School of Business, Economics and Law
GÖTEBORG UNIVERSITY

WORKING PAPERS IN ECONOMICS

No 281

The Roots of Ethnic Diversity

by

Pelle Ahlerup
Ola Olsson

December, 2007

ISSN 1403-2473 (print)
ISSN 1403-2465 (online)

SCHOOL OF BUSINESS, ECONOMICS AND LAW, GÖTEBORG UNIVERSITY

Department of Economics
Visiting adress Vasagatan 1,
Postal adress P.O.Box 640, SE 405 30 Göteborg, Sweden
Phone + 46 (0) 31 786 0000

The Roots of Ethnic Diversity

Pelle Ahlerup* Ola Olsson

December 10, 2007

Abstract

The level of ethnic diversity is believed to have significant consequences for economic and political development within countries. In this article, we provide a theoretical and empirical analysis of the determinants of ethnic diversity in the world. We introduce a model of cultural and genetic drift where new ethnic groups endogenously emerge among peripheral populations as a response to an insufficient supply of public goods. In line with our model, we find that the duration of human settlements has a strong positive association with ethnic diversity. Ethnic diversity decreases with the length of modern 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 the probability that two randomly chosen individuals belong to different ethnic groups only 0.2 percent in South Korea, whereas the same probability is roughly 93 percent in Uganda?[1]

The broad aim of this article is to offer theory and evidence to improve our understanding of the determinants of ethnic diversity across the world. 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

*Corresponding author: Dept of Economics, Göteborg University, Box 640, 405 30 Göteborg, Sweden. Email: pelle.ahlerup@economics.gu.se. We have received useful comments from Carl-Johan Dalgaard, Stelios Michalopoulos, Andrea Mitrut, Karl-Ove Moene, Brian Wood, and seminar participants at University of Copenhagen, Göteborg University, and the Nordic Workshop in Development Economics. Olsson gratefully acknowledges financial support from SIDA, Vetenskapsrådet, and the Wallander-Hedelius foundation.

[1] The estimates are taken from Alesina et al. (2003).

primordial view, contending that ethnic divisions have deep roots in history and ecology and should be analyzed in an evolutionary framework. A key prediction from our 'kinship model' of ethnicity is that the antiquity of human settlement should be positively correlated with current levels of ethnic fractionalization. The intuition behind this hypothesis is that among more or less geographically isolated prehistorical hunter-gatherer populations, random genetic and cultural drift that accumulated over time repeatedly caused new groups to form in order to secure an efficient provision of public goods. We argue that the ethnic legacy from prehistory should still be detectable in the current didstribution of ethnic groups.

In order to empirically identify this effect, 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. The hypothesis regarding the effects of the duration of human settlements receives strong empirical support, also when we have controlled for other relevant variables. In fact, the 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 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. Arcand et al. (2000) later disputed Easterly and Levine's proposed link from ethnic diversity to poor economic performance, a lower provision of public goods, yet find that the negative 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 uniform. Collier (2000) finds it only in non-democratic countries and Easterly (2001) claims that the effect is weaker where institutional quality is higher. Alesina and La Ferrara (2005) confirm Easterly and Levine's (1997) basic results and find that the negative effect is less pronounced in rich countries, and that after controlling for this effect the impact of democracy is nonsignificant. Moreover, 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] The provision of public goods in ethnically diverse societies tend to be biased towards excludable goods rather than on non-excludable goods, such as education and defence (Alesina and Wacziarg, 1998; Kimenyi, 2006). 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 creation of two new indices on ethnic fractionalization: Alesina et al (2003) and Fearon (2003). In our empirical section, we use the index created by Alesina et al (2003) but our results are robust to using Fearon's (2003) index.

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 (since circa 1800) for the purpose of uniting disparate nations into states (Gellner, 1983; Anderson, 1983; Hobsbawm and Ranger, 1983). The vehicles for the achievement of nation-states were to be found in a combination of efficient printing technology, universal literacy, and industrialization that broke up traditional societies.

The primordial tradition, on the other hand - with somewhat less influential proponents like Smith (1986) and van den Berghe (1981) - argues that ethnic identification is a natural and indeed rational behavior that has existed throughout history. In van den Berghe's sociobiological model of ethnicity, ethnic groups are regarded as *extended kinships* that are successful as a means of social organization because cooperation based on kin has evolved as an evolutionary favorable strategy for solving collective action problems. The primordial kinship model 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 the scientific community's view 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 research on the human genome has so far had almost no impact in social science.

Ethnicity is almost always taken as a given in economics. A recent exception to this rule is Leeson (2005) who argues on the basis of historical examples that colonial policy was often purposefully directed towards increasing ethnic fractionalization. In an interesting study of a neighboring communities in Kenya and Tanzania, Miguel (2004) shows that even among former colonies, govern-

[2] The results in Alesina et al. (2003) are not robust to including latitude as a control variables. Since we show that latitude is a factor *causing* ethnic diversity their approach, to exclude latitude, is not without problems.

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

ments might have an important role to play in fostering a national identity over tribal identification. 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.

Despite the widespread recognition of the importance of ethnicity in politics and economic development, we are aware of only one other systematic attempt to account for the international variation in ethnic diversity. In a current working paper, Michalopoulos (2007) argues that geographical variation in a given land area should reduce inter-regional migration and lead to more ethnic groups.[4] This prediction concerning the role of geographical frictions receives empirical support in a cross-country analysis and in a study of a large number of adjacent pairs of regions. Our approach differs in some important ways from that of Michalopoulos (2007), especially in that our theory of ethnic diversity is centered on public goods-provision and on cultural drift, rather than on human capital transmission. Our empirical focus is further on the time since original settlement which we find strong support for in the empirical analysis.[5]

The only other paper that also explicitly takes into account genetic and cultural drift is Spolaore and Wacziarg (2007). Using data on genetic distances from Cavalli-Sforza et al (1994), the authors show that income convergence appears to be faster among genetically 'close' countries due to a stronger diffusion of technology. In our model, on the other hand, genetic and cultural drift cause the public goods-provision from the core to the periphery to deteriorate, which eventually results in that people in the periphery choose to break away and form new groups.

This brief literature overview suggests that our article makes at least three contributions to the existing literature: Firstly, our article is the first to offer a public goods-based model of endogenous ethnic group formation with genetic and cultural drift as the engine of ethnic fractionalization. Secondly, our article provides the first attempt at measuring the duration of uninterrupted human settlement in a cross-country setting. Thirdly, we provide a comprehensive empirical assessment of the determinants of ethnic diversity across the world. Our main findings 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 and that state experience during modernity is another key factor. These results are particularly relevant for our understanding of the reasons behind the dismal performance of economic growth in Africa and how ethnic diversity fits into the overall picture.

[4]See also Ashraf and Galor (2007) for a model where geography has an important effect on the level of cultural assimilation.

[5]Our empirical findings also show that there is an important role for geographical diversity, as predicted by our own model as well as by Michalopoulos (2007).

The article is structured as follows. In section two, we argue that the diverse literature on the constructivist and primordial explanations can be combined with recent ecological and genetic research into one joint framework for understanding ethnic diversity. In section three, we outline our model of ethnic fractionalization and formation. In section four, we briefly discuss the construction of our measure for the duration of human settlements and present the main empirical results. In section five we conclude the discussion and allow ourselves to speculate about the theoretical and empirical implications of our findings.

2 Literature overview

In this section we discuss the main theories of ethnic diversity in some detail, but first we need to define what an ethnic group is. An ethnic group is a social entity with two basic features; 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, and the group currently forms a cultural community, manifested for instance in common language, religion, and customs (Fearon, 2003; Bates, 2005). Ethnicity is distinct from concepts such as race and nation. Although both latter 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 primarily to be used in conjunction with discussions on (nation-)state formation during modernity (Gellner, 1983).

2.1 The primordial view

At the core of the modern primordial view lies an emphasis of the history and origins of ethnic groups. 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. 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?

One potential explanation is provided by the sociobiological 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 (Jones, 2000). 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. The evolutionary logic dictates that individuals 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.[6]

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 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. In line with this logic, we conjecture that a primary reason for the existence of ethnic groups is their role in organizing the provision of public goods.

A straightforward implication of van den Berghe's theory is that we should expect that distinguishable extended kinships of the type described above have existed throughout most of human history. It is by now generally agreed that the history of 'anatomically modern humans' (AMH) goes back roughly 200,000 years.[7] Genetic research on human origins have suggested that all human beings in the world today have originated from a founding population of a few thousand individuals living in East Africa (Oppenheimer, 2003).

As AMH migrated from their East African home to other parts of sub-Saharan Africa, an inevitable process of ethnic (and genetic) fractionalization started. Since public goods could not be effectively provided over long distances, groups necessarily soon organized in relatively small units. A result of this

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

[7] 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).

process was *genetic* and *cultural drift*. Genetic drift is a general tendency for genetic diversity to be reduced among isolated populations as time passes.[8] If there were initially for instance five lineages in a founding group - labelled A, B, C, D, and E - there would after say ten generations maybe just be the D-lineage left so that all subsequent offspring had D as their ancestor. As we shall see, although genetic drift is a random process, the rate at which it occurs has an estimated expected value.[9] Cultural drift is in an analogous manner the tendency for multiple cultural traits to be reduced within an isolated population. Cultural drift implies that two groups that initially shared the same culture should - after a number of generations in isolation - display two quite distinct sets of cultural characteristics, often different enough for all parties to recognize them as two different ethnic groups (Cavalli-Sforza et al, 1994). However, genetic and cultural drift does not only occur between groups, it is also present within groups, and can emerge over time as a result of clustering at for instance village level when men do not interbreed with women from other villages.[10] Eventually, such drift causes even non-migrating peoples to split up into distinct ethnic groups.

In order to get an idea of how such a fractionalized prehistorical society might have looked like, it is illustrative to consider Papua New Guinea (PNG), where isolated primitive peoples have populated the greater part of the country to this day. In PNG, it is estimated that about 820 different languages are currently spoken among its 5.6 million inhabitants (CIA, 2007). Two factors appear to have contributed to this enormous diversity; first, PNG:s extreme geography with mountains and impenetrable rain forests where groups easily became isolated, and second, PNG is believed to have been populated for some 65,000 years and is therefore one of the countries with the oldest presence of AMH outside Africa.

Our view of ethnic fractionalization thus implies that there should be strong linkages between the formation of genetic and ethnic groups. The most intuitive reason for this is of course that both cultural and genetic characteristics are essentially transmitted from parents to children. 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

[8] A closely related concept is the "founder's effect" which arises when only a small fraction of the whole population moves on to establish a new colony. Clearly, such a small group 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.

[9] A third factor behind genetic changes, distinct from migration and genetic drift, is *natural selection*.

[10] Cavalli-Sforza et al (1994, p 113) discuss an investigation showing that observable genetic differences can be found in a population of 37 villages in the upper Parma Valley in Italy.

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.

Recent studies on the human genome have produced genealogical trees showing how related or genetically distant populations around the world are. Figure 1 shows one such phylogenetic tree from Oppenheimer (2003), based on an often-cited study that uses mitochondrial DNA (mtDNA) from 53 individuals from around the world (Ingman et al, 2000). As expected, the tree shows that the deepest genetic branches are found in Africa and that all non-Africans descend from a branch that emerged relatively recently some 83,000 years ago (L3). One of the most recent fissions in the figure occurred around 36,000 BP when the English, French, and Dutch individuals in the sample had their latest common ancestor down the female line. On the basis of the evidence above, we argue that since cultural and genetic drift appear to be inevitable features of all hunting-gathering societies we should observe that territories with a long settlement history during pre-historical times are ethnically more diverse than countries with a shorter settlement history, all other factors held constant.

2.2 The constructivist view

Contrasting the reasoning behind this in essence primordial hypothesis is the constructivist discourse, which provides us with a plethora of more recent factors with potential to affect current levels of ethnic diversity. Figure 2 outlines some of these factors. A dramatic turning point in human history was 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 a few 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, what is usually referred to as 'civilization'. 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 from the turn of the nineteenth century was not possible. On the other hand, Smith (1986) finds that 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. A reasonable conjecture from these observations is that within states extended kinships partly lose their raison d'être, the role as the most efficient mode of organizing collective action. State institutions like codified law, courts, taxation, and military protection substitute for the services provided by extended kinships which is the reason why many small ethnic groups disappear in such an environment. The implied hypothesis is that the length of statehood experience (and the associated time since the transition to agriculture and civilization) will exercise a negative influence on ethnic diversity.

States have not only created institutions that passively reduced heterogeneity, but have also actively pursued policies designed to result in more homogenous populations. This process gained momentum as the modern industrial European states at the turn of the eighteenth century acquired both means and motivation for nation-building. The modern industrialized society's increasing division of labor created, in combination a rapidly changing production, problems for which the creation of a dynamic and mobile workforce was a solution. The industrial society required strangers to easily communicate with and understand each other, and therefore demanded sufficient homogeneity in both language and culture (Gellner, 1983).

Another driving force behind deliberate attempt to obtain a homogenous population was related to ability to wage and win wars. Refering to the European experience Tilly (1992:106) finds 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." In Europe this process started well before the era of industrialization, when direct rule replaced indirect rule by intermediaries. The extent of homogeneity in the population was in fact both the final outcome of the process as well as a factor that could make the process faster and more effective - it is easier to unite a population that is not to diverse to begin with. The process whereby the European states encouraged national rather than ethnic/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 primary provider of services previously provided at the local level.[11]

Another major historical event with potentially dramatic repercussions was European colonialism from the fifteenth century onwards. This extremely heterogeneous process, coherently analyzed in Osterhammel (2005) and Olsson (2007), is often though to have 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. In stark contrast to the homogenizing domestic policies of the time the Europeans that ruled the colonial states had no strong incentives towards ethnic homogenization in the colonial countries since these were created only to benefit the colonial power. 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. However, 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 were forcefully introduced in the form of slaves of African descent. Whether or not ethnic diversity should increase with colonial experience is an empirical issue.

2.3 Geography and ecology

Apart from these influences there are factors at the micro level that have the potential to influence the degree of ethnic diversity. A stylized fact from ecology is that species richness, or diversity, is a product of isolation and adaptation, and that it increases as we get closer to the equator. 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 as well as for

[11] Tilly (1992:115-6) finds that 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."

birds. They also find that in this pre-colonial environment, linguistic diversity was higher in areas with more habitat diversity. 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 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, since paleoanthropology and medical science have shown that variation in human skin color comes partly from differences in UV radiation, which in turn is determined 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 yet had enough time to adjust (Diamond 2005), which implies that similarity of skin color is a weak predictor of close genetic connections (Jablonski 2004).

The impact of latitude on ethnic diversity is 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. Over time these differences between habitats, on matters such as soil and vegetation types, should lead to the creation of ethnic groups in premodern societies, as discussed above. Collard and Foley (2002) find that the number of 'cultures' within a certain area follows geographical patterns, falls with latitude, rises with temperature and rainfall, and that this pattern holds both in 'new continents' such as the Americas, and 'old continents' such as Africa.

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

[12]Skin color affects production of vitamin D, and protection againts skin cancer (Diamond 2005).

[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

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 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.

As any species spread out from its origin, genetic diversity declines naturally due to genetic drift and founder's effects, as discussed above. 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 suggests 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. It is reasonable to assume that over the millennia the latter effect should eventually come to dominate.

3 A model of ethnic fractionalization

In this section, we develop a model of ethnic fractionalization where ethnic groups essentially are a kinship-based type of social organization with the primary purpose of supplying public goods. Members of the initial group might potentially break away and supply their own public goods if their distance to the initial provider in terms of geography, kinship, or preferences have become too large. Two settings will be analyzed; fractionalization among hunter-gatherers and fractionalization under agriculture and statehood.

3.1 Basics

Let us assume an ethnic group of hunter-gatherers whose population is uniformly spread along a one-dimensional geographical area of total size s. The area has just been colonized by the group and no fissions have yet occurred. Individuals in this group receive positive utility from consumption and leisure. The primary reason that the ethnic group exists is to solve the collective action problem inherent in public goods provision. The public good could for instance be primitive institutions for hunting coordination, dispute settlement, religious ceremonies, and so on, i.e. the services provided by a chieftainship or primitive government. The chief is the real or imagined 'founding father' of the ethnic group and the group is thus an extended kinship. Cooperation among its

easily domesticated mammals (such as the wild ancestors of sheep, goat, and cattle).

members (kinship nepotism) has evolved as an evolutionary favorable strategy throughout the history of human development and will be taken as given here. As we shall see, however, those members who have the strongest family ties with the chief also live closer to him and benefit more from the public good.

The utility of each individual i in the ethnic group j is given by the utility function

$$U_i(c_i, l_i) = \alpha \ln c_i + (1-\alpha) \ln l_i \tag{1}$$

where c_i is private consumption, l_i is leisure, and $\alpha \in (0,1)$ is a parameter. The loglinear form ensures diminishing marginal utility in each of the two arguments.

Consumption is the difference between the value of individual production y_i and the tributes paid for the provision of the public good τ_i. For mathematical convenience, we define consumption as

$$c_i = \frac{y_i}{\tau_i} \geqslant \bar{c} > 1 \tag{2}$$

The ratio of production to tributes must exceed a subsistence level $\bar{c}$ which in turn must obviously be greater than unity. Only regions where this subsistence condition can be met will be populated.

Individual production y_i is given by the production function

$$y_i = g_i e_i^{\gamma} \left(\omega L_i\right)^{1-\gamma}. \tag{3}$$

In this expression, g_i is the level of the public good individual i benefits from, e_i is the individual effort induced into production, L_i is the land available to each individual, ω is the quality of the land, and $\gamma \in (0,1)$ is an elasticity parameter that yields constant returns to scale in effort and land. For simplicity, we will normalize land to be $L_i = 1$ for all i.[14] Each individual's total endowment of time is normalized to unity, implying that $l_i = 1 - e_i$.

3.2 Public goods and cultural distance

The key factor in the model is g_i, the level of the public good experienced by individual i. This level is

$$g_i = g\left(1 - \phi d_i m_i\left(t\right)\right) \geq 0 \tag{4}$$

where g is the level of the public good at its source (i.e. where the chief lives), ϕ is a parameter indicating the general geographical frictions for the spatial

[14] We implicitly assume that land is not an excludable, rival, or constraining factor for hunter-gatherers during this era of very low population densities. This is not an entirely realistic assumption but allows us to focus on more central aspects.

diffusion of the public good in the region, d_i is the geographical distance between individual i:s location and that of the chief, and $m_i(t) \geq 1$ is a function of time t to be specified below.

Starting with ϕ, it is obviously the case that it is easier for a chief or government to rule over areas with a small spatial resistance. For instance, the Nile River Valley had a small ϕ since information and decisions were very easily spread in such a territory whereas the spatial frictions to public goods provision are much higher in a tropical jungle area or a mountainous terrain as in Papua New Guinea.

Distance from individual i to the chief's location depends on the size of the area, s, and on the location of the chief. In line with Alesina and Spolaore (1997) and the Hotelling result, we will assume throughout that the provider of public goods is always located in the middle of the ethnic group's territory. Initially, with only one ethnic group, that implies that the location of the chief is at $s/2$. Hence, we must have that $d_i \in [0, s/2]$. We assume that people located closer to the chief also are closer relatives to him and that distance thus also, by the logic of kinship nepotism, reflects their sympathies for each other. Due to the uniform distribution of people across this territory, the average member of the ethnic group will initially be located at a distance of $s/4$ from the chief. In order for g_i to be positive at all d_i, we assume that levels are scaled such that $\phi d_i < 1$.[15] The fact that we specify an exact d_i is not meant to imply, however, that people are sedentary. Their 'location' on the line should simply be thought of as the center in their movements as nomadic hunter-gatherers.

Lastly, $m_i(t) \geq 1$ reflects internal cultural and genetic drift that occurs over time between the chief and individual i within the group. In our model, such drift increases the cultural distance for individuals to the core of the ethnic group. This has the general effect of weakening the communication from the core.[16] In line with the arguments above, we further assume that the process of genetic and cultural drift is one and the same.

The basic dynamics of genetic drift has been successfully characterized by genetic research. The general formula for genetic distance between two separate populations is

$$F_{ST} = 1 - \exp\left(-\frac{t}{2N}\right) \tag{5}$$

where $F_{ST} \in (0, 1)$ is genetic distance between the two populations, t is time (in generations) since the two populations became separated, and N is the number of individuals within the new group in reproductive age (Cavalli-Sforza

[15] Inserting the extreme value $d_i = s/2$ further gives us that $s\phi < 2$.

[16] One type of cultural drift that quickly emerges between separated populations is language change, as discussed above.

et al, 1994). F_{ST} is thus a positive, convex function of time. Genetic drift also decreases with the size of the reproducing population N.[17]

It can be shown that for relatively small values of t, the expression for genetic drift and distance will be approximately linear with respect to time (Cavalli-Sforza et al, 1994):

$$F_{ST} \approx \frac{t}{2N} \tag{6}$$

Since we are primarily interested in cultural distances within the group from individual i to the core, N should in our model be thought of as a subpopulation inside group j that practice assortative mating in the form of homogamy. The relevant homogamous population might for instance be the people in a band or village. We assume here that there are no behavioral differences between ethnic groups in their preference for the size of homogamous clusters of people. The constraining factor in mating behavior is rather spatial frictions, i.e. geography forces populations into separate breeding clusters, which we henceforth refer to as *niches*. We assume for simplicity that permanent migrations between niches are not possible, although temporary movements are.[18]

In line with this reasoning, let us assume that individual i is located at some spot $z \in [0, s/2]$, i.e. to the left of the chief, as shown in Figure 3a. The geographical distance to the core is here $d_i = s/2 - z$. At z, i is not a member of the chief's niche. We assume that all niches in the region include a reproductive population of size

$$N = \frac{n\bar{p}}{2\phi} < p_j \tag{7}$$

where n is a constant, ϕ is geographical frictions, p_j is the total level of population in ethnic group j (initially equal to $s\bar{p}$), and $\bar{p}$ is the uniformly distributed population density. The expression in (7) shows the geographical extension of each niche n/ϕ times population density $\bar{p}$ times the fraction of the population in reproductive age, which we assume to be 1/2 for simplicity. The size of the niche is a decreasing function of ϕ, which implies that groups inhabiting inaccessible terrains with high frictions will generally have a smaller pool of potential sexual partners. For simplicity, let us assume that total territory can initially be divided into exactly $q = s\phi/n$ equal-sized niches where q is an odd integer.[19] Figure 3a shows one such configuration with $q = 3$ where i belongs to the niche

[17] This is why a small founding population typically will experience rapid genetic drift, i.e. a 'founder effect'.

[18] We could have used a more relaxed assumption regarding migration, but it would have complicated the model and potentially obscured our main point. ϕ might in general be considered as a parameter indicating the ease of migration within a region. See Michalopoulos (2007) for a model for a model where geographical frictions influence migration patterns.

[19] We make this assumption in order to keep the chief in the middle of his own cluster. With an even number of clusters, the chief would have to be located at a border between clusters to maintain his position at $s/2$.

to the left.

We will assume throughout that $q \geq 3$ so that cultural drift will always be present in the initial situation. Specifically, we propose on the basis of (6) and (7) that cultural drift for any individual i at location z_i within group j is

$$m_i(t) : \begin{cases} = 1 + \frac{\phi t}{n\bar{p}} & \text{if } z_j \notin \left[\frac{s-n/\phi}{2}, \frac{s+n/\phi}{2}\right] \\ = 1 & \text{otherwise} \end{cases} . \quad (8)$$

In other words, individual i will experience cultural drift in relation to the core of the ethnic group if i is not a member of the chief's breeding cluster. [20] The level of drift outside the core niche will be the same regardless of the number of other niches. If i is a member of the chief's cluster, she will experience no drift and $m_i(t) = 1$. Note also that immediately when new groups emerge, we will that have $m(0) = 1$, i.e. new group formation is always associated with cultural consolidation even in the old group.[21]

The structure outlined in (8) implies that the effective level of public goods will be a discontinuous function of geographical distance from the core, as illustrated in Figure 3b. When the boundary of the chief's niche is passed, g_i will make a discontinuous jump downwards and get a more pronounced negative slope. The width of the jump will further increase with time. Cultural drift thus gradually causes a deterioration in the effective supply of public goods which, as we shall see, will make a fission event more and more likely.

The total cost of providing the public good is k. In line with Alesina and Spolaore (1997), we assume that each individual pays the same tribute for the provision of this good $\tau_i = k/p_j$, perhaps in the form of hunted game, artwork, or firewood. As above, p_j is the level of the population in group j. For now, we will simply hold the total level of the population constant. Making the fertility decision endogenous is straightforward but we refrain from that here in order to keep the model simple. The fact that the tribute for each individual decreases with the level of the population, is a source of economies of scale.

[20] In the case of $q = 3$ as in Figure 3a, we will have that $(s - n/\phi)/2 = n/\phi$, but that will not apply when $q > 3$.

[21] We believe that this assumption of cultural homogenization holds for most societies that experience a split. If it was not the case, the dynamics of continued cultural fractionalization would have been very fast, a pattern which we do not seem to observe. Even without the homogenization assumption above, the central prediction regarding the influence of time on cultural distance would be qualitatively the same.

3.3 The hunter-gatherer equilibrium

After having specified all these functional forms, we can reformulate the utility function as

$$U_i = \alpha \ln \left(g \left(1 - \phi d_i m_i\right) e_i^{\gamma} \omega_j^{1-\gamma} \right) - \alpha \ln \left(\frac{k}{p_j} \right) + (1 - \alpha) \ln (1 - e_i) .$$

We are also ready to study optimal individual behavior. The model has two stages, and the individual's first choice is whether to remain within the ethnic group where they currently belong or to form a new group together with their nearest neighbors. In the second stage the individual decides on optimal levels of effort, leisure, and production within the chosen group. The model is solved through backward induction and we therefore start in the second stage.

By taking the usual first-order conditions for maximum, we can solve for the optimal level of effort e^*:

$$\frac{\partial U_i}{\partial e_i} = \frac{\alpha\gamma}{e_i^*} - \frac{(1-\alpha)}{1-e_i^*} = 0$$

After some manipulations, the equilibrium levels of effort and leisure:

$$\begin{aligned} e_i^* &= \frac{\alpha\gamma}{1-\alpha+\alpha\gamma} \\ l_i^* &= \frac{1-\alpha}{1-\alpha+\alpha\gamma} \end{aligned}$$

Note in particular that the optimal level of effort will be independent of the quality of the land and the spatial frictions.

The implied indirect level of utility is thus

$$V_i = \ln \left(\frac{(\alpha\gamma)^{\alpha} (1-\alpha)^{1-\alpha}}{1-\alpha+\alpha\gamma} \right) + \alpha \ln \left(g \left(1 - \phi d_i m_i\right) \omega^{1-\gamma} \right) - \alpha \ln \left(\frac{k}{p_j} \right) . \quad (9)$$

The more complicated and crucial decision is taken in the first stage. Taking the second-stage level of utility into account, each individual considers whether they should remain in the ethnic group or leave the ethnic group and form a new one. We refer to the break-up of one existing ethnic group into more than one group as an *ethnic fission*. The existence of this kind of decision means that the chief is unable to prevent kinsmen from breaking away, i.e. fission decisions can be made unilaterally even though such fissions cause a greater per capita cost of public goods for the kinsmen who stay in the old group. We argue that this regime is the most reasonable one for primitive hunter-gatherer societies but is not well applied to the later agricultural or industrial eras.[22]

In our model, it is intuitively clear that an individual will be more inclined to form a new ethnic group the greater the distance to the chief and the greater the accumulated level of cultural drift within her niche. Obviously, the chief and the people who are close to him in terms of space and kinship will never attempt to form a new group since they have a small distance, belong to the same niche, and pay a smaller tribute in the situation with one group than with two or more groups. Thus, people in the geographical or kinship periphery will typically be the founders of new groups.

Formally, the decision hinges upon the relative indirect utilities from the two choices for the people in the group's periphery. The general mathematical condition for an individual at any location $z < s/2$ to be willing to form a new ethnic group is

$$\begin{aligned} & V_i^{new} - V_i = \\ \alpha \ln\left(1-\frac{\phi z}{2}\right) - \alpha \ln\left(\frac{k}{z\bar{p}}\right) & - \alpha \ln\left(1-\frac{m_i \phi\left(s-2z\right)}{2}\right) + \alpha \ln\left(\frac{k}{s\bar{p}}\right) = \\ & = \alpha \ln\left(\frac{\left(1-\frac{\phi z}{2}\right) z}{\left(1-\frac{m_i\phi(s-2z)}{2}\right) s}\right) > 0 \end{aligned} \tag{10}$$

where V_i^{new} is the indirect utility for i after having joined the new ethnic group and V_i is the utility from status quo.[23] This condition will obviously be satisfied if the expression inside the parenthesis exceeds unity, or analogously, if

$$\begin{aligned} \left(1-\frac{\phi z}{2}\right) z - \left(1-\frac{m_i \phi\left(s-2z\right)}{2}\right) s = \\ z - s + \frac{\phi}{2}\left(m_i s\left(s-2z\right) - z^2\right) = \Omega\left(\phi, m_i\right) > 0 \end{aligned} \tag{11}$$

The lower expression can be broken up in two parts: The first part, the term $z - s$, is clearly negative and shows the increase in costs paid for public goods if i joins the new group. This negative effect of a fission is potentially dominated by the second term which reflects the gains from a shorter effective distance to the core. These gains will grow with time due to internal cultural drift since $m_i(t)$ is a linear function of time.

We will assume that a new group can only be formed if all members of the most peripheral clusters to the left and right of the core group agree to leave

[22] See Alesina and Spolaore (2003) for an extensive analytical treatment of different potential regimes.

[23] In the expression above, it should be remembered that V_i^{new} is independent of the parameters in m since $m\left(0\right) = 1$.

the old group.[24] The key persons will thus be the individuals at $z = n/\phi$ and at $z = (q-1)\,n/\phi$. Since their choice situations are identical, we will only consider the person at $z = n/\phi$. The person at this location will be the one who is worst off if she joins the new group since she will have the greatest distance to the new core. Whereas costs for public goods will obviously be higher in a new group, her potential gain lies in a smaller cultural distance. If we insert $z = n/\phi$ and the cultural drift equation into (11) and set $\Omega = 0$, we can fully solve for the general condition that defines the level of $m_i(t)$ when an ethnic fission will occur:

$$m^* = \frac{\left(2s\phi - 2n + n^2\right)}{s\phi\,(s\phi - 2n)} = \frac{(2q - 2 + n)}{qn\,(q-2)} > 1 \tag{12}$$

The expression to the right-hand side is obtained by substituting in $q = s\phi/n$.

If we combine this critical level of m with the cultural drift equation in (8), we can derive a key results of the model:

Proposition 1 *In the hunter-gatherer era, the number of generations from the foundation of an ethnic group to a fission event is*

$$t^* = \frac{n\bar{p}}{\phi}\left(\frac{2s\phi - 2n + n^2}{s\phi\,(s\phi - 2n)} - 1\right) = \frac{n\bar{p}}{\phi}\left(\frac{2q - 2 + n}{qn\,(q-2)} - 1\right) \tag{13}$$

where t^* increases with $\bar{p}$ and decreases with $q = s\phi/n$ at all $q \geq 3$.

Proof. *The date when a fission occurs is given by $m^* = \frac{\left(2s\phi - 2n + n^2\right)}{s\phi(s\phi - 2n)} = m_i(t) = 1 + \frac{\phi t}{n\bar{p}}$. Solving for t yields the expression in (13). Comparative statics with regard to $\bar{p}$ is straightforward. The result for q is obtained by the derivative $\frac{\partial t^*}{\partial q} = \frac{n\bar{p}}{\phi} \cdot \frac{\partial m^*}{\partial q} = \frac{n\bar{p}}{\phi} \cdot \frac{2(q(2-q)+n(1-q)-2)}{nq^2(q-2)^2} < 0$ at all $q \geq 3$.* ■

A high population density $\bar{p}$ inside each niche means slower genetic and cultural drift and hence a longer time until a fission is viable. Furthermore, the greater the initial number of niches q, the smaller the t^* and the faster the rate of new group formation. Since $q = s\phi/n$, we can infer from the expression above that t^* will also decrease with s and ϕ.

As time passes, there is a sequence of fissions in the old group, each time with two territories the size of n/ϕ breaking away. None of these new ethnic groups experience any further ethnic fissions.[25] Note that each time a fission occurs, the size of the territory that the old group retains and the number of niches that it populates, will both shrink. At the most recent θ:th fission event,

[24] Other rules regarding the minimum size of new ethnic groups would certainly have been possible. However, we find it plausible that a new group will strive to include all members of each involved breeding cluster. The assumption regarding the minimum size are not central to the main results.

[25] The maximum number of ethnic groups on a given territory is obviously $q = s\phi/n$.

a land area of size $s-2(\theta-1)n/\phi$ is split between the new and the old groups, an area which of course is substantially smaller than the original size s.[26] After the θ:th fission, there will be $q-2\theta\geq 1$ niches left in the old ethnic group. From Proposition 1, this in turn means that t_{θ}^{*} will increase with each new fission and hence that the rate of ethnic fissions will fall over time:

$$t_{\theta}^{*}>t_{\theta-1}^{*}>...>t_{1}^{*}\quad \text{where } \theta=1,2,3,...(q-1)/2.$$

The rate of ethnic fissions will thus be fastest in the early days of settlement and then gradually decline, implying longer and longer periods of ethnic stability.

Lastly, and most importantly, the expressions above provide a very clear hypothesis regarding the impact of time on the level of ethnic fractionalization. Let $T_r>0$ be the time interval that a certain region r has been settled by AMH. The most recent fission event in r, referred to as the θ_r:th, is more formally defined as

$$\theta_r=\arg\max_{\theta}\left(\sum_{v=1}^{\theta}t_v^{*}-T_r\leq 0\right) \tag{14}$$

The important insight from (14) is simply that there is a positive, non-linear relationship between θ_r and T_r. The longer the time span that a region has been settled, the greater the number of fissions that have occurred, holding everything else constant. Since θ_r fissions implies the existence of $2\theta_r+1$ ethnic groups, we can express the key proposition:

Proposition 2 *At any given time during the hunter-gatherer era, ethnic diversity within a region increases with the time elapsed since initial human settlement.*

Proof. *See text.* ■

Figure 4 shows an example of two otherwise identical regions, A and B, with the only distinguishing feature that A has been settled twice as long; $T_A=2T_B$. Since the process of ethnic fissions starts much later in region B, ethnic diversity at time T_A will be smaller in region B than in A (3 fissions (implying 7 groups) in B and 5 fissions (implying 11 groups) in A). This is the key hypothesis that we bring to the empirical section.

3.4 Ethnic fractionalization under statehood

Since the rise of civilization around 4000 BC, many ethnic groups have lived in agricultural states characterized by centralized government and more or less

[26] The fission outlined above happens at $\theta=1$ so that $s-(1-1)n/\phi=s$ is the relevant territory.

distinct notions of geographical and cultural boundaries. Although there are many differences between primitive and civilized societies, we argue that there are three primary factors that affect the process of ethnic fractionalization in our model: Firstly, agricultural, sedentary societies have a much higher population density than hunter-gatherer societies. Secondly, governments in agricultural states have a greater capacity and often pursue efforts aimed at cultural homogenization, for instance manifested in the construction of a national cultural identity.[27] Thirdly, a unilateral fission by a population cluster within a state can only be achieved at a substantial fixed cost.[28]

In our model, we capture these features by assuming that governments in every time period t made investments h aimed at constructing a national identity that reduce the distance from i to the core and counterbalance the process of internal cultural drift m.[29] More specifically, we propose that the level of the public good and the distance function for individual i at z now is given by $g_i = g\left(1 - \frac{\phi(s-2z)}{2} \cdot \frac{m_i(t)}{(h_0+ht)}\right)$ where h_0 is the initial level of distance-reducing public investments, and h is the (exogenous) size of the investment in each period.

As before, $m_i(t) = 1 + \frac{\phi t}{n\bar{p}_s}$ captures cultural and genetic drift within homogamous subpopulations that are not part of the core cluster. The core cluster now centres around the king or the ruling elite in the capital city. The uniform level of population density in agriculture $\bar{p}_s$ is substantially higher than in hunter-gatherer times. This in turn implies that cultural drift within clusters is much less severe under agriculture, which makes $m_i(t)$ lower. If homogenization investments h are larger than $\phi/n\bar{p}_s$, it will even be the case that the term $m_i(t)/(h_0 + ht)$ shrinks with time so that the cultural distance falls.

We assume that the public investments at each point in time h also contribute to increasing the fixed costs of a fission so that $f(h)$ where $f'(h) > 0$. The costs of a secession usually include the risk of military retribution by the government or a refusal to recognize the new state.

Taken together, these features of the model enables us to make the following

[27] Ashraf and Galor (2007) present a model where cultural assimilation during the agricultural era is more likely to occur in areas that are less 'geographically vulnerable' to cultural diffusion from nearby areas.

[28] The issue of how agriculture and states arose from more primitive production is beyond the scope of this paper. See for instance Smith (1998), Olsson (2001), or Olsson and Hibbs (2005) for literature overviews and theoretical models of these transitions.

[29] A government might alternatively pursue a kind of transfer system of public funds to the periphery, as analyzed in Alesina and Spolaore (2003).

indirect utility comparison:

$$V_i^{new'} - V_i^{state} = \tag{15}$$

$$+\alpha \ln\left(1 - \frac{\phi z}{2}\right) - \alpha \ln\left(\frac{k}{z\bar{p}}\right) -$$

$$-\alpha \ln\left(1 - \frac{\phi(s-2z)}{2} \cdot \frac{m_i(t)}{(h_0 + ht)}\right) + \alpha \ln\left(\frac{k}{s\bar{p}}\right) - \ln f = \tag{16}$$

$$= \alpha \ln\left(\frac{\left(1 - \frac{\phi z}{2}\right) z}{\left(1 - \frac{\phi(s-2z)}{2}\frac{m_i}{(h_0+ht)}\right) s} \cdot \frac{1}{f(h)}\right).$$

Obviously, the break-up of an existing ethnic community (which is also a state) will prove less beneficial if h and $\bar{p}_s$ are large and if the fixed costs of a secession $f(h)$ are high. Indeed, what actually seemed to happen in the era of nation-states was rather an amalgamation of ethnicities and a reduction of ethnic diversity. If $h > \phi/n\bar{p}_s$ cultural distance within the state would eventually go to zero so that $m_i = 1$ at all z. In this case, existing small ethnic groups on the fringe of the agricultural state with an indirect utility equivalent of $V_i^{new''}$ would typically find that $V_i^{new''} - V_i^{state} < 0$ so that it would actually be rational for them to join the state and gradually give up their own ethnic identity.

Proposition 3 *If $h \geq \phi/n\bar{p}_s$ so that cultural distances among the subpopulations within the state are non-increasing with time, ethnic fissions will cease to occur and some existing ethnic groups will decide to join the state. Ethnic diversity should thus decrease with the duration of state experience.*

Proof. *Intuitive.* ■

In summary, the results in the two last sections thus lead to the prediction that ethnic diversity should increase with the duration of human settlements and with the strength of geographical frictions, and decrease with the duration and intensity of statehood experience. In the next section, we bring these predictions to the data.

4 Empirical analysis

4.1 Coding initial human settlement

Our empirical analysis introduces the historical duration of human settlements, *Origtime*, and since we have coded this variable ourselves and since our major results concern this variable, we will briefly discuss how it has been constructed.

What we have tried to establish is the date of the first uninterrupted settlement by anatomically modern humans 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).[30] Oppenheimer (2003) provides a synthesis of genetic, archeological, climatological, and fossil 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 - 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.[31] 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 in 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 south-

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

[31] 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).

ern '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 (Macauley et al., 2005), as well as very early archeological findings of AMH in Australia, appear to support a beachcombing route.[32]

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, modern humans 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 9 contains the estimated *Origtime* for the 191 countries in our sample.

4.2 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 since the focus in this analysis is not on the effects of ethnic diversity but on where it comes this issue is of minor impor-

[32] 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).

tance. In the years following Easterly and Levine (1997) researchers generally used ethnolinguistic fractionalization (*Elf*), constructed using data collected by Soviet ethnographers in the 1960s. 'Fractionalization' refers to the probability that two randomly selected individuals from a population come from different groups, and the larger the number of groups above the chosen threshold value for inclusion, the higher the fractionalization. More recent indices of ethnic diversity include the fractionalization-indices created by Fearon (2003) and Alesina et al. (2003), and in the analysis to follow we use the latter measure.[33] A full list of the variables included in this section as well as sources and detailed descriptions are presented in table 8.

4.3 Regression results

In line with the predictions from our model we find in Table 2 that ethnic fractionalization is higher in countries with a longer duration of human settlement (*Origtime*). An earlier transition from hunter-gatherer to agricultural production (*Agriculture*) as well as the ultimate success of this transition (measured as population density in year 1 AD) (*Civilization*), a longer history of a strong state apparatus (*State Antiquity*), a stronger state in the era of the modern nation-state (*Nation*), and a geographical location further from the equator (*Latitude*) are factors that are significantly negatively correlated with ethnic diversity.[34] Though illuminating, bivariate correlations hold limited persuasive power and so the further investigation of the empirical strength of our theoretical and historical discussion is performed in a linear regressions framework.

<Table 1 about here>

<Table 2 about here >

In column 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, we believe this is a remarkable result. The size of the coefficient implies that 10000 years earlier human settlement is associated with on average 0.028 points higher ethnic fractionalization. The influence of the introduction of sedentary agriculture and the rise of civilization can be captured in two ways; by the timing of the transition from hunter-gatherer to agricultural production (*Agriculture*) from Putterman (2007a) and by the effect it had on productivity, the ultimate determinant of both population density and social stratification.

[33] These two are available for larger set of countries than *Elf* and our results are 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.

[34] For ease of exposition we have in tables 2-7 scaled *Origtime* to be in 100.000 years units, *Agriculture* to be in 10.000 years units, *Civilization* to be in 100 times the population density in year 1, *Independence* to be in 100 years units, and *Latitude* to be in 10 degrees units.

This effect is captured by population density in year 1 (*Civilization*). Since there is no *a priori* reason to expect that early transition must equal successful transition, a measure for the latter effect should better capture the impact on ethnic diversity. This is indeed what happens - when we include *Agriculture* and *Civilization* simultaneously *Agriculture* shrinks and becomes nonsignificant.

<**Table 3 about here** >

The measure for historical state capacity developed by Bockstette et al. (2002) captures the extent to which states have controlled their present territory, and we include this measure (*State Antiquity*) in column 5.[35] The effect is the expected - having had more control of the present territory for a considerable time is associated with less diversity. Yet, the deliberate homogenizing efforts of the kind that Gellner (1983) and Tilly (1992) discuss (see section 2.2 for details) requires means and motivation only available in the last centuries. In order to investigate whether the result reflects millennia-long unintended effects of state activities or the more recent phenomena of the rise of the modern nation-state, we use the underlying data used to create *State Antiquity* to create two new variables, one representing statehood before the modern era, i.e. between year 1 and year 1800, (*Premodern*), and one representing statehood in the modern nation-state era (*Modern*). A variable closely related to *Modern* is for how long the state has been sovereign and independent (*Independence*). Columns 6-8 show how the reduction in ethnic diversity associated with *State Antiquity* took place in the modern era rather than in the nearly two millennia preceding it, as suggested by the facts that *Premodern* is nonsignificant in column 6 and that the explanatory power of the model is not affected when we literally exclude the information on state capacity between year 1 and year 1800. Longer time as an independent country is also associated with lower ethnic diversity. *Modern*, like *State Antiquity*, assigns lower values to countries when under colonial rule, but since this is only part of the information used *Modern* and *Independence* do not add the exact same information. In the remaining regressions we use *Nation*, created as the first principal component of *Modern* and *Independence,* to pick up the most variation possible associated with being an independent state free to set its policies, as well as having the capacity to pursue them.

In columns 10-12, we include dummies for Europe, America, and exclude the countries in Sub-Saharan Africa and the results on *Origtime*, *Civilization*, and *Nation* remain largely unaffected, and that is also the result when we include a Sub-Saharan Africa dummy in column 11, as well as when our preferred geographical variable, *Latitude*, is added. The specification with region dummies are included for completeness only. If the ambition is to explain a causal relationship, including variables like region dummies makes little sense, and if

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

one wishes to test the strength of *Origtime* it is a particularly ill-suited specification. The reason for this is that two factors make our measure *Origtime* discontinuously distributed over the continents; the fact that AMH did not settle the Earth in a smooth and continuous fashion, and the problem that there is very limited data on the exact dates for each country, especially within Africa. Together these imply that a test of the strength of *Origtime* should use the full variation in *Origtime* (omitting a region is an unwise strategy) and not use other variables that too closely track the variation in *Origtime* (including region dummies may effectively cripple the model).

Despite these potential problems the coefficient for *Origtime* remain significant in all regressions. The variables *Civilization* and *Nation* do not share this problem and remain significant at the 1% level. As is shown in the rest of the section the results presented in table 3, that lower levels of ethnic diversity is found where humans have lived longer, where sedentism and civilization proved most successful, and where the state was more potent in the era of modern nation-states, are very robust to alternative specifications

Any serious investigation of the international variation in ethnic diversity must address the effects of colonialism, not least since it is around these that much of the constructivist debate has taken place. This debate is fuelled by the casual observation that former European colonies tend to have higher ethnic fractionalization than other countries. Among the 143 countries included in columns 11 of table 3 the 84 countries that are coded as colonies in Olsson (2007) have an average ethnic fractionalization 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. The apparent relationship between colonial status and ethnic diversity is thus entirely driven by the difference in ethnic diversity between Europe and sub-Saharan Africa.

<**Table 4 about here** >

As we see in columns 1 and 2 of table 4, the dummy for being a former European colony (*Colony*) enters positively and significantly only when *Latitude* is not controlled for. Since we, on the basis of solid prior research, argue that there indeed is a geographically determined component of ethnic diversity, column 2 is a more correct specification than column 1.[36] That the dummy for being a former colony does not enter as significant does not prove that colonialism has not affected the ethnic diversity in the countries that suffered under it, but it does suggest that there is neither a positive nor a negative effect on average.

[36]Noting that the density of colonization follows a latitudinal gradient an alternative hypothesis, which we do not investigate further in this article, is that latitude is a determinant both for being colonized and for becoming ethnically fractionalized.

Therefore, it becomes even more pressing to investigate differences within the sample of former colonies.

The result from this investigation is presented in columns 4-12 of table 4. The effects of *Origtime*, *Civilization*, and *Nation* are present and strong both among countries that were never colonized and among former colonies, and in the latter group the time spent under colonial rule (*Duration*) is positive and so is the effect of being among the first countries, before 1608, to be colonized. It seems reasonable to assume that both these variables pick up the same variation. A notable difference between the two groups is that former colonies have had less control of their territory and hence get lower values on *Nation*. Our results show that it is this fact, rather than the colonial status itself, that is related to the level of ethnic diversity. Excluding the early colonies from the sample (not reported) or including both variables simultaneously result in *Duration* becoming nonsignificant.[37] Other than the effect of very early colonization, the results in columns 8 and 9 suggest that the effect of colonization is not determined by what year the colonization process started.[38] Columns 10-12 show that the identity of the colonizer may have some importance also when we have controlled for geography and length of colonization. Controlling if there is an effect of the identity of the colonizing power, we include dummies for being subject to colonization by Great Britain, France, or Spain, leaving former colonies of the Netherlands, Portugal and Belgium as the control group. The French colonial dummy is significant, but the magnitude of the coefficient seem unreasonably high. In many cases it is acceptable to treat colonial status as exogenous but dealing with time spans like the ones we have here, tens or even hundreds of thousands of years, historical variables become inseparable from geographical, as shown in figure 2. 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.[39]

That geographical factors influence the level of ethnic diversity was predicted by our model. While we have no direct empirical equivalent of ϕ, the geographical friction of a territory, we proxy for its effects by diversity of vegetation and soil types as well as within-country variation in temperature, precipitation, and altitude. The results are presented in tables 5 and 6. What results can we expect the inclusion of these geographical variables to have on the estimated co-

[37]Finding the exact mechanism behind these two effects is surely an interesting task, but not the subject of this paper.

[38]The thresholds 1608, 1820, and 1886 corresponds to the 25th percentile, median, and 75th percentile of years since start of colonization respectively.

[39]A hypothetical mechanism that suggests that the causality may run from fractionalization to colonization is that an ethnically fractionalized society may face collective action problems in the defence against colonizing powers.

efficient for *Origtime*? As should be evident from the discussion on the creation of *Origtime,* the first waves of human settlements were largely directed by geographical factors such as climate and vegetation, and areas closer to the equator were generally populated first. That the effect on ethnic diversity from *Latitude* should be negative is evident from the detailed discussed in section 2.3. When *Latitude* is included as regressor the coefficient for *Origtime* should therefore shrink. Other variables measuring the extent to which the geographic characteristics of a country promote isolation should enter with positive coefficients while simultaneously reducing the coefficient for *Origtime*.

As expected, the general ecological pattern of higher species richness closer to the equator is found also for human diversity, and the magnitude of the effect of *Origtime* is reduced by the inclusion of *Latitude* in the regressions. That ethnic diversity follows a latitudinal gradient is, as is shown by the stability of the coefficient to the inclusion of other geographical controls in tables 5 and 6, a robust finding. The first specifications in table 5 add controls for size of physical territory and population. There is no effect of neither a larger territory nor latitudinal stretch, indicating that if within-country physical distances affect group formation they do not do so to the extent that it becomes visible in our data. If population density in the modern nation-state era affects the level of diversity we should find an effect from the population density in 1900, yet we do not. Since the present population density cannot reasonably affect diversity to any measurable extent, the fact that population density in 2005 is significantly negatively associated with present ethnic diversity indicates either reversed causality or a mere spurious relationship.

< **Table 5 about here**>

Next we turn to factors more directly associated with habitat diversity as proxies for ϕ. Within-country differences 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.[40] The variety of vegetation and soil types (see table 5) and within-country differences in temperature, altitude, and precipitation (table 6) are all positively associated with diversity.

We argued above that there are two effects associated with a larger part of a country's area being close to a sea coast or a river: A positive effect on diversity from higher values on *Origtime* following the beachcombing hypothesis, and a negative effect on diversity from population in these areas being less isolated. We

[40] 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 gives stronger support.

predicted that the latter (negative) effect should be likely to dominate. Testing this argument requires that we include both the share of area and the share of population in coastal/riverine areas. Since the share of land and the share of population are highly correlated (0.92) including only, say, 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.[41] When we thus include both these measures, in column 4 of table 6, a larger fraction of the population living in less isolated areas is associated with lower levels of ethnic diversity. Since we control for *Latitude* as well as differences in temperature this does not simply capture the fact that populations in many tropical countries choose to live in the cooler inland areas. Evidently, the result that geographical factors enabling isolation affect ethnic diversity is not sensitive to which variable we include. To capture the essential variation while controlling for region dummies we create *Geography-induced isolation* as the principal component of diversity of vegetation types, differences in temperature and altitude, and latitude, and add this to columns 5 and 6. The overall results are neither sensitive to the inclusion of region dummies, nor to running the regression with Sub-Saharan Africa excluded.

< **Table 6 about here**>

4.4 Robustness

In table 7 we investigate whether our results are sensitive to the inclusion of particularly influential or unusual observations, or to our choice of dependent variables. We start by running *Origtime* against *Fission*, a measure we construct from data provided in Cavalli-Sforza et al (1994). Using statistical clustering analysis, Cavalli-Sforza et al (1994) are able to divide the world population into 6 main population groups and provide their approximate genetic distance. Using this data and the formula in equation (5), we can calibrate the approximate time since each group split from the others, *Fission* (see table 8 for details). As an example, we find that the now non-African populations split from the African population about 86,000 years ago, closely matching the dating of the African exodus provided by Oppenheimer (2003). *Fission* is included in column 2 of table 7. The result shows that *Origtime* does not merely reflect the variation in time since separation *between* these 6 world populations but also the variation in the duration of settlements *within* these populations.[42]

[41] Including present day population does cause causality problems We justify our approach with the fact that only quite recent population figures with the proper disaggregation are available are available and encourage the reader to interpret with care.

[42] If *Fission* is used instead of *Origtime* as our proxy for settlement duration, it is strongly significant in the predicted direction. (Results available on request).

In columns 3 and 4 we drop observations identified with high DFBETA or leverage.[43] In columns 5 and 6 we employ two methods designed to create as normally distributed variables as possible. Both these methods show that our results are not driven by non-normally distributed variables.[44] Lastly, in column 7 we replace ethnic fractionalization from Alesina et al. (2003) with ethnic fractionalization from Fearon (2003). As indicated by the significance of the variables and the quite stable beta coefficients, which are reported in parentheses, none of the tests or alternative specifications in table 7 change the results previously obtained.

< **Table 7 about here**>

5 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

[43] We obtain very similar results when we use alternative measures and rules for omission of influential and unusual observations, such as DFITS, or Cook's or Welsch distance.

[44] In column 5 we analyse the distribution of our variables without transformation (X) as well as with cubic (X^3), square (X^2), square-root $\left(\sqrt{X}\right)$, log $(\ln X)$, reciprocal root $\left(\frac{1}{\sqrt{X}}\right)$, reciprocal $\left(\frac{1}{X}\right)$, reciprocal square $\left(\frac{1}{X^2}\right)$, and reciprocal cubic $\left(\frac{1}{X^3}\right)$. We test these for skewness and kurtosis and use the one with best test scores. The specification in column 5 is therefore on; ethnic = sqrt(ethnic), origtime = log(origtime), civilization = sqrt(civilization), nation = nation, latitude = sqrt(latitude).

In column 6 we use the zero skewness Box-Cox power transformation to create variables with zero skewness. We report the results when this procedure is applied only to *Origtime*. The procedure requires positive values wherefore it cannot be applied to *Nation*, but we have checked the effects of using this procedure on the other variables and including these and a non-tranformed *Nation* in the regression, as well as replacing *Nation* with its components *Modern* or *Independence,* after subjecting them to the procedure, and the results are very similar.

countries where early civilization proved more successful and where the state was stronger during the modern nation-state era.

Our result shows that a serious understanding of ethnic diversity requires a synthesis of primordial and constructivist arguments. Supporting the primordialists is the finding of long–run factors underlying the differences in ethnic diversity, but that a stronger state apparatus in the modern nation-state era is associated with increased ethnic homogeneity supports the constructivist side. The analysis shows that social indicators such as the rise of civilization and the modern state 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 nation-state. But the analysis also shows that 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).

A reasonable projection of the results implies that the effect of *Origtime* eventually will vanish in all regions and in all countries. However, we have deliberately used the terms *modern* and *nation-state* since the evidence we present here potentially have a weaker bearing on the effects of how the developed (or post-modern) state works in a global economy. 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 folklore will be forgotten, and with them the aggregated experience of a large part of humanity.

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. An often employed method for assessing the effect of ethnic diversity on economic and political performance has been to include a measure of ethnic fractionalization as one of many potential regressors. Since a stronger state in the era of the nation-state is associated with a lower degree of ethnic diversity, and since there is a positive correlation between indicators of this strength and many indicators of economic and political performance, the negative coefficient on ethnicity obtained in these regressions could reflect an omitted variable bias - they may be but statistical artefacts created by the omission of long-term state strength from the regression.

References

[1] Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat and R. Wacziarg (2003) "Fractionalization", *Journal of Economic Growth*, 8:155-94.

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

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

[4] Alesina, A. and E. Spolaore (2003) *The Size of Nations*, Boston: MIT Press.

[5] Alesina, A. and R. Wacziarg (1998) "Openness, Country Size, and Government", *Journal of Public Economics*, 69:305-321.

[6] Anderson, B. (1983) *Imagined Communities*, London: Verso.

[7] Arcand, J-L., P. Guillaumont, and S. Guillaumont Jeanneney (2000) "How to make a tragedy: on the alleged effects of ethnicity on growth", *Journal of International Development*, 12:925-938.

[8] Ashraf, Q. and O. Galor (2007) "Cultural Assimilation, Cultural Diffusion, and the Origin of the Wealth of Nations", *mimeo*, Brown University.

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

[10] van den Berghe, P.L. (1981) *The Ethnic Phenomenon*, Westport: Praeger.

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

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

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

[14] Caselli, F. and W.J. Coleman II. (2006) "On the Theory of Ethnic Conflict" London School of Economics, mimeo.

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

[16] Cavalli-Sforza, L., A. Piazza, P. Menozzi, and J. Mountain (1988) "Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data" *Proceedings of the National Academy of Science*, 85:6002-6006.

[17] Cavalli-Sforza, L., P. Menozzi, and A. Piazza (1994) *The History and Geography of Human Genes*, Princeton University Press.

[18] Cavalli-Sforza, L., and F. Cavalli-Sforza (1995) *The Great Human Diasporas: The History of Diversity and Evolution*, Cambridge, MA: Perseus Books.

[19] Chanda, A. and L. Putterman (2006) "Early Starts, Reversals and Catchup in The Process of Economic Development", mimeo.

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

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

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

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

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

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

[26] Collier, P. and A. Hoeffler (2004) "Greed and Grievance in civil war", *Oxford Economic Papers*, 56:563-95.

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

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

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

[30] Easterly, W. (2001) ?

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

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

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

[34] Gallup, J.L., J.D. Sachs, and A.D. Mellinger (1998) "Geography and Economic Development", NBER Working Paper No. 6849.

[35] Galor, O., and O. Moav (2002) "The Origins of Economic Growth" *Quarterly Journal of Economics* 117, 1133-1191.

[36] Gellner, E. (1983) *Nations and Nationalism.* Oxford: Blackwell

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

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

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

[40] Hobsbawm, E.J. and T. Ranger (1983) *The Invention of Tradition*, Cambridge University Press.

[41] 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.

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

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

[44] Leeson, P.T. (2005) "Endogenizing fractionalization", *Journal of Institutional Economics* 1(1):75–98.

[45] Liu, H., F. Prugnolle, A. Manica and F. Balloux (2006) "A Geographically Explicit Genetic Model of Worldwide Human-Settlement History", *American Journal of Human Genetics*, 79:230-237.

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

[47] Mace, R. and M. 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.

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

[49] Michalopoulos, S. (2007) "The Origins of Ethnolinguistic Diversity: Theory and Evidence", mimeo, Brown University.

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

[51] Olsson, O. (2001) "The Rise of Neolithic Agriculture", mimeo, Göteborg University.

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

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

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

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

[56] Pagel, M., and R. Mace (2004) "The cultural wealth of nations", *Nature*, 428(18 March):275-8.

[57] Putterman, L. (2007a) `http://www.econ.brown.edu/fac/Louis_Putterman`.

[58] Putterman, L. (2007b) `http://www.econ.brown.edu/fac/Louis_Putterman`.

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

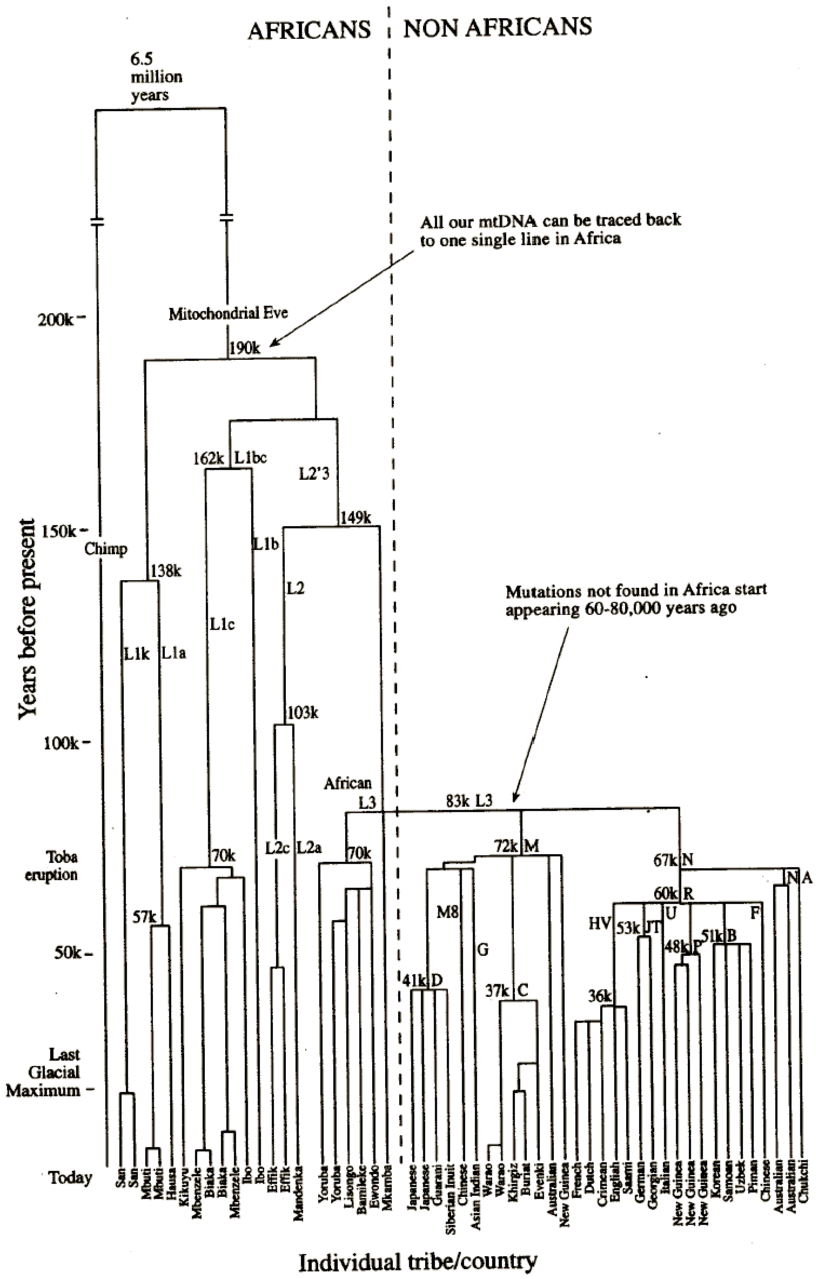
[60] Smith, B.D. (1998) *The Emergence of Agriculture*, New York: ScientificAmerican Library.

[61] Smith, A.D. (1986) *The Ethnic Origin of Nations*, Oxford: Basil Blackwell.

[62] Tilly, C. (1992) *Coercion, Capital, and European States*, AD 990-1992, Oxford: Blackwell.

[63] 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: The human genealogy



Source: Oppenheimer (2003), based on Ingman et al (2000).
Note: The figure shows the genetic distance (using mitochondrial DNA) between ethnic groups across the world, based on a sample of 53 individuals. Information such as "138k" refers to the approximate date when genetic branches split up, whereas for instance "L1k" is the name of a particular genetic branch.

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

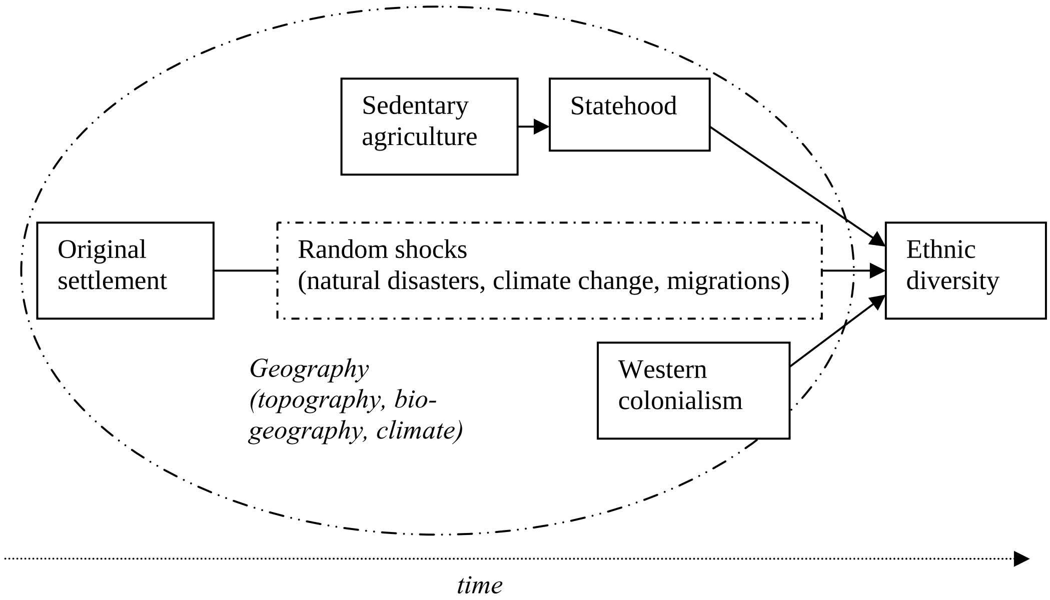


Figure 3a: Geographical breeding clusters within ethnic group j (q=3).

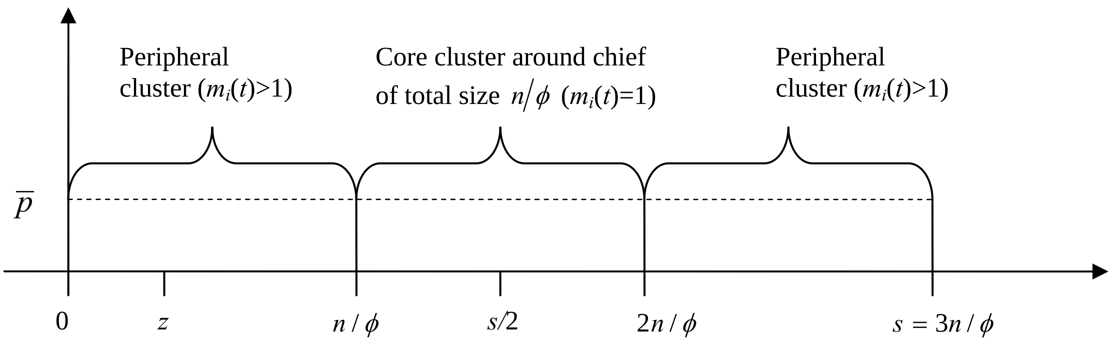


Figure 3b: Effective level of public goods supply at different locations (q=3).

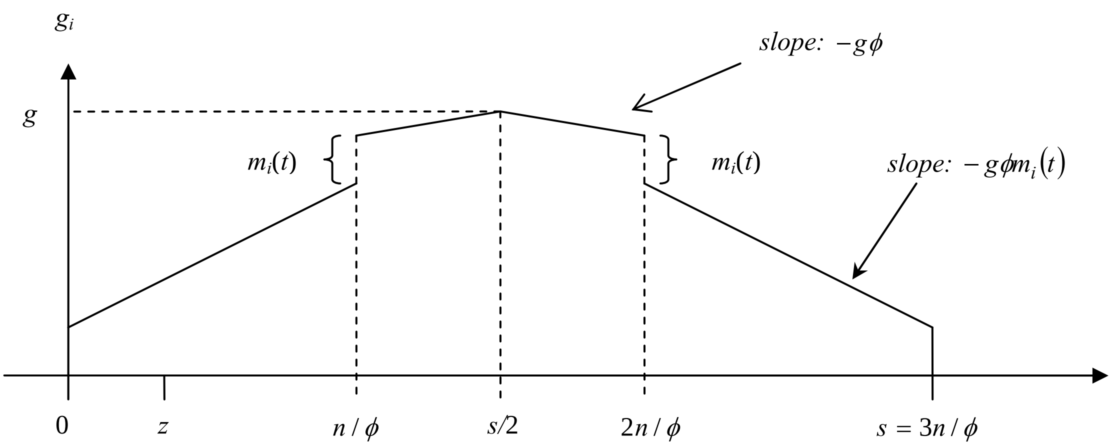


Figure 4: The evolution of ethnic diversity in two regions with different settlement times (T_A= $2T_B$).

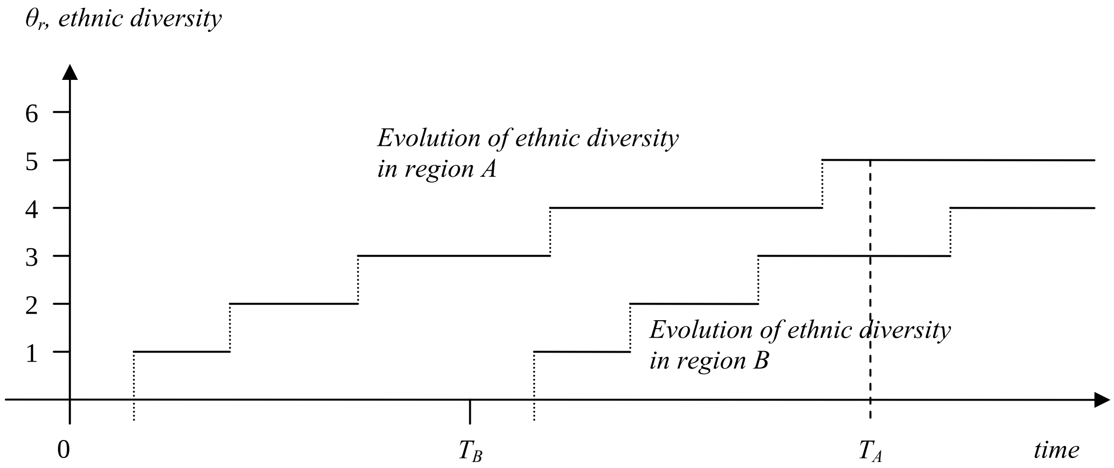


Table 1. Descriptive statistics

Variable	Observations	Mean	Median	St.deviation	Min value	Max value
Ethnic fractionalization	143	0.45	0.46	0.25	0.00	0.93
Origtime	143	56573	40000	49455	200	160000
Civilization	143	3.88	0.84	8.45	0.00	71.93
State Antiquity	143	0.44	0.41	0.24	0.02	0.96
Nation	143	0.01	-0.29	1.01	-1.71	3.53
Latitude	143	27.56	25.97	17.62	0.23	64.15

Table 2. Pair-wise correlations

	Ethnic	Origtime	Civilization	Agriculture	State Antiquity	Nation
Origtime	0.5267					
	0.0000					
	184					
Civilization	-0.3250	-0.1987				
	0.0000	0.0065				
	180	186				
Agriculture	-0.2725	-0.2869	0.1623			
	0.0006	0.0003	0.0436			
	156	157	155			
State Antiquity	-0.2655	-0.1518	0.1121	0.6346		
	0.0013	0.0673	0.1794	0.0000		
	144	146	145	138		
Nation	-0.3681	-0.3011	0.0254	0.1665	0.5053	
	0.0000	0.0002	0.7619	0.0510	0.0000	
	144	147	145	138	146	
Latitude	-0.4231	-0.3810	0.0797	0.4766	0.4085	0.3269
	0.0000	0.0000	0.2797	0.0000	0.0000	0.0001
	182	188	186	146	146	146

Notes: Correlation, p-value, and number of observations are lister for each pair.

Table 3. Determinants of ethnic diversity

	Full sample (1)	Full sample (2)	Full sample (3)	Full sample (4)	Full sample (5)	Full sample (6)	Full sample (7)	Full sample (8)	Full sample (9)	SSA excluded (10)	Full sample (11)	Full sample (12)
Origtime	0.277***	0.246***	0.224***	0.251***	0.235***	0.215***	0.214***	0.241***	0.209***	0.237**	0.187**	0.148*
	(0.031)	(0.036)	(0.036)	(0.032)	(0.034)	(0.036)	(0.036)	(0.032)	(0.035)	(0.112)	(0.076)	(0.085)
Agriculture		-0.138*	-0.108									
		(0.079)	(0.075)									
Civilization			-0.522***	-0.543***	-0.723***	-0.758***	-0.792***	-0.514***	-0.783***	-0.623***	-0.716***	-0.769***
			(0.145)	(0.148)	(0.209)	(0.209)	(0.231)	(0.134)	(0.225)	(0.122)	(0.140)	(0.153)
State Antiquity					-0.166**							
					(0.075)							
Premodern						-0.084						
						(0.071)						
Modern						-0.186**	-0.223***					
						(0.091)	(0.084)					
Independence								-1.879**				
								(0.797)				
Nation									-0.588***	-0.534***	-0.560***	-0.497***
									(0.152)	(0.169)	(0.163)	(0.166)
America										0.151**	0.135**	0.074
										(0.061)	(0.055)	(0.069)
Europe										-0.028	-0.035	0.022
										(0.051)	(0.049)	(0.056)
Sub-Saharan Africa											0.082	0.045
											(0.079)	(0.081)
Latitude												-0.036*
												(0.019)
Constant	0.294***	0.390***	0.409***	0.328***	0.415***	0.514***	0.507***	0.356***	0.359***	0.302***	0.329***	0.461***
	(0.025)	(0.054)	(0.052)	(0.027)	(0.042)	(0.066)	(0.067)	(0.030)	(0.029)	(0.060)	(0.049)	(0.090)
Observations	184	156	154	180	143	143	143	180	143	103	143	143
Adjusted R-squared	0.27	0.27	0.31	0.32	0.35	0.36	0.36	0.34	0.37	0.21	0.42	0.44

Notes: Dependent variable in all regressions: ethnic fractionalization from Alesina et al. (2003). Coefficients are reported with robust standard errors in parentheses. ***, ** , and * indicate significance at the 1, 5, and 10 percent levels. For ease of exposition we have in tables 2-7 scaled Origtime to be in 100.000 years units, Agriculture to be in 10.000 years units, Civilization to be in 100 times the population density in year 1, Independence to be in 100 years units, and Latitude to be in 10 degrees units.

Table 4. Ethnic diversity and colonial history

	Full sample (1)	Full sample (2)	Non-colonies (3)	Former colonies (4)	Former colonies (5)	Former colonies (6)	Former colonies (7)	Former colonies (8)	Former colonies (9)	Former colonies (10)	Former colonies (11)	Former colonies (12)
Origtime	0.186***	0.129***	0.328***	0.156***	0.194***	0.191***	0.197***	0.189***	0.156***	0.164***	0.130**	0.169**
	(0.036)	(0.040)	(0.081)	(0.045)	(0.052)	(0.051)	(0.052)	(0.052)	(0.051)	(0.051)	(0.056)	(0.064)
Civilization	-0.758***	-0.806***	-1.083**	-0.769***	-0.783***	-0.681***	-0.726***	-0.813***	-0.769***	-0.678***	-0.703***	-0.709***
	(0.192)	(0.176)	(0.493)	(0.200)	(0.189)	(0.208)	(0.216)	(0.211)	(0.201)	(0.195)	(0.173)	(0.161)
Nation	-0.498***	-0.461***	-0.466***	-0.690*	-0.664*	-0.877**	-0.784*	-0.786**	-0.690*	-1.043*	-0.609	-0.544
	(0.156)	(0.156)	(0.167)	(0.397)	(0.385)	(0.395)	(0.453)	(0.386)	(0.400)	(0.543)	(0.525)	(0.506)
Colony	0.089**	-0.057										
	(0.036)	(0.061)										
Latitude		-0.059***									-0.061**	-0.055**
		(0.019)									(0.025)	(0.025)
Duration					0.033*		0.018					0.035*
					(0.017)		(0.027)					(0.019)
Colonized before 1608						0.100*	0.058					
						(0.056)	(0.092)					
Colonized before 1820								0.077				
								(0.054)				
Colonized before 1886									-0.001			
									(0.063)			
British colony										-0.020	0.030	0.050
										(0.061)	(0.070)	(0.069)
French colony										0.058	0.089	0.129*
										(0.059)	(0.072)	(0.075)
Spanish colony										0.090	0.068	0.070
										(0.086)	(0.084)	(0.082)
Constant	0.319***	0.601***	0.281***	0.425***	0.336***	0.360***	0.339***	0.359***	0.426***	0.378***	0.484***	0.358***
	(0.031)	(0.096)	(0.044)	(0.037)	(0.068)	(0.059)	(0.067)	(0.065)	(0.076)	(0.068)	(0.081)	(0.124)
Observations	143	143	59	84	84	84	84	84	84	84	84	84
Adjusted R-squared	0.40	0.45	0.22	0.34	0.35	0.35	0.35	0.35	0.33	0.34	0.39	0.40

Notes: Dependent variable in all regressions: ethnic fractionalization from Alesina et al. (2003). Coefficients are reported with robust standard errors in parentheses. ***, ** , and * indicate significance at the 1, 5, and 10 percent levels. For ease of exposition we have in tables 2-7 scaled Origtime to be in 100.000 years units, Agriculture to be in 10.000 years units, Civilization to be in 100 times the population density in year 1, Independence to be in 100 years units, and Latitude to be in 10 degrees units.

Table 5. Geographical determinants of ethnic diversity

	Full sample (1)	Full sample (2)	Full sample (3)	Full sample (4)	Full sample (5)	Full sample (6)	Full sample (7)
Origtime	0.135***	0.132***	0.129***	0.134***	0.133***	0.124***	0.114***
	(0.039)	(0.039)	(0.039)	(0.039)	(0.038)	(0.038)	(0.038)
Civilization	-0.788***	-0.763***	-0.568**	-0.692***	-0.737***	-0.648***	-0.594***
	(0.171)	(0.162)	(0.251)	(0.145)	(0.162)	(0.188)	(0.215)
Nation	-0.444***	-0.497***	-0.430***	-0.437***	-0.512***	-0.506***	-0.532***
	(0.154)	(0.162)	(0.154)	(0.155)	(0.167)	(0.165)	(0.163)
Latitude	-0.046***	-0.047***	-0.046***	-0.049***	-0.046***	-0.050***	-0.055***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.010)	(0.011)	(0.011)
Area		0.000					
		(0.000)					
Population density in 1900			-0.000				
			(0.000)				
Population density in 2005				-0.000***			
				(0.000)			
Latitudinal stretch					0.003		
					(0.002)		
Soil type diversity						0.172**	
						(0.086)	
Vegation type diversity							0.249***
							(0.089)
Constant	0.528***	0.523***	0.537***	0.540***	0.505***	0.419***	0.354***
	(0.049)	(0.049)	(0.050)	(0.049)	(0.050)	(0.064)	(0.071)
Observations	143	143	143	140	142	142	142
Adjusted R-squared	0.45	0.45	0.45	0.46	0.45	0.46	0.47

Notes: Dependent variable in all regressions: ethnic fractionalization from Alesina et al. (2003). Coefficients are reported with robust standard errors in parentheses. ***, ** , and * indicate significance at the 1, 5, and 10 percent levels. For ease of exposition we have in tables 2-7 scaled Origtime to be in 100.000 years units, Agriculture to be in 10.000 years units, Civilization to be in 100 times the population density in year 1, Independence to be in 100 years units, and Latitude to be in 10 degrees units.

Table 6. Geographical determinants of ethnic diversity

	Full sample	Full sample	Full sample	Full sample	Full sample	SSA excluded
	(1)	(2)	(3)	(4)	(5)	(6)
Origtime	0.132***	0.147***	0.107**	0.127***	0.147**	0.213*
	(0.038)	(0.037)	(0.046)	(0.044)	(0.071)	(0.118)
Civilization	-0.722***	-0.727***	-0.740***	-1.107***	-0.574***	-0.523***
	(0.174)	(0.175)	(0.159)	(0.242)	(0.147)	(0.139)
Nation	-0.575***	-0.541***	-0.518***	-0.507***	-0.722***	-0.706***
	(0.156)	(0.159)	(0.163)	(0.160)	(0.171)	(0.177)
Latitude	-0.058**	-0.045***	-0.058***	-0.048***		
	(0.023)	(0.010)	(0.014)	(0.011)		
Average temperature	-0.002					
	(0.005)					
Temperature differences	0.006***			0.006**		
	(0.002)			(0.003)		
Average altitude		-0.033				
		(0.045)				
Altitude difference		0.048**				
		(0.019)				
Average precipitation			-0.001**			
			(0.000)			
Precipitation difference			0.000**			
			(0.000)			
Coastal/riverine land				0.217*		
				(0.128)		
Coastal/riverine population				-0.263**		
				(0.122)		
Geography-induced isolation					0.567***	0.531***
					(0.185)	(0.198)
America					0.156***	0.176***
					(0.052)	(0.059)
Europe					0.002	0.011
					(0.049)	(0.050)
Sub-Saharan Africa					0.174**	
					(0.081)	
Constant	0.549***	0.473***	0.608***	0.544***	0.302***	0.271***
	(0.153)	(0.049)	(0.088)	(0.078)	(0.044)	(0.060)
Observations	142	142	142	130	142	103
Adjusted R-squared	0.48	0.47	0.46	0.54	0.46	0.26

Notes: Dependent variable in all regressions: ethnic fractionalization from Alesina et al. (2003). Coefficients are reported with robust standard errors in parentheses. ***, ** , and * indicate significance at the 1, 5, and 10 percent levels. For ease of exposition we have in tables 2-7 scaled Origtime to be in 100.000 years units, Agriculture to be in 10.000 years units, Civilization to be in 100 times the population density in year 1, Independence to be in 100 years units, and Latitude to be in 10 degrees units.

Table 7. Robustness with respect to influential observations and alternative measures for Origtime and ethnic diversity

	No restriction	Cavalli-Sforza control	Omit if |DForigtime| > 2/sqrt(n)	Omit if leverage > (2k+2)/n	"Normalizing" trans. of all variables	Box-Cox trans. of origtime	Fearon's ethnic diversity
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Origtime	0.135***	0.185***	0.182***	0.143***	0.0283***	0.0739***	0.136***
	(0.26)	(0.36)	(0.35)	(0.29)	(0.17)	(0.23)	(0.26)
Fission		-0.0928					
		(-0.13)					
Civilization	-0.788***	-0.786***	-0.749***	-1.304***	-0.477***	-0.823***	-0.916***
	(-0.26)	(-0.27)	(-0.26)	(-0.21)	(-0.31)	(-0.28)	(-0.21)
Nation	-0.444***	-0.447***	-0.456***	-0.374*	-0.477***	-0.478***	-0.454**
	(-0.18)	(-0.18)	(-0.19)	(-0.13)	(-0.23)	(-0.19)	(-0.18)
Latitude	-0.0459***	-0.0481***	-0.0425***	-0.0411***	-0.111***	-0.0509***	-0.0409***
	(-0.32)	(-0.34)	(-0.30)	(-0.29)	(-0.31)	(-0.36)	(-0.28)
Constant	0.528***	0.536***	0.500***	0.524***	0.905***	0.675***	0.527***
	(2.09)	(2.10)	(1.99)	(2.11)	(4.23)	(2.67)	(2.04)
Observations	143	130	135	137	143	143	138
Adjusted R-squared	0.45	0.46	0.54	0.41	0.42	0.44	0.39

Notes: Dependent variable in all regressions except (7): ethnic fractionalization from Alesina et al. (2003). Coefficients are reported with robust normalized beta coefficients in parentheses. ***, ** , and * indicate significance at the 1, 5, and 10 percent levels. Specification (4): ethnic = sqrt(ethnic), origtime = log(origtime), civilization = sqrt(civilization), nation = nation, latitude = sqrt(latitude). For ease of exposition we have in tables 2-7 scaled Origtime to be in 100.000 years units, Agriculture to be in 10.000 years units, Civilization to be in 100 times the population density in year 1, Independence to be in 100 years units, and Latitude to be in 10 degrees units.

Table 8. Variable descriptions

Variable name	**Description**	**Source**
Agriculture	Years since Neolithic revolution (in 10 000s years)	Putterman (2007a)
Altitude difference	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
Average altitude	Average grid cell altitude*	Authors' calculation based on the G-Econ Dataset (2006)
Average precipitation	Average grid-cell precipitation*	Authors' calculation based on the G-Econ Dataset (2006)
Average temperature	Average grid-cell temperature*	Authors' calculation based on the G-Econ Dataset (2006)
British colony	Dummy for longer British colonization	Cepii
Civilization	Population density in year 1 AD	Estimation from Worldmapper (2006)
Coastal/riverine land	Share of surface area within 100km of Sea or river	Gallup, Sachs, & Mellinger (1998)
Coastal/riverine population	Share of population within 100km of Sea or river	Gallup, Sachs, & Mellinger (1998)
Colonized before 1608	Dummy for being colonized by Europeans before 1608	Olsson (2007)
Colonized before 1820	Dummy for being colonized by Europeans before 1820	Olsson (2007)
Colonized before 1886	Dummy for being colonized by Europeans before 1886	Olsson (2007)
Colony	Dummy for being colonized by Europeans	Olsson (2007)
Duration	Duration of colonization by Europeans	Olsson (2007)
Ethnic fractionalization	Ethnic fractionalization	Alesina et al. (2003)
Europe	Dummy for Europe	Cepii
Fearon's ethnic diversity	Ethnic fractionalization	Fearon (2003)
Fission	Time as separate groups, based on genetic data.	The formula in equation (5) can be restated as $t = -2N \cdot \ln\left(1 - F_{ST}\right)$. By assuming a founding population of 7500 individuals (roughly the same magnitude as used in the specialized literature), we can convert F_{ST}-values into calendar years for the six major genetic clusters. Source of genetic distances: Cavalli-Sforza et al (1994, figure 2.3.3, p 80)
French colony	Dummy for longer French colonization	Cepii
Geography-induced isolation	First principal component of Vegetation diversity, Altitude & Temperature differences, and Latitude	
Independence	Years since independence	ICOW
Latitude	Absolute latitude in degrees	Cepii

Latitudinal stretch	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)
Nation	First principal component of Modern and Independence	
Origtime	Duration of human settlement (in 100 000s years)	See detailed description in section 4
Population density in 1900	Population density in year 1900	Estimation from Worldmapper (2006)
Population density in 2005	Population density in year 2005	WDI online
Precipitation difference	Max. difference in mean grid-cell precipitation*	Authors' calculation based on the G-Econ Dataset (2006)
Premodern	Statehood strength before 1800	Authors' calculation based on Putterman (2007b)
Soil type diversity	Diversity of soil types**	Authors' calculation based on the G-Econ Dataset (2006)
Spanish colony	Dummy for longer Spanish colonization	Cepii
State Antiquity	Statehood strength	Putterman (2007b)
Sub-Saharan Africa	Dummy for sub-Saharan Africa	WDI online
Temperature difference	Maximum difference in mean grid-cell temperature*	Authors' calculation based on the G-Econ Dataset (2006)
Vegetation type diversity	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 soil/vegetation type (of a list of 27 types) are listed for each grid-cell. The measure is calculated as 1 - (#different types listed for each country/#grid cells for each country). Thus if all the same soil/vegetation type dominates all grid cell "measure"=1, if there are two grid cells in the country with different soil/vegetation types "measure"=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>.For ease of exposition we have in tables 2-7 scaled Origtime to be in 100.000 years units, Agriculture to be in 10.000 years units, Civilization to be in 100 times the population density in year 1, Independence to be in 100 years units, and Latitude to be in 10 degrees units.