-302.

[18] CIA, World Factbook, available online at `https://www.cia.gov/cia/publications/factbook/index.html`.

[19] Cohen, V.G. (1977) The Food Crisis in Prehistory: Overpopulation and the Origins of Agriculture, New Haven: Yale University Press.

[20] Collard, Ian F. and Robert A Foley (2002) "Latitudinal patterns and environmental determinants of recent human cultural diversity: do humans follow biogeographical rules?", Evolutionary Ecology Research,4:371-83.

[21] Collier, Paul (2000) "Ethnicity, Politics and Economic Performance", Economics and Politics, 12(3):225-45.

[22] Collier, Paul (2001) "Implications of Ethnic Diversity", Economic Policy, April:129-66.

[23] Collier, Paul and Anke Hoeffler (2004) "Greed and Grievance in civil war", Oxford Economic Papers, 56:563-95.

[24] Diamond, Jared (1997) Guns, Germs and Steel: The Fates of Human Societies, New York: W.W. Norton.

[25] Diamond, Jared (2005) "Geography and skin colour", Science, 435(May 19):283-284.

[26] Dunn, M., A. Terrill, G. Reesink, R. Foley, S. Levinson (2005) "Structural Phylogenetics and the Reconstruction of Ancient Language History" Science 309, 23 September.

[27] Easterly and Levine (1997) "Africa's Growth Tragedy: Policies and Ethnic Divisions", The Quarterly Journal of Economics, 112(4):1203-1250.

[28] Encyclopedia Britannica (2007) `http://www.britannica.com`.

[29] Fearon, James D. (2003) Ethnic and Cultural Diversity by Country, Journal of Economic Growth, 8:195-222.

[30] Gallup, John Luke Jeffrey D. Sachs, and Andrew D. Mellinger (1998) "Geography and Economic Development", NBER Working Paper No. 6849.

[31] Gellner, Ernest (1983) "Nations and Nationalism", Oxford: Blackwell

[32] Hansson, Gustav and Ola Olsson (2007) "Country Size and the Rule of Law: Resuscitating Montesquieu", Working Papers in Economics No 200, Göteborg University.

[33] Herbst, Jeffrey (2000) States and Power in Africa: Comparative Lessons in Authority and Control, Princeton University Press.

[34] Jablonski, Nina G. (2004) " The Evolution of Human Skin and Skin Color", Annual Review of Antropology, 33:585-623.

[35] Jones, D. (2000) "Group Nepotism and Human Kinship" Current Anthropology 41(5): 779-809.

[36] Hobsbawm, Eric J. and Terence Ranger (1983) The Invention of Tradition, Cambridge University Press.

[37] Ingman, M., H. Kaessman, S. Pääbo, and U. Gyllensten (2000) "Mitochondrial Genome Variation and the Origin of Modern Humans" Nature 408, 7 December, 708-713.

[38] Kimenyi, Mwangi S. (2006) "Ethnicity, Governance and the Provision of Public Goods", Journal of African Economies, 15(Supplement 1):62-99

[39] Kremer, M. (1993) "Population Growth and Technological Change: One Million B.C. to 1990" Quarterly Journal of Economics 108(3): 681-716.

[40] La Porta, Lopez-de-Silanes, Schleifer and Vishny (1999) "The Quality of Government", The Journal of Law, Economics & Organization, 15(1):222-79

[41] Leeson, Peter T. (2005) Endogenizing fractionalization, Journal of Institutional Economics 1(1):75–98.

[42] Liu, Hua, Franck Prugnolle, Andrea Manica and Francois Balloux (2006) "A Geographically Explicit Genetic Model of Worldwide Human-Settlement History", American Journal of Human Genetics, 79:230-237.

[43] Macaulay, Vincent, and others (2005) "Single, Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes", Science, 308(May 13):1034-6.

[44] Mace, Ruth and Mark Pagel (1995) "A latitudinal gradient in the density of human languages in North America", Proceedings of the Royal Society of London, B-Series, 261:117-21.

[45] McDougall, I., F. Brown, and J. Fleagle (2005) "Stratigraphic Placement and Age of Modern Humans from Kibish, Ethiopia" Nature, 433, 17 February.

[46] Miguel, Edward (2004) "Tribe or Nation? Nation Building and Public Goods in Kenya versus Tanzania", World Politics, 65(april):327-62.

[47] Olsson, O. and D.A. Hibbs (2005) "Biogeography and Long-Run Economic Development", European Economic Review, 49(4):909-938

[48] Olsson, O. (2007) "On the Institutional Legacy of Mercantilist and Imperialist Colonialism" Working Papers in Economics No 247, Göteborg University.

[49] Oppenheimer, Stephen (2003) "Out of Eden: The peopling of the World", London: Constable.

[50] Osterhammel, Jurgen (2005) "Colonialism" Princeton: Markus Wiener Pub.

[51] Pagel, Mark, and Ruth Mace (2004) "The cultural wealth of nations", Nature, 428(18 March):275-8.

[52] Putterman, Louis (2007a) http://www.econ.brown.edu/fac/Louis_Putterman.

[53] Putterman, Louis (2007b) http://www.econ.brown.edu/fac/Louis_Putterman.

[54] Shils, E. (1957) "Primordial, Personal, Sacred, and Civil Ties" British Journal of Sociology, 7:113-45.

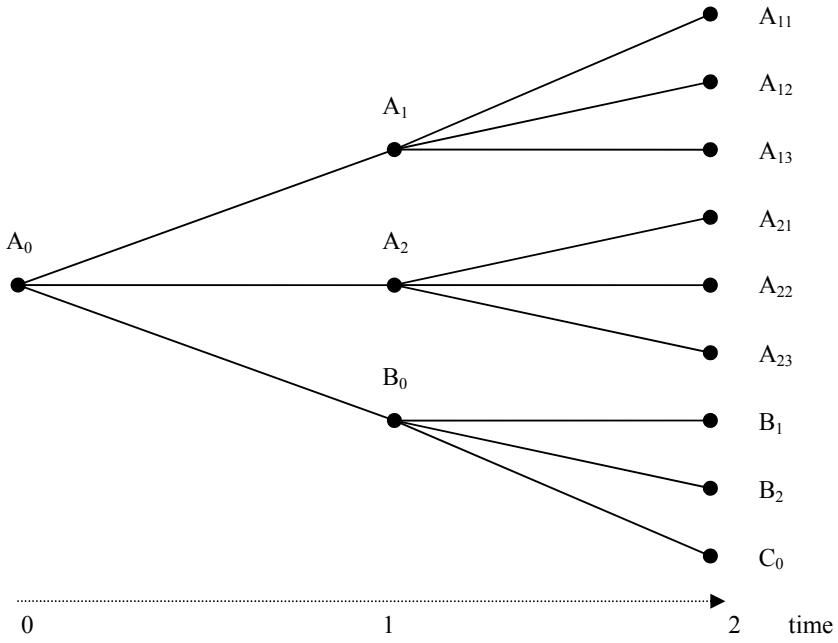
[55] Smith, B.D. (1998) The Emergence of Agriculture, New York: ScientificAmerican Library.

[56] Smith, Anthony D. (1986) The Ethnic Origin of Nations, Oxford: Basil Blackwell.

[57] Tilly, Charles (1992) Coercion, Capital, and European States, AD 990-1992, Oxford: Blackwell.

[58] Tishkoff, S. and B. Verelli (2003) "Patterns of Human Genetic Diversity: Implications for Human Evolutionary History and Disease", Annual Review of Genomics and Human Genetics, 4:293-340.

Figure 1: Ethnic fractionalization on continents A, B, and C as a function of time according to the kinship model of ethnicity.



Note: Founding population A_0 that emerges at time 0 splits into three ethnic groups at time 1 when B_0 becomes the ancestor of all 'out-of-A'-groups. At time 2, ethnic group C_0 populates continent C. The model assumes no barriers to within continent-migration of peoples but significant barriers to migrations between continents. All groups have the same size at time 2. Ethnic diversity on the three continents increase with time since original settlement. On continent A, the Herfindahl index of ethnic fractionalization is 1-6*(1/36) = 5/6, on continent B, 1-2*(1/4) = 1/2, and on continent C 1-1 = 0.

Figure 2: Primordial, social constructivist, and geographical influences on ethnic diversity over time.

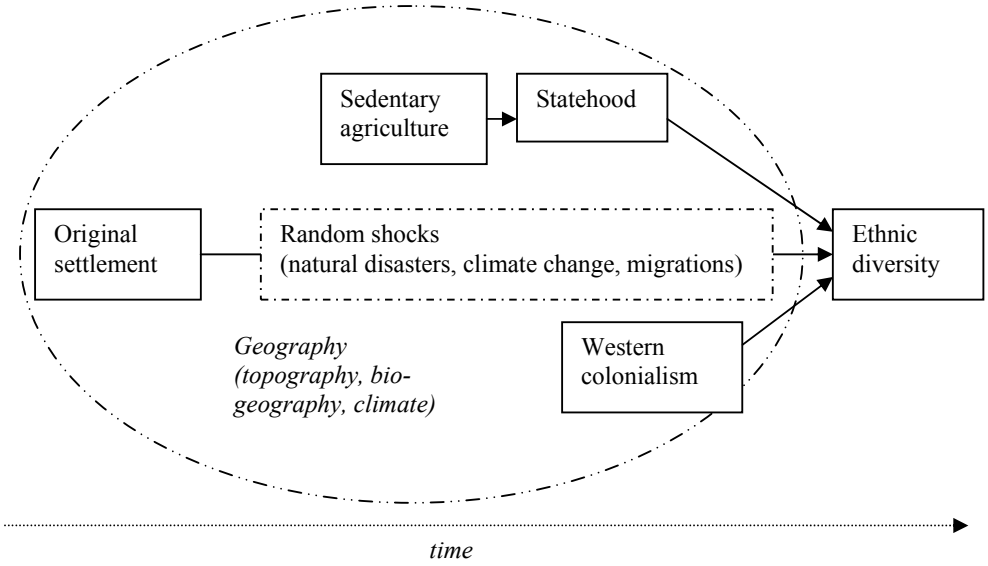


Figure 3: Ethnic fractionalization and time since first human settlement.

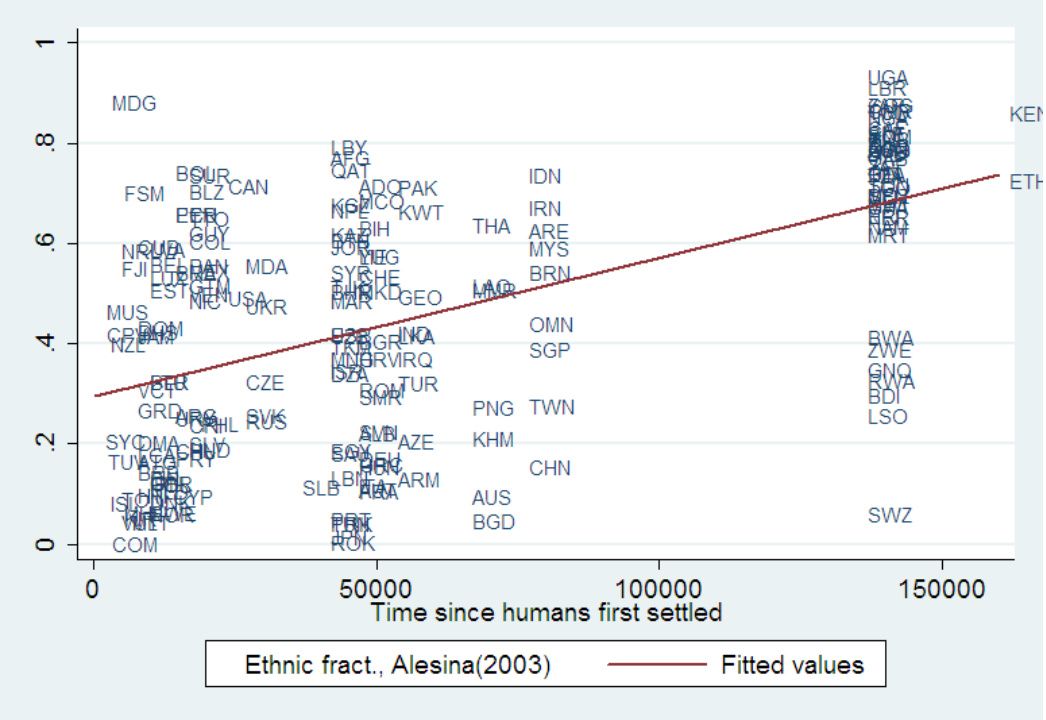


Figure 4: Ethnic fractionalization and time since first human settlement. America and SSA excluded.

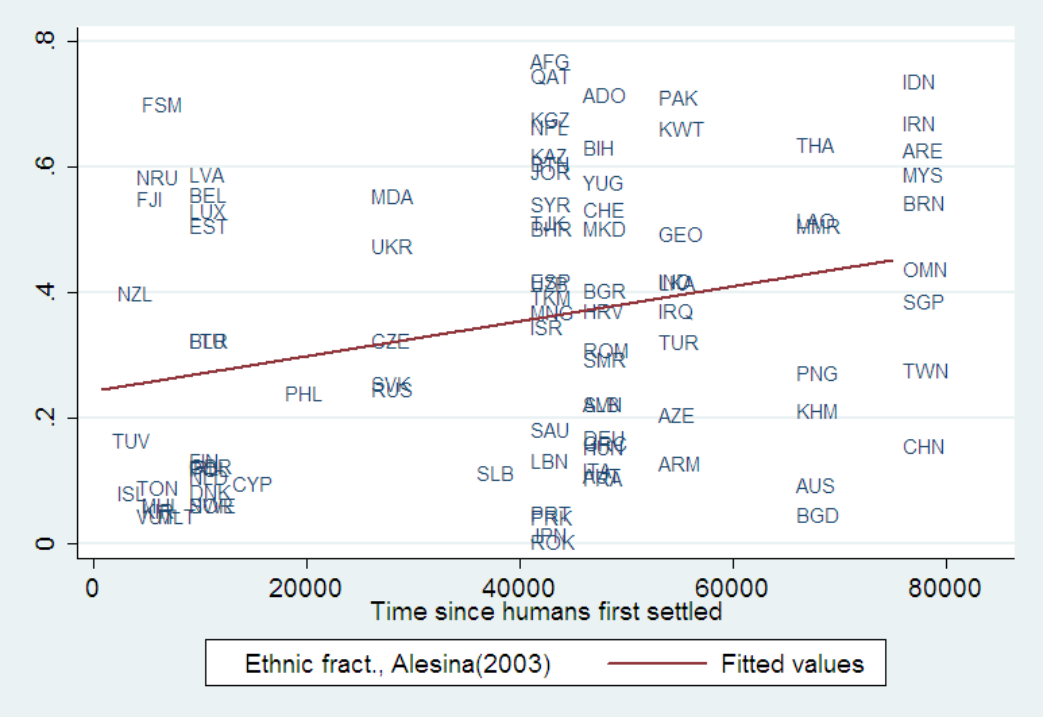


Table 1: Descriptive statistics.

Variable	N	Mean	Sd	Min	Max
AEF	142	0.45	0.25	0.00	0.93
ORIGTIME	142	5.64	4.96	0.02	16.00
POPDEN	142	3.85	8.47	0.00	71.93
MODERN	142	0.67	0.21	0.13	1.00
ABSLAT	142	27.58	17.68	0.23	64.15
VEGDIV	142	0.81	0.20	0.00	1.00

The sample is the same as in Specification (5.8)

Table 2: Correlations

	AEF	ORIGTIME	POPDEN	MODERN	ABSLAT
ORIGTIME	0.5267*				
	(0.0000)				
POPDEN	-0.3250*	-0.1987*			
	(0.0000)	(0.0065)			
MODERN	-0.3291*	-0.3201*	0.0095		
	(0.0001)	(0.0001)	(0.9094)		
ABSLAT	-0.4231*	-0.3810*	0.0797	0.2663*	
	(0.0000)	(0.0000)	(0.2797)	(0.0012)	
VEGDIV	0.1341	0.1085	-0.3459*	0.1984	0.1898
	(0.0770)	(0.1458)	(0.0000)	(0.0157)	(0.0107)

The sample is the same as in Specification (5.8)

Table 3. Historical factors.

	(3.1)	(3.2)	(3.3)	(3.4)	(3.5)	(3.6)	(3.7)	(3.8)	(3.9)	(3.10)	(3.11)	(3.12)
	AEF	**AEF**	**AEF**	**AEF**	**AEF**	**AEF**	**AEF**	**AEF**	**AEF**	**AEF**	**AEF**	**AEF**
ORIGTIME	0.028	0.025	0.025	0.022	0.023	0.021	0.021	0.022	0.019	0.020	0.018	0.010
	(8.36)**	(6.65)**	(7.67)**	(6.13)**	(6.64)**	(5.84)**	(5.81)**	(6.26)**	(5.04)**	(5.48)**	(4.94)**	(2.52)*
AGRITIME		-0.138		-0.108								
		(1.86)+		(1.49)								
POPDEN			-0.005	-0.005	-0.007	-0.008	-0.008	-0.007	-0.008	-0.007	-0.008	-0.006
			(3.73)**	(3.44)**	(3.52)**	(3.70)**	(3.89)**	(3.68)**	(3.81)**	(3.76)**	(3.83)**	(3.13)**
STATEHIST					-0.166							
					(2.33)*							
PREMODERN						-0.084		-0.064				
						(1.29)		(0.98)				
MODERN						-0.186	-0.223		-0.103		-0.197	-0.230
						(2.10)*	(2.66)**		(1.01)		(2.38)*	(2.92)**
INDTIME								-0.026	-0.017	-0.024		
								(2.75)**	(1.55)	(2.66)**		
EXCOL									0.087	0.085	0.101	-0.084
									(2.37)*	(2.33)*	(2.80)**	(1.57)
ABSLAT												-0.008
												(4.64)**
VEGDIV												0.268
												(2.96)**
Constant	0.294	0.390	0.328	0.409	0.415	0.514	0.507	0.412	0.411	0.346	0.445	0.618
	(12.17)**	(7.78)**	(12.85)**	(8.36)**	(9.46)**	(7.35)**	(7.25)**	(10.68)**	(5.51)**	(9.32)**	(6.21)**	(5.73)**
N	184	156	180	154	143	143	143	143	143	143	143	142
R^2	0.28	0.28	0.33	0.33	0.36	0.38	0.37	0.39	0.41	0.41	0.40	0.49

Absolute value of t statistics in parentheses, + significant at 10%; * significant at 5%; ** significant at 1%.

Table 4: Colonialism

	(4.1)	(4.2)	(4.3)	(4.4)	(4.5)	(4.6)	(4.7)	(4.8)	(4.9)	(4.10)	(4.11)	(4.12)
ORIGTIME	0.017	0.034	0.027	0.013	0.026	0.020	0.024	0.017	0.016	0.016	0.011	0.011
	(3.76)**	(3.30)**	(2.46)*	(2.71)**	(5.44)**	(4.17)**	(4.28)**	(3.54)**	(2.57)*	(2.56)*	(2.54)*	(0.68)
POPDEN	-0.007	-0.011	-0.007	-0.008	-0.010	-0.008	-0.010	-0.007	-0.007	-0.007	-0.005	-0.010
	(3.42)**	(1.74)+	(2.29)*	(2.98)**	(3.24)**	(3.53)**	(2.75)**	(2.88)**	(3.81)**	(3.74)**	(2.48)*	(1.61)
MODERN	-0.201	-0.229	-0.233	-0.318	-0.253	-0.198	-0.311	-0.264	-0.263	-0.265	-0.166	-0.270
	(1.64)	(1.88)+	(2.27)*	(2.30)*	(2.55)*	(1.63)	(1.70)+	(1.83)+	(3.05)**	(3.02)**	(1.45)	(2.00)+
EXCOL					0.014					-0.010		
					(0.28)					(0.15)		
DURATION						0.000	0.000					
						(1.73)+	(0.66)					
BRI_COL								-0.008				
								(0.13)				
ESP_COL								0.072				
								(0.85)				
FRA_COL								0.071				
								(1.11)				
SSA									0.229	0.228		
									(2.50)*	(2.48)*		
AMERICA									0.267	0.267		
									(3.26)**	(3.25)**		
ASIA									0.152	0.145		
									(1.99)*	(1.63)		
EUROPE									0.077	0.067		
									(1.00)	(0.66)		
ABSLAT											-0.008	-0.008
											(3.80)**	(1.97)+
VEGDIV											0.280	0.265
											(2.79)**	(1.14)
Constant	0.561	0.426	0.432	0.705	0.469	0.468	0.508	0.560	0.394	0.406	0.478	0.663
	(5.58)**	(4.61)**	(5.09)**	(5.86)**	(6.08)**	(4.14)**	(3.93)**	(4.55)**	(4.63)**	(3.56)**	(4.22)**	(2.57)*
Sample	Colonies	non-Colonies	ssa & america excl.	only ssa & america	america excluded	Colonies	Colonies, america excl.	Colonies	Full Sample	Full sample	Colonies	non-Colonies
N	84	59	76	66	116	84	57	84	142	142	84	58
R^2	0.36	0.24	0.19	0.40	0.47	0.38	0.41	0.38	0.46	0.46	0.47	0.30

Absolute value of t statistics in parentheses, + significant at 10%; * significant at 5%; ** significant at 1%

Table 5. Geographical factors

	(5.1)	(5.2)	(5.3)	(5.4)	(5.5)	(5.6)	(5.7)	(5.8)	(5.9)	(5.10)	(5.11)	(5.12)
ORIGTIME	0.013	0.013	0.012	0.013	0.013	0.013	0.011	0.012	0.014	0.010	0.011	0.006
	(3.38)**	(3.29)**	(3.18)**	(3.41)**	(3.21)**	(3.29)**	(2.81)**	(3.28)**	(3.55)**	(2.19)*	(2.21)*	(0.99)
POPDEN	-0.008	-0.008	-0.005	-0.007	-0.008	-0.007	-0.006	-0.007	-0.007	-0.008	-0.011	-0.008
	(4.19)**	(4.06)**	(1.94)+	(3.70)**	(4.25)**	(3.89)**	(3.08)**	(3.93)**	(3.96)**	(3.99)**	(3.86)**	(4.14)**
MODERN	-0.183	-0.207	-0.182	-0.176	-0.186	-0.215	-0.222	-0.257	-0.232	-0.235	-0.222	-0.366
	(2.32)*	(2.55)*	(2.32)*	(2.26)*	(2.36)*	(2.58)*	(2.80)**	(3.17)**	(2.93)**	(2.87)**	(2.66)**	(4.12)**
ABSLAT	-0.005	-0.005	-0.005	-0.005	-0.006	-0.005	-0.006	-0.006	-0.005	-0.006	-0.005	-0.005
	(4.63)**	(4.75)**	(4.61)**	(4.94)**	(3.71)**	(4.70)**	(5.35)**	(3.00)**	(4.67)**	(4.39)**	(2.28)*	(2.13)*
AREA		0.000										
		(1.25)										
POPDEN1900			-0.001									
			(1.27)									
POPDEN2005				-0.000								
				(1.77)+								
EXCOL					-0.050							
					(0.93)							
LATDIFF						0.003						
						(1.39)						
VEGDIV							0.244					
							(2.72)**					
TEMPMEAN								-0.002			0.001	0.003
								(0.55)			(0.17)	(0.57)
TEMPDIFF								0.006			0.006	0.011
								(2.85)**			(2.28)*	(4.13)**
ATLMEAN									-0.027			
									(0.65)			
ALTDIFF									0.047			
									(2.68)**			
PRECMEAN										-0.001		
										(2.25)*		

Continued ...

Table 5. Geographical factors. *Continued* ...

	(5.1)	*(5.2)*	*(5.3)*	*(5.4)*	*(5.5)*	*(5.6)*	*(5.7)*	*(5.8)*	*(5.9)*	*(5.10)*	*(5.11)*	*(5.12)*
PRECDIFF										0.000		
										(2.05)*		
LAND100CR											0.196	
											(1.61)	
POP100CR											-0.261	
											(2.34)*	
SSA												-0.011
												(0.13)
AMERICA												-0.066
												(0.86)
ASIA												-0.125
												(1.98)*
PACIFIC												-0.309
												(2.80)**
NAFRICA												-0.228
												(2.18)*
Constant	0.661	0.674	0.672	0.669	0.728	0.662	0.519	0.740	0.641	0.792	0.703	0.752
	(9.03)**	(9.13)**	(9.13)**	(9.20)**	(7.09)**	(9.04)**	(5.90)**	(4.92)**	(8.93)**	(7.61)**	(4.32)**	(4.47)**
N	143	143	143	140	143	142	142	142	142	142	130	142
R^2	0.45	0.46	0.46	0.47	0.46	0.46	0.48	0.49	0.49	0.48	0.56	0.54

Absolute value of t statistics in parentheses, + significant at 10%; * significant at 5%; ** significant at 1%

Table 6. Variable descriptions

Variable name	Description	Source
ABSLAT	Absolute latitude in degrees	Cepii
AGRITIME	Years since Neolithic revolution (in 10 000s years)	Putterman (2007a)
ALTDIFF	Maximum difference in mean grid cell altitude*	Authors' calculation based on the G-Econ Dataset (2006)
AMERICA	Dummy for (South and North) America	Cepii
AREA	Surface area in 1000's square kilometers	Cepii
ASIA	Dummy for Asia	Cepii
ALTMEAN	Average grid cell altitude*	Authors' calculation based on the G-Econ Dataset (2006)
BRI_COL	Dummy for longer British colonization	Cepii
DURATION	Duration of colonization by Europeans	Olsson (2007)
ESP_COL	Dummy for longer Spanish colonization	Cepii
EUROPE	Dummy for Europe	Cepii
EXCOL	Dummy for being colonized by Europeans	Olsson (2007)
FRA_COL	Dummy for longer French colonization	Cepii
INDTIME	Years since independence	ICOW
LAND100CR	Share of surface area within 100km of Sea or river	Gallup, Sachs, & Mellinger (1998)
LATDIFF	Maximum difference in mean grid-cell latitude*	Authors' calculation based on the G-Econ Dataset (2006)
MODERN	Statehood strength after 1800	Authors' calculation based on Putterman (2007b)
NAFRICA	Dummy for North Africa	Cepii
ORIGTIME	Duration of human settlement (in 10 000s years)	See detailed description in section 4
PACIFIC	Dummy for the Pacific	Cepii
POP100CR	Share of population within 100km of Sea or river	Gallup, Sachs, & Mellinger (1998)
POPDEN	Population density in year 1 AD	Estimation from Worldmapper (2006)
POPDEN1900	Population density in year 1900	Estimation from Worldmapper (2006)
POPDEN2005	Population density in year 2005	WDI online
PRECDIFF	Max. difference in mean grid-cell precipitation*	Authors' calculation based on the G-Econ Dataset (2006)
PRECMEAN	Average grid-cell precipitation*	Authors' calculation based on the G-Econ Dataset (2006)
PREMODERN	Statehood strength before 1800	Authors' calculation based on Putterman (2007b)
SSA	Dummy for sub-Saharan Africa	WDI online
STATEHIST	Statehood strength ("state antiquity")	Putterman (2007b)
TEMPDIFF	Maximum difference in mean grid-cell temperature*	Authors' calculation based on the G-Econ Dataset (2006)
TEMPMEAN	Average grid-cell temperature*	Authors' calculation based on the G-Econ Dataset (2006)
VEGDIV	Diversity of vegetation types**	Authors' calculation based on the G-Econ Dataset (2006)

* In the G-Econ dataset variables are reported for each grid cell within countries. Each grid cell corresponds to an area of 1 degree latitude times 1 degree longitude, which is approximately 100km by 100km and according to G-econ (2006) "approximately the same size as the second level political entities in most countries (e.g., counties in the United States)." ** Dominant vegetation type (of a list of 27 types) are listed for each grid-cell. VEGDIV is calculated as 1 - (#different types listed for each country/#grid cells for each country). Thus if all the same vegetation type dominates all grid cell VEGDIV=1, if there are two grid cells in the country with different vegetation types VEGDIV=0.5. The G-Econ Data set is available online: <http://gecon.yale.edu/>. The Bradshaw Foundation (2007) is an online source <http://www.bradshawfoundation.com>. The Worlmapper (2006) is an online source <http://www.sasi.group.shef.ac.uk/worldmapper/index.html>. The Cepii data set is available online: <http://www.cepii.fr/>. Independence is coded as "the date on which this state became independent - i.e., acquired control of its own foreign policy, without being ruled by a foreign power" and is drawn from the Issue Correlates of War (ICOW) Project <http://garnet.acns.fsu.edu/%7Ephensel/icowdata.html#names>.

Table 7. Origtime

Country / Economy	Origtime	Country / Economy	Origtime	Country / Economy	Origtime
Afghanistan	40000	China	75000	Haiti	6000
Albania	45000	Colombia	15000	Honduras	15000
Algeria	40000	Comoros	1500	Hungary	45000
Andorra	45000	Congo, DRC	135000	Iceland	1200
Angola	135000	Congo, Republic of the	135000	India	52000
Antigua and Barbuda	6000	Costa Rica	15000	Indonesia	75000
Argentina	12500	Cote d'Ivoire	135000	Iran	75000
Armenia	52000	Croatia	45000	Iraq	52000
Australia	65000	Cuba	6000	Ireland	8000
Austria	45000	Cyprus	12000	Israel	40000
Azerbaijan	52000	Czech Republic	25000	Italy	45000
Bahamas, The	6000	Denmark	8000	Jamaica	6000
Bahrain	40000	Djibouti	135000	Japan	40000
Bangladesh	65000	Dominica	6000	Jordan	40000
Barbados	6000	Dominican Republic	6000	Kazakhstan	40000
Belarus	8000	Ecuador	12500	Kenya	160000
Belgium	8000	Egypt	40000	Kiribati	3500
Belize	15000	El Salvador	15000	Korea, North	40000
Benin	135000	Equatorial Guinea	135000	Korea, Rep.	40000
Bhutan	40000	Eritrea	135000	Kuwait	52000
Bolivia	12500	Estonia	8000	Kyrgyzstan	40000
Bosnia and Herzegovina	45000	Ethiopia	160000	Laos	65000
Botswana	135000	Fiji	3000	Latvia	8000
Brazil	12500	Finland	8000	Lebanon	40000
Brunei	75000	France	45000	Lesotho	135000
Bulgaria	45000	Gabon	135000	Liberia	135000
Burkina Faso	135000	Gambia, The	135000	Libya	40000
Burma	65000	Georgia	52000	Liechtenstein	45000
Burundi	135000	Germany	45000	Lithuania	8000
Cambodia	65000	Ghana	135000	Luxembourg	8000
Cameroon	135000	Greece	45000	Macedonia	45000
Canada	22000	Grenada	6000	Madagascar	1300
Cape Verde	500	Guatemala	15000	Malawi	135000
Central African Republic	135000	Guinea	135000	Malaysia	75000
Chad	135000	Guinea-Bissau	135000	Maldives	2500
Chile	12500	Guyana	15000	Mali	135000

Table 7. Origtime. *Continued…*

Country / Economy	Origtime	Country / Economy	Origtime	Country / Economy	Origtime
Malta	5000	Saint Vincent and the Grenadines	6000	United Arab Emirates	75000
Marshall Islands	3500	Samoa	3000	United Kingdom	8000
Mauritania	135000	San Marino	45000	United States	22000
Mauritius	500	Sao Tome and Principe	500	Uruguay	12500
Mexico	15000	Saudi Arabia	40000	Uzbekistan	40000
Micronesia, Federated States of	3500	Senegal	135000	Vanuatu	3000
Moldova	25000	Serbia and Montenegro	45000	Venezuela	15000
Monaco	45000	Seychelles	200	Vietnam	75000
Mongolia	40000	Sierra Leone	135000	Yemen	85000
Morocco	40000	Singapore	75000	Zambia	135000
Mozambique	135000	Slovakia	25000	Zimbabwe	135000
Namibia	135000	Slovenia	45000		
Nauru	3000	Solomon Islands	35000		
Nepal	40000	Somalia	135000		
Netherlands	8000	South Africa	135000		
New Zealand	1200	Spain	40000		
Nicaragua	15000	Sri Lanka	52000		
Niger	135000	Sudan	135000		
Nigeria	135000	Suriname	15000		
Norway	8000	Swaziland	135000		
Oman	75000	Sweden	8000		
Pakistan	52000	Switzerland	45000		
Palau	4500	Syria	40000		
Panama	15000	Taiwan	75000		
Papua New Guinea	65000	Tajikistan	40000		
Paraguay	12500	Tanzania	135000		
Peru	12500	Thailand	65000		
Philippines	17000	Togo	135000		
Poland	8000	Tonga	3000		
Portugal	40000	Trinidad and Tobago	15000		
Qatar	40000	Tunisia	40000		
Romania	45000	Turkey	52000		
Russia	25000	Turkmenistan	40000		
Rwanda	135000	Tuvalu	700		
Saint Kitts and Nevis	6000	Uganda	135000		
Saint Lucia	6000	Ukraine	25000		