The Nature of Conflict*

Cemal Eren Arbatlı† Quamrul H. Ashraf‡ Oded Galor§

April 2015

Abstract

This research establishes that the emergence, prevalence, recurrence, and severity of intrastate conflicts in the modern era reflect the long shadow of prehistory. Exploiting variations across national populations, it demonstrates that genetic diversity, as determined predominantly during the exodus of humans from Africa tens of thousands of years ago, has contributed significantly to the frequency, incidence, and onset of both overall and ethnic civil conflict over the last half-century, accounting for a large set of geographical and institutional correlates of conflict, as well as measures of economic development. Furthermore, the analysis establishes the significant contribution of genetic diversity to the intensity of social unrest and to the incidence of intra-group factional conflict. These findings arguably reflect the contribution of genetic diversity to the degree of fractionalization and polarization across ethnic, linguistic, and religious groups in the national population; the adverse influence of genetic diversity on interpersonal trust and cooperation; the contribution of genetic diversity to divergence in preferences for public goods and redistributive policies; and the potential impact of genetic diversity on economic inequality within a society.

Keywords: Civil conflict, genetic diversity, fractionalization, polarization, interpersonal trust, preferences for public goods, economic inequality

JEL classification codes: D74, N30, N40, O11, O43, Z13

---

*We thank Ran Abramitzky, Alberto Alesina, Yann Algan, Sascha Becker, Moshe Buchinsky, Matteo Cervellati, Carl-Johan Dalgaard, David de la Croix, Emilio Depetris-Chauvin, Paul Dower, James Fenske, Raquel Fernández, Boris Gershman, Avner Greif, Pauline Grosjean, Elhanan Helpman, Murat Iyigun, Noel Johnson, Garett Jones, Marc Klemp, Mark Koyama, Stelios Michalopoulos, Nathan Nunn, John Nye, Ömer Özak, Elias Papaioannou, Sergey Popov, Stephen Smith, Enrico Spolaore, Uwe Sunde, Nico Voigtländer, Joachim Voth, Romain Wacziarg, Fabian Waldinger, David Weil, Ludger Wößmann, Noam Yuchtman, Alexei Zakharov, Fabrizio Zilibotti, and seminar participants at George Mason University, George Washington University, HSE/NES Moscow, the AEA Annual Meeting, the conference on "Deep Determinants of International Comparative Development" at Brown University, the workshop on "Income Distribution and Macroeconomics" at the NBER Summer Institute, the conference on "Culture, Diversity, and Development" at HSE/NES Moscow, the conference on "The Long Shadow of History: Mechanisms of Persistence in Economics and the Social Sciences" at LMU Munich, the fall meeting of the NBER Political Economy Program, the session on "Economic Growth" at the AEA Continuing Education Program, the workshop on "Biology and Behavior in Political Economy" at HSE Moscow, and the Economic Workshop at IDC Herzliya for valuable comments. We are especially grateful to Avner Grief, Nathan Nunn, and Enrico Spolaore, all of whom have served as discussants for our paper on different occasions. Gregory Casey, Ashwin Narayan, Daniel Prinz, and Jeffrey Wang provided excellent research assistance. Arbatlı acknowledges research support from the Center for Advanced Studies (CAS) at the National Research University Higher School of Economics (HSE). Ashraf acknowledges research support from the NSF (SES-1338738) and the Hellman Fellows Program. Galor acknowledges research support from the NSF (SES-1338738) and the Population Studies and Training Center (PSTC) at Brown University. The PSTC receives core support from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (5R24HD041020).

†Department of Theoretical Economics, National Research University Higher School of Economics, 26 Shabolovka Street, Building 3, Room 3116A, Moscow, Russia (e-mail: darbatli@hse.ru).

‡Department of Economics, Williams College, Schapiro Hall, 24 Hopkins Hall Drive, Williamstown, MA 01267 (email: Quamrul.H.Ashraf@williams.edu).

§Department of Economics, Brown University, Robinson Hall, 64 Waterman Street, Providence, RI 02912 (email: Oded_Galor@brown.edu).

1 Introduction

Over the course of the 20th century, in the period following World War II, civil conflicts have been responsible for more than 16 million casualties worldwide, well surpassing the cumulative loss of human life associated with international conflicts. Nations plagued by civil conflict have experienced significant fatalities from violence, substantial loss of productive resources, and considerable declines in their standards of living. More than a quarter of all nations across the globe encountered the incidence of civil conflict for at least 10 years during the 1960–2013 time horizon, and although the number of countries experiencing conflict has declined from its peak of 54 in the early 1990s, as many as 35 nations have been afflicted by the prevalence of civil conflict since 2010.[1]

This research establishes that a significant portion of the emergence, prevalence, recurrence, and severity of various forms of intrastate conflict in the modern era reflects the long shadow of prehistory. The analysis demonstrates that the genetic diversity of a contemporary national population – determined predominantly during the "out of Africa" expansion of anatomically modern humans tens of thousands of years ago – has been a significant contributor to the frequency, incidence, and onset of both overall and ethnic civil conflict in society over the last half-century. Moreover, the analysis establishes the significant predictive power of genetic diversity for the intensity of social unrest and for the incidence of intragroup factional conflict. These findings account for the potentially confounding influence of a substantially large set of covariates, including the geographical and institutional correlates of conflict, measures of ethnolinguistic fragmentation, and outcomes of economic development.

The genetic diversity of a national or subnational population can contribute to conflicts in society through several mechanisms. First, genetic diversity may have an adverse effect on the prevalence of mutual trust and cooperation (Ashraf and Galor, 2013a), and excessive diversity can therefore depress the level of social capital below a threshold that otherwise subdues the emergence of social, political, and economic grievances and prevents the culmination of such grievances to violent hostilities. Second, to the extent that genetic diversity captures interpersonal divergence in preferences for public goods and redistributive policies, overly diverse societies may find it difficult to reconcile such differences through collective action, thereby intensifying their susceptibility to internal antagonisms. Third, insofar as genetic diversity reflects interpersonal heterogeneity in traits that are differentially rewarded by the geographical, institutional, or technological environment, it can potentially cultivate grievances that are rooted in economic inequality, thereby magnifying society's vulnerability to internal belligerence.

In addition to the aforementioned mechanisms that apply to both intergroup and intragroup conflict, genetic diversity can also manifest a link with intergroup conflict in society through its potential role in facilitating the endogenous formation of coalitional groups in prehistory and the subsequent differentiation of their respective collective identities over a long expanse of time

[1] These figures are based on the UCDP/PRIO Armed Conflict Dataset, Version 4-2014a (Gleditsch et al., 2002; Themnér and Wallensteen, 2014), compiled by the Uppsala Conflict Data Program (UCDP) and the Peace Research Institute Oslo (PRIO).

(Ashraf and Galor, 2013b). Specifically, following the "out of Africa" migration of humans, the initial endowment of genetic diversity in a given location may have catalyzed the formation of distinct groups at that location through a process of endogenous group selection, reflecting the trade-off between the costs and benefits of intragroup heterogeneity and group size. On the one hand, interpersonal heterogeneity could have adversely affected group productivity by reducing mutual trust and cooperation, but on the other, it could also have stimulated group productivity via interpersonal complementarities across diverse traits and preferences. In particular, since diminishing marginal returns to both intragroup diversity and homogeneity implies a socially optimal size for each group, a larger initial endowment of genetic diversity in a given location may have given rise to a larger number of groups. Over time, due to the forces of "cultural drift" and "biased transmission" of cultural markers that serve to distinguish "insiders" from "outsiders" of a group (e.g., language dialects, customs and traditions, norms of social conduct), intergroup divergence in such markers would have become more pronounced, leading to the formation of distinct collective identities along ethnic lines, and thereby, linking prehistoric genetic diversity with the degree of ethnolinguistic fragmentation observed in a given location today. The resultant fragmentation can then facilitate intergroup conflict in society either directly, by fueling excessive intergroup competition and dissension, or indirectly, by creating more fertile grounds for political elites to exploit ethnic mobilization strategies.

Beyond establishing the salience of a population's genetic diversity as a significant reduced-form contributor to its risk of experiencing various forms of internal conflict, the analysis uncovers evidence suggesting that this reduced-form influence may indeed potentially operate through some of the aforementioned hypothesized mechanisms. In particular, it documents that the quantitative importance of genetic diversity as a deep determinant of the potential for civil conflict in society becomes noticeably diminished in both magnitude and explanatory power when conditioned on the influence of each of three more proximate determinants – namely, the number of ethnic groups in the national population, the prevalence of generalized interpersonal trust at the country level, and the intracountry dispersion in revealed political preferences.

In order to measure the extent of diversity in genetic material across individuals in a given population (e.g., an ethnic group), population geneticists employ an index called expected heterozygosity, which can be interpreted simply as the probability that two individuals, selected at random from the relevant population, are genetically different from one another with respect to a given spectrum of traits. Specifically, the construction of the measure starts with incorporating information on the allelic frequencies for a particular gene or DNA locus – i.e., the proportional representations of different alleles or variants of a given genetic trait in the population. This permits the computation of a gene-specific expected heterozygosity index (i.e., the probability that two randomly selected individuals differ with respect to the genetic trait in question), and upon measuring heterozygosity for a large number of genes or DNA loci, the information is averaged to yield an overall expected heterozygosity for the relevant population.

Although population geneticists provide data on genetic diversity at the ethnic group level, many national populations in the modern world are composed of multiple ethnic groups, some of which have not been indigenous to their current locations since before the great human migrations of the past half-millennium; a pattern that is particularly germane to national populations in the New World. Given that the analysis aims to reveal the influence of genetic diversity on the contemporary risk of intrastate conflict at the national level, the main explanatory variable employed by the analysis is the ancestry-adjusted genetic diversity of a contemporary national population, as constructed by Ashraf and Galor (2013a). In particular, this measure not only accounts for the genetic diversity of a country's prehistorically indigenous subnational ethnic groups as well as the diversity of those descended from immigrant settlers over the past 500 years, but it also accounts for the additional component of national genetic diversity that stems from the genetic distances prevalent amongst these subnational groups.

To the extent that interregional migration flows in the post-1500 era may have been spurred by historically persistent spatial patterns of conflict risk, or by other unobserved (or even observed but noisily measured) correlates thereof, the employment of the ancestry-adjusted measure of national genetic diversity may raise concerns regarding its potential endogeneity in an empirical model that explains the contemporary risk of intrastate conflict. To mitigate these potential concerns, the analysis develops and exploits two alternative empirical strategies that arguably yield better-identified estimates of the influence of genetic diversity on the risk of intrastate conflict in the modern era. As will become evident, both strategies yield remarkably similar results. The first strategy entails confining the regression analysis to exploiting variations in a sample of countries that only belong to the Old World (i.e., Africa, Europe, and Asia), where the genetic diversity of contemporary national populations overwhelmingly reflects the genetic diversity of populations that have been native to their current locations since well before the colonial era. Fundamentally, this strategy rests on the empirical fact that post-1500 population movements within the Old World did not result in the significant admixture of populations that are genetically very distant from one another. Under the second strategy, however, the regression analysis is permitted to exploit variations in a globally representative sample of countries using a two-stage estimator, wherein the migratory distance of a country's prehistorically native population from East Africa is employed as an excluded instrument for the ancestry-adjusted genetic diversity of its contemporary national population. This strategy exploits the extraordinarily strong and negative first-stage impact of migratory distance from the cradle of humankind on the contemporary worldwide distribution of genetic diversity across prehistorically indigenous ethnic groups. In addition, the strategy rests on the identifying assumption that the migratory distance of a country's prehistorically native population from East Africa is exogenous to the risk of intrastate conflict faced by the country's overall population in the last half-century, plausibly satisfying the necessary exclusion restriction in light of a substantially large set of second-stage controls for geographical and institutional factors as well as outcomes of economic development.

The measurement of the genetic diversity of a country's prehistorically native population and the aforementioned two-stage estimation strategy are both based on two widely accepted assertions from the field of population genetics – the "out of Africa" hypothesis of human origins and the existence of a serial founder effect associated with the subsequent demic expansion of humans to the rest of the globe. According to the well-established "out of Africa" hypothesis, the human species, having evolved to its anatomically modern form in East Africa some 150,000 years ago, embarked on populating the entire globe in a stepwise migration process beginning around 90,000–70,000 BP. In addition, the contemporary worldwide distribution of genetic diversity across prehistorically indigenous ethnic groups overwhelmingly reflects a serial founder effect – i.e., a chain of ancient population bottlenecks – originating in East Africa. Specifically, because the spatial diffusion of humans to the rest of the world occurred in a series of discrete steps, where in each step, a subgroup of individuals left their parental colony to establish a new settlement farther away, carrying with them only a subset of the genetic diversity of their parental colony, the genetic diversity of a prehistorically indigenous ethnic group as observed today decreases with increasing distance along ancient human migratory paths from East Africa (e.g., Ramachandran et al., 2005; Prugnolle, Manica and Balloux, 2005).[2]

To place the findings of this research into perspective, conditional on the full set of controls for geographical characteristics, institutional factors, ethnolinguistic fragmentation, and outcomes of economic development, the better-identified estimates obtained from exploiting variations in a globally representative sample of countries suggest that a move from the 10th to the 90th percentile of the global cross-country genetic diversity distribution (equivalent to a move from the diversity level of the Republic of Korea to that of the Democratic Republic of Congo) is associated with an increase in the temporal frequency of civil conflict during the 1960–2008 time horizon by 0.052 new conflict outbreaks per year (or 164 percent of a standard deviation in the global cross-country conflict frequency distribution). In addition, a similar move along the global distribution of genetic diversity across countries leads to (i) an increase in the likelihood of observing the incidence of civil conflict in any given 5-year interval during the 1960–2008 time horizon from 19.2 percent to 34.9 percent; (ii) an increase in the likelihood of observing the onset of a new civil conflict in any given year during the 1960–2008 time horizon from 0.834 percent to 4.23 percent; and (iii) an increase in the likelihood of observing the incidence of one or more intragroup factional conflict events at any point in the 10-year interval between 1990 and 1999 from 14.7 percent to

[2]The measure of expected heterozygosity for prehistorically indigenous ethnic groups is constructed by population geneticists using data on allelic frequencies for a particular class of DNA loci called microsattelites, residing in non-protein-coding or "neutral" regions of the human genome – i.e., regions that do not directly result in phenotypic expression. This measure therefore possesses the advantage of not being tainted by the differential forces of natural selection that may have operated on these populations since their prehistoric exodus from Africa. Critically, however, as argued and empirically established by Ashraf and Galor (2013a,b), the observed socioeconomic influence of expected heterozygosity in microsattelites can indeed reflect the latent impact of heterogeneity in phenotypically and cognitively expressed genomic material, in light of mounting evidence from the fields of physical and cognitive anthropology on the existence of a serial founder effect on the observed worldwide patterns in various forms of intragroup phenotypic and cognitive diversity, including phonemic diversity (Atkinson, 2011) as well as interpersonal diversity in skeletal features pertaining to cranial characteristics (Manica et al., 2007; von Cramon-Taubadel and Lycett, 2008; Betti et al., 2009), dental attributes (Hanihara, 2008), and pelvic traits (Betti et al., 2013).

72.6 percent. Finally, depending on the measure of intrastate conflict severity employed, a move from the 10th to the 90th percentile of the global cross-country genetic diversity distribution is associated with an increase in the intensity of social unrest by either 30.4 percent or 51.2 percent of a standard deviation from the observed distribution of the relevant measure of intrastate conflict severity across countries and 5-year intervals during the 1960–2008 time horizon. These findings are generally shown to be qualitatively insensitive to a wide range of robustness checks, including but not limited to exploiting alternative estimation techniques; accounting for spatial dependence across observations; eliminating a priori statistically influential world regions from the estimation sample; accounting for additional correlates of civil conflict potential that have received attention in the relevant literature; considering alternative definitions or types of intrastate conflict (e.g., large-scale conflicts and conflicts involving only nonstate actors); and employing alternative data sources for proxies of intrastate conflict potential and conflict intensity.

The remainder of the paper is organized as follows: Section 2 highlights the added value of this research to the related literature; Section 3 discusses the data and empirical framework employed for identifying the reduced-form causal influence of genetic diversity on various outcomes associated with intrastate conflict over the past half-century; Section 4 reveals for each of several conflict outcomes examined, the baseline findings as well as the results from robustness checks, and it also presents the findings from an investigation of some of the mechanisms that can potentially mediate the reduced-form influence of genetic diversity; and finally, Section 5 offers some concluding remarks and directions for future research.

2 Advancements with Respect to the Related Literature

This study is the first to empirically establish the salience of diversity in intergenerationally persistent traits across individuals in a given population – namely, its genetic diversity – as an economically and statistically significant reduced-form contributor to the risk of civil discordance faced by the population. In so doing, however, our analysis in this paper speaks to several well-established lines of inquiry.

First and foremost, in light of the fact that the contemporary variation in genetic diversity at the national level predominantly reflects the influence of ancient population bottlenecks that occurred during the "out of Africa" demic expansion of humans to the rest of the world tens of thousands of years ago, our paper contributes to a burgeoning literature – surveyed by Galor (2011), Spolaore and Wacziarg (2013b), and Nunn (2014) – that emphasizes the deeply rooted geographical, sociocultural, and institutional determinants of contemporary variations across populations in economic development and other socioeconomic outcomes. Specifically, amongst other dimensions, our analysis adds value to a major research program within this literature that highlights the importance of the prehistorically determined macrogenetic structure of human populations (e.g., Spolaore and Wacziarg, 2009, 2013a, 2014; Ashraf and Galor, 2013a,b). In contrast to our study, which focuses squarely on the link between the genetic diversity of a national population and the potential of social conflict *within* that population, in a complementary analysis, Spolaore and Wacziarg (2013a) study how the genetic distance between any two national populations is associated with their proclivity to engage in interstate war, empirically documenting a negative relationship between genetic distance and interstate warfare – a finding that is broadly consistent with the view that if genetic relatedness proxies for unobserved similarity in preferences over rival and excludable goods and, thus, the resources necessary for producing them, then violent contentions over the ownership and control of such resources would be more likely to arise between populations that are genetically closer to one another.[3]

Second, considering the fact that the genetic diversity of a national population in part reflects interpersonal differences that are associated with heterogeneity amongst subnational groups in ethnic markers, our study contributes to a vast literature – originating in seminal works like Easterly and Levine (1997) and Alesina, Baqir and Easterly (1999), and surveyed by Alesina and La Ferrara (2005) – that establishes the adverse influence of the ethnolinguistic fragmentation of a national population on various societal outcomes, including the rate of economic growth, the quality of national institutions, the extent of financial development, efficiency in the provision of public goods, the level of social capital, and the potential for civil conflict. As elaborated below, however, because genetic diversity at the national level additionally captures information on interpersonal differences even *within* ethnolinguistically heterogeneous subnational groups, our analysis is able to link outcomes related to social dissonance and aggregate inefficiency with a previously unexplored but nevertheless critical element of intrapopulation diversity.

Third, with respect to the specific societal outcomes that we examine, our paper adds value to the vast literature on the empirical determinants of civil conflict. Notable surveys of work on the subject (e.g., Sambanis, 2002; Collier and Hoeffler, 2007; Blattman and Miguel, 2010) indicate that the origins of civil conflict have been the focus of intensive research over the past two decades. One of the major ongoing debates in this literature – originating from seminal studies by Collier and Hoeffler (1998, 2004) and Fearon and Laitin (2003) – concerns the role of social, political, and economic grievances, relative to the importance of other factors like the capability of the state to subdue armed opposition groups, the conduciveness of geographical characteristics towards rebel insurgencies, or the opportunity cost of engaging in rebellions, as determinants of the risk of civil conflict in society. By highlighting the fact that the genetic diversity of a national population can proxy for both intragroup and intergroup social divisions amongst subnational groups, thereby possibly serving as a deeply rooted catalyst for the manifestation of social, political, and economic grievances along such cleavages, the present study advances our understanding of the nature of grievance-related mechanisms in civil conflict.

[3]It is worth noting that because dissensions in the context of intrastate conflict often arise from grievances associated with incompatibilities in preferences over public (rather than private) goods, to the extent that the genetic diversity of a national population is associated with divergence in such preferences across subnational groups in that population, the empirical findings of Spolaore and Wacziarg (2013a) are not necessarily inconsistent with those from our analysis.

In particular, motivated by the conventional wisdom that intergroup competition over ownership of productive resources and exclusive political power, along with conflicting preferences for public goods and redistributive policies, are more difficult to reconcile in societies that are ethnolinguistically more fragmented, measures based on *fractionalization* indices were initially at the forefront of empirical analyses of grievance-related mechanisms in civil conflict. Nevertheless, early evidence regarding the influence of ethnic, linguistic, and religious fractionalization on the risk of civil conflict in society had been largely inconclusive (Collier and Hoeffler, 1998, 2004; Fearon and Laitin, 2003), arguably due in part to certain conceptual limitations associated with fractionalization indices. In line with this assertion, the subsequent introduction of *polarization* indices to empirical analyses of civil conflict has led to more affirmative findings (Montalvo and Reynal-Querol, 2005; Esteban, Mayoral and Ray, 2012), demonstrating that the notion of intergroup grievances as contributors to the risk of civil conflict in society becomes apparent in the data only when one employs conceptually more meaningful proxies for such grievances. An important shortcoming, however, of all existing measures of ethnolinguistic fragmentation (based on either intergroup fractionalization or polarization indices) is that they are unable to account for the potentially critical role of intragroup heterogeneity in augmenting the risk of conflict in society at large. For instance, important theoretical models of conflict (e.g., Esteban and Ray, 2008b, 2011a) have generated interesting predictions regarding the role of intragroup heterogeneity (in individual income or wealth) in promoting the risk of intergroup conflict due to complementarities between human and material inputs, suggesting a positive link between intragroup diversity, broadly defined, and intergroup conflict in society; a link that cannot be directly tested in a cross-country framework using conventional measures of ethnolinguistic fragmentation. As such, a key advantage of genetic diversity over existing measures of ethnolinguistic fragmentation is that it captures, amongst other elements of diversity at the national level, heterogeneity across individuals even *within* ethnolinguistically differentiated subnational groups – an advantage that a priori permits genetic diversity to retain both economically and statistically significant explanatory power for the potential of civil conflict in society, even after being conditioned on the influence of conventional measures of ethnolinguistic fragmentation in our analysis.

As mentioned previously, genetic diversity at the national level additionally subsumes information on interpersonal heterogeneity associated with differences in ethnic markers, and as such, part of its reduced-form influence on the potential for conflict in society arguably reflects the more proximate impact of intergroup cleavages that conventional measures of ethnolinguistic fragmentation have aimed to capture.[4] Nevertheless, even as a proxy for interethnic divisions, the employment of genetic diversity in empirical analyses of conflict potential can add substantial value in a couple of dimensions over the use of existing proxies that are based on fractionalization and polarization indices. First, notwithstanding some notable exceptions (e.g., Fearon, 2003; Desmet, Ortuño-Ortín and Weber, 2009; Esteban, Mayoral and Ray, 2012), the commonly used measures

[4]Indeed, when exploring some of the mechanisms that can potentially mediate our reduced-form findings, we uncover suggestive evidence consistent with this assertion.

of ethnolinguistic fragmentation typically do not exploit information beyond the proportional representations of ethnolinguistically differentiated groups in the national population – namely, they implicitly assume that these groups are socioculturally "equidistant" from one another. In contrast, genetic diversity at the national level incorporates information on pairwise intergroup genetic distances that were predominantly determined over the course of the "out of Africa" demic diffusion of humans to the rest of the globe tens of thousands of years ago.[5] Second, conventional measures of ethnolinguistic fragmentation are potentially tainted by both measurement error and endogeneity in empirical analyses of civil conflict. Specifically, the individual shares of different ethnolinguistic groups in the national population may be noisily observed in general and may even be systematically mismeasured in more conflict-prone societies, owing to the "fuzzy" and often inconsistent nature of ethnic categories in the data from national censuses (Fearon, 2003), the endogenous "political economy" of national census categorizations of subnational groups, and the endogenous constructivism of individual self-identification with an ethnic group (Eifert, Miguel and Posner, 2010; Caselli and Coleman, 2013; Besley and Reynal-Querol, 2014). In addition, due to atrocities and voluntary or forced migrations associated with historical conflict events, to the extent that temporal persistence in the potential for conflict in society is driven by factors other than interethnic cleavages, the ethnolinguistic configuration of a national population cannot be considered exogenous to the contemporaneous risk of civil conflict in the population (Fletcher and Iyigun, 2010). Although our national-level measure of genetic diversity exploits information on the population shares of subnational groups possessing ethnically differentiated ancestries, the fact that the endowment of genetic diversity in a given location was overwhelmingly determined during the prehistoric "out of Africa" expansion of humans permits our analysis to exploit a plausibly exogenous source of the contemporary cross-country variation in genetic diversity, thereby mitigating the biases associated with many of the measurement and endogeneity issues that may continue to plague the more widely used proxies of ethnolinguistic fragmentation.

Last but not least, it is worth noting how our analysis connects with various perspectives from the social sciences on the formation of ethnic identities, the manifestation of grievances across ethnic boundaries, and the culmination of such grievances to intergroup conflict in society. On this issue, social theory features at least three distinct but not necessarily independent approaches. In particular, the primordialist or essentialist perspective (e.g., Shils, 1957; Geertz, 1973; Brewer, 1979, 1991, 1997; Van den Berghe, 1981, 1995; Horowitz, 1985, 1999; Connor, 1994) harbors

[5] In this respect, the more sophisticated measures of ethnolinguistic fragmentation – such as (i) the Greenberg index of "cultural diversity," as measured by Fearon (2003) and Desmet, Ortuño-Ortín and Weber (2009), or (ii) the ethnolinguistic polarization index of Esteban and Ray (1994), as measured by Desmet, Ortuño-Ortín and Weber (2009), and Esteban, Mayoral and Ray (2012) – incorporate information on pairwise linguistic distances, wherein pairwise linguistic proximity is defined as being monotonic in the number of shared branches between any two languages in a hierarchical linguistic tree, relative to the maximum possible number of branches, which is 15. This information, however, is still somewhat constrained by the nature of a hierarchical linguistic tree, in the sense that the languages that reside at the same level of branching of the tree are all necessarily equidistant from one another. On the other hand, the genetic distance between any two ethnic groups in a contemporary national population predominantly reflects the prehistoric migratory distance between their respective ancestral populations (from the precolonial era), and as follows from the smooth continuity of geographical distances, genetic diversity ends up incorporating a more continuous measure of intergroup distances.

the view that because ethnicity is ultimately rooted in perennial notions of kinship and group-belonging, interethnic relations in society can be charged with the potential for conflict, especially when "groupthink" is conditioned by deep sources of enmity against other groups or the desire to dominate them. On the other hand, the instrumentalist-constructivist approach (e.g., Barth, 1969; Bates, 1983; Horowitz, 1985, 1999; Hardin, 1995; Brass, 1997; Fearon and Laitin, 2000; Brubaker, 2004) argues that although ethnic identities can be conditioned by past conflict events, conflict in society may simply manifest itself along endogenous interethnic boundaries for pragmatic reasons, including but not limited to the mobilization of ethnic networks by "ethnic entrepreneurs" as devices for effective monitoring, enforcement against free-riding, and easier access to financing. Finally, advocates of the modernist viewpoint (e.g., Bates, 1983; Gellner, 1983; Wimmer, 2002) stress that interethnic conflict arises from increased competition over scarce resources, especially when previously marginalized groups that were excluded from the nation-building process experience socioeconomic modernization and, thus, begin to challenge the status quo.

Although our contribution is not immediately relevant for the assessment of the modernist approach, it is not inconsistent with either primordialist or instrumentalist perspectives. Because the initial endowment of interpersonal diversity at a given location – as governed by the "out of Africa" demic diffusion process – may have facilitated the endogenous formation of coalitional groups at that location in prehistory, with collective identities thereafter diverging over time under the forces of "cultural drift," a reduced-form link between the prehistorically determined genetic diversity of a national population and its contemporary risk of interethnic conflict may well be apparent in the data, regardless of whether these groups today are mobilized into conflict by "ethnic entrepreneurs" that aim to reinforce ethnic identities for their private interests or whether such identities entirely reflect primordial notions of kinship and group-belonging, with conflict between groups being driven by deeply rooted grievances.[6]

3 Data and Empirical Framework

This section discusses our data and empirical framework for identifying the reduced-form causal influence of genetic diversity on various conflict outcomes. We first describe the key variables employed by our analysis and then introduce the empirical models that we estimate by exploiting variations in either cross-country or repeated cross-country data.

[6] The argument that genetic diversity in the distant past may have contributed to ethnic heterogeneity as observed in the modern era is consistent with the sociobiological perspective of ethnic origins (e.g., Van den Berghe, 1981, 1995), rooted mainly in the dual-inheritance theory of gene-culture coevolution from the field of evolutionary anthropology (e.g., Durham, 1991; Cavalli-Sforza, Menozzi and Piazza, 1994). Accordingly, like most other mammals, human beings exhibit nepotistic behavior, including greater loyalty to their immediate kin, extended family, or clan, because such behavior can ultimately serve to maximize the likelihood of passing on one's genes successfully to future generations. As such, the formation of collectives in prehistory is partly viewed as a manifestation of "extended nepotism," with the subsequent intergroup differentiation of collective identities occurring over a long expanse of time through the forces of "cultural drift" (Cavalli-Sforza and Feldman, 1981) and "biased transmission" (Boyd and Richerson, 1985; Heinrich and McElreath, 2003) of cultural markers – e.g., language dialects, customs and traditions, and norms of social conduct – that partly serve to distinguish the "outsiders" from the "insiders" of a group.

3.1 Data

Our baseline sample contains information on 143 countries. We follow the norm in the empirical literature on civil conflict by focusing our analysis on the post-1960 time period. Since most of the previous European colonies in Sub-Saharan Africa, the Middle East, and South and Southeast Asia had become independent nation states by 1960, this time horizon permits an assessment of the correlates of civil conflict at the national level, independently of their interactions with the presumably distorting contemporaneous influence of the colonial power on the potential for internal conflict. Due to constraints on the availability of data for some of our baseline covariates, the sample period for our analyses of civil conflict is 1960–2008. In the following sections, we describe our main outcome variables, discuss the measurement of genetic diversity, and briefly introduce the covariates included in our baseline specifications.

3.1.1 Outcome Variables: Frequency, Incidence, and Onset of Overall and Ethnic Civil Conflict

The core dependent variables in our analysis reflect various outcomes related to either overall civil conflict (i.e., of the type that occurs regardless of interethnic divisions) or ethnic civil conflict. The main data source that we rely on for conflict events is the UCDP/PRIO Armed Conflict Dataset, Version 4-2012 (Gleditsch et al., 2002; Themnér and Wallensteen, 2012). By definition, a civil conflict refers to an internal armed conflict between the government of a state and one or more internal opposition groups, fighting over a given incompatibility. This definition includes internationalized internal conflicts, involving intervention from other states on either side of a civil conflict. Not all events that satisfy this definition, however, are included in the data set. Specifically, the UCDP/PRIO data set only records those conflict events that have led to at least 25 battle-related deaths in a given year. For our main analysis, we employ the most comprehensive conflict coding – namely, PRIO25 – encompassing all conflict events that resulted in 25 or more battle-related deaths in a given year. Table B.3 in Appendix B lists all the countries in our baseline sample that experienced at least one PRIO25 civil conflict outbreak during the 1960–2008 time horizon, along with the number of distinct conflict outbreaks and the fraction of years of active conflict experienced by each country in the sample period.[7]

Recent evidence uncovered by Ashraf and Galor (2013b) supports the notion that following the prehistoric "out of Africa" migration of humans, the genetic diversity of an indigenous settlement may have served as a domain over which endogenous group selection and subsequent intergroup cultural or ethnic differentiation had taken place at that location. As already mentioned, this deeper mechanism may indeed be one of the primary channels through which genetic diversity influences the potential for intergroup conflict in contemporary national populations. One way to assess the validity of this argument is to investigate the influence of genetic diversity on the potential for those types of civil conflict in which interethnic divisions are presumably a more

[7]For the interested reader, the table also reports the specific decile from the global cross-country genetic diversity distribution to which each country belongs.

germane issue. As such, our analysis also focuses on outcomes associated with *ethnic* civil conflict, as defined by Wimmer, Cederman and Min (2009) (henceforth referred to as WCM09). Using the UCDP/PRIO Armed Conflict Dataset as their primary data source for civil conflict events, WCM09 additionally apply an "ethnic" categorization to the data, identifying those conflict events wherein the opposition group(s) either explicitly pursue ethnonationalist aims (e.g., attempt to secure ethnonational self-determination, ethnoregional autonomy, or language and other cultural rights) or are motivated by ethnic concerns (e.g., ethnic balance of power in the government or ethnic discrimination), and in which they recruit fighters and forge political alliances on the basis of ethnic affiliations. Since WCM09 employ an earlier version – namely, Version 3-2005b – of the UCDP/PRIO Armed Conflict Dataset, the sample employed by our analyses of ethnic civil conflict contains information limited to 141 countries in the 1960–2005 time horizon. Corresponding to Table B.3, Table B.4 in Appendix B provides some descriptive statistics for those countries in our baseline sample that experienced at least one WCM09 ethnic civil conflict outbreak during this time period.

Depending on the unit of analysis, our outcome variables capture different dimensions of either overall or ethnic civil conflict. In our cross-country regressions, for instance, the outcome variables record the annual frequency (i.e., the average number per year) of "new" civil conflict outbreaks – involving a new issue of incompatibility and/or a new set of nonstate actors fighting against the government – over the relevant sample period. These outcome variables therefore reflect the number of distinct incompatibilities between state actors and armed opposition groups that have, on average, escalated to a full-blown conflict – as defined by the battle-related death threshold – on a yearly basis. Many civil conflicts, however, span several years and may even comprise multiple conflict episodes that are separated by one or more years of inactivity – i.e., years of actual peace or in which the annual battle-related death toll is below the specified threshold. Our regressions based on repeated cross-country data exploit this temporal dimension of civil conflict. Specifically, in our regressions explaining the incidence (prevalence) of civil conflict, the outcome variable is an indicator, coded 1 for each country-period (a period typically being a 5-year time interval) in which there is at least one active conflict-year observed, and 0 otherwise. On the other hand, in our regressions explaining the onset of civil conflict, we employ two qualitatively distinct indicators of conflict outbreak. Our first outcome measure is the standard PRIO2 onset variable, coded 1 for any given country-year when a (possibly recurrent) conflict episode erupts after at least two years of uninterrupted civil peace, whereas our second measure – namely, PRIO-NC – is coded 1 for any given country-year if and only if the country experienced the eruption of (the first episode of) a new civil conflict in that year.

3.1.2 Main Explanatory Variable: Genetic Diversity

The measurement of observed genetic diversity at the ethnic group level is based on an index referred to by population geneticists as expected heterozygosity. Like the interpretation of many other measures of diversity, this index conceptually reflects the probability that two individuals, selected at random from the relevant population, are genetically different from one another with respect to a given spectrum of genetic traits. The index itself is constructed by population geneticists using data on allelic frequencies – i.e., the frequency with which a gene variant or allele (e.g., the brown versus blue variant of the eye-color genetic trait) occurs in a given population.[8] Using information on the allelic frequencies in a given population for a particular gene or DNA locus, it is possible to compute a gene-specific heterozygosity statistic (i.e., the probability that two randomly selected individuals will differ with respect to the gene in question), which when averaged across multiple genes, yields an aggregate expected heterozygosity for the population. Formally, consider the case of a single gene or DNA locus, l, with k_l observed variants or alleles in the population, and let p_i denote the frequency of occurrence of the ith allele. Then, the expected heterozygosity of the population, H^l_{exp}, with respect to locus l is

$$H^l_{exp} = 1 - \sum_{i=1}^{k_l} p_i^2, \tag{1}$$

which given allelic frequencies for each of m different genes or DNA loci, can be averaged across these loci to yield an aggregate measure, H_{exp}, of expected heterozygosity as

$$H_{exp} = 1 - \frac{1}{m} \sum_{l=1}^{m} \sum_{i=1}^{k_l} p_i^2. \tag{2}$$

Like standard measures of ethnolinguistic fragmentation, based on fractionalization or polarization indices, *observed* genetic diversity would potentially be endogenous in an empirical model of civil conflict, since it could be tainted by genetic admixtures resulting from the movement of populations across space, triggered by cross-regional differences in patterns of historical conflict potential, the nature of political institutions, and levels of economic prosperity. To circumvent this problem, we employ the measure of contemporary genetic diversity previously introduced to the literature by Ashraf and Galor (2013a). Specifically, this measure captures (amongst other dimensions of contemporary genetic diversity at the national level, as explained below) the component of observed interpersonal diversity associated with a country's indigenous ethnic groups that is *predicted* by migratory distance from Addis Ababa, Ethiopia to the country's modern-day capital city, along prehistoric land-connected human migration routes.[9]

[8] In molecular genetics, an allele is defined as any one of a number of viable DNA codings (formally, a sequence of nucleotides) that occupy a given locus (or position) in a chromosome. Chromosomes themselves are packages for carrying strands of DNA molecules in cells, and they comprise multiple loci that typically correspond to some of the observed discrete units of heredity (or genes) in living organisms. For additional details on basic concepts and definitions from molecular and population genetics, the interested reader is referred to Griffiths et al. (2000) and Hartl and Clark (2007).

[9] These routes incorporate five obligatory intermediate waypoints, namely Cairo, Egypt; Istanbul, Turkey; Phnom Penh, Cambodia; Anadyr, Russia; and Prince Rupert, Canada. In contrast to a measure of direct geodesic distance from East Africa, the use of these intermediate waypoints ensures that the measure of migratory distance more accurately reflects the fact that humans did not cross large bodies of water in the course of their prehistoric exodus from Africa.

Exploiting the explanatory power of a serial founder effect associated with the "out of Africa" migration process, the predicted genetic diversity of a country's prehistorically indigenous population is generated by applying the coefficients obtained from an ethnic-group-level regression (e.g., Ramachandran et al., 2005; Prugnolle, Manica and Balloux, 2005) of expected heterozygosity on migratory distance from Addis Ababa, Ethiopia, in a sample comprising 53 globally representative ethnic groups from the Human Genome Diversity Cell Line Panel, compiled by the Human Genome Diversity Project (HGDP) in collaboration with the Centre d'Etudes du Polymorphisme Humain (CEPH). According to population geneticists, these ethnic groups have not only been prehistorically native to their current geographical locations, but they have also been largely isolated from genetic flows from other ethnic groups.[10]

It is especially relevant to note that the measure of expected heterozygosity in the sample of 53 HGDP-CEPH ethnic groups is constructed using data on allelic frequencies for a particular class of DNA loci called microsattelites, residing in non-protein-coding or "neutral" regions of the human genome – i.e., regions that do not directly result in phenotypic expression. This measure of observed genetic diversity therefore possesses the advantage of not being tainted by the differential forces of natural selection that may have operated on these populations since their prehistoric exodus from Africa. Importantly, however, we expect that the observed socioeconomic influence of expected heterozygosity in microsattelites should reflect the unobserved impact of diversity in phenotypically and cognitively expressed genomic material, in light of mounting evidence from the fields of physical and cognitive anthropology regarding the existence of a serial founder effect – associated with the prehistoric "out of Africa" migration process – on worldwide spatial patterns in various forms of intragroup phenotypic and cognitive diversity, including phonemic diversity (Atkinson, 2011) as well as interpersonal diversity in skeletal features pertaining to cranial characteristics (Manica et al., 2007; von Cramon-Taubadel and Lycett, 2008; Betti et al., 2009), dental attributes (Hanihara, 2008), and pelvic traits (Betti et al., 2013).

In the absence of systematic and large-scale population movements across geographically (and, thus, genetically) distant regions, as had been largely true during the precolonial era, the genetic diversity of the prehistorically native population in a given location serves as a good proxy for the contemporary genetic diversity of that location. While this continues to remain true to a large extent for nations in the Old World (i.e., Africa, Europe, and Asia), post-1500 population flows from the Old World to the New World has had a considerable impact on the ethnic composition and, thus, the contemporary genetic diversity of national populations in the Americas and Oceania. Thus, instead of employing the genetic diversity of prehistorically native populations (i.e., precolonial genetic diversity) at the expense of limiting our entire analysis to the Old World, we adopt the measure of ancestry-adjusted genetic diversity (i.e., contemporary genetic diversity) from Ashraf and Galor (2013a) as our main explanatory variable. Using the shares of different groups in a country's modern-day population, this measure not only accounts for the

[10]For a more detailed description of the HGDP-CEPH Human Genome Diversity Cell Line Panel data set, the interested reader is referred to Cann et al. (2002). A broad overview of the Human Genome Diversity Project is provided by Cavalli-Sforza (2005).

genetic diversity of the ethnic groups that can trace native ancestry to the year 1500 as well as the diversity of those descended from immigrant settlers over the past half-millennium, but it also accounts for the additional component of genetic diversity at the national level that arises from the pairwise genetic distances that exist amongst these different subnational groups.[11]

The pairwise correlation between the measures of precolonial and contemporary genetic diversity is about 0.993 across countries in the Old World. Due to the influence of post-1500 migrations from the Old World to the New World, however, the same correlation is 0.750 in the global sample. These correlations establish that, on the one hand, post-1500 population flows did not have any significant impact on the genetic diversity of contemporary national populations in the Old World. On the other hand, the difference between the two measures of genetic diversity are far from negligible amongst nations in the New World, indicating that in a global analysis of civil conflict, the ancestry-adjusted measure of genetic diversity should indeed be the relevant explanatory variable of interest. At the same time, however, the ancestry-adjusted measure may not be fully immune from potential endogeneity issues in an empirical model of civil conflict. Specifically, one concern is that it may partly reflect endogenous cross-country migrations in the post-1500 era – namely, migrations that were driven by past conflict events or other unobserved (or even observed but noisily measured) correlates of historical conflict potential. In Section 3.2, we discuss our strategies for mitigating this concern.

3.1.3 Control Variables: Geography, Institutions, Ethnolinguistic Fragmentation, and Development Outcomes

The vast empirical literature on the determinants of civil conflict has emphasized a large number of potentially contributing factors. Drawing on this literature, we include a substantial set of control variables in our baseline specifications. Apart from previous studies on the subject, our specific measure of intrapopulation diversity raises the need to account for the potentially confounding influence of geography, an issue that we address by controlling for a wide range of measurable geographical characteristics. This section only briefly describes these and other covariates in our analysis. A discussion of the deeper conceptual justifications for considering them will be deferred until Section 4, when we reveal our main empirical results.

Geographical Characteristics Given that the predicted intragroup component of our ancestry-adjusted measure of genetic diversity varies linearly with prehistoric migratory distance from East Africa, we consider a large set of geographical attributes that may be correlated with migratory distance and that can also reasonably impart a reduced-form influence on conflict risk through channels unrelated to genetic diversity at the national level. To this end, absolute latitude, total land area, terrain ruggedness, distance to the nearest waterway, and the mean and range of both

[11]The data on the population shares of these different subnational groups at the country level are obtained from the World Migration Matrix, 1500–2000 of Putterman and Weil (2010), who compile for each country in their data set, the share of the country's population in 2000 that is descended from the population of every other country in 1500. For an in-depth discussion of the methodology underlying the construction of the ancestry-adjusted measure of genetic diversity, the reader is referred to the data appendix of Ashraf and Galor (2013a).

agricultural land suitability and elevation are included in our baseline set of covariates.[12] Our baseline specifications additionally account for a complete set of continent fixed effects to ensure that the estimated reduced-form impact of genetic diversity on conflict potential is not simply reflecting the latent influence of unobserved time-invariant cultural, institutional, and geographical factors at the continent level.[13]

Institutional Factors We consider two different sets of covariates in our baseline specifications to control for the impact of colonial legacies on contemporary political institutions and conflict potential. Depending on the unit of analysis, the first set comprises either binary indicators for the historical prevalence of colonial rule (as is the case in our cross-country regressions) or time-varying measures of the lagged prevalence of colonial rule (as is the case in our regressions using repeated cross-country data), but in either case, we distinguish between colonial rule by the U.K., France, and any other major colonizing power. Regardless of the unit of analysis, however, the second set of covariates comprises time-invariant binary indicators for British and French legal origins, included to account for any latent influence of legal codes and institutions that may not necessarily be captured by colonial experience.[14]

Our baseline specifications additionally include three control variables, all based on yearly data at the country level from the Polity IV Project (Marshall, Gurr and Jaggers, 2009), in order to account for the direct influence of contemporary political institutions on the risk of civil conflict. The first variable is based on an ordinal index that reflects the degree executive constraints in any given year, whereas the other two variables are based on binary indicators for the type of political regime, reflecting the prevalence of either democracy (when the polity score is above 5) or autocracy (when the polity score is below -5) in a given year.[15] In light of the time-varying nature of these variables, our regressions based on repeated cross-country data are able to exploit their temporal variations by controlling for either their respective temporal means over the previous 5-year interval (in the case of quinquennially repeated data) or simply their respective lagged values (in the case of annually repeated data). Our cross-country regressions, however, control for the temporal means of these variables over the entirety of the relevant sample period.

[12]The data for absolute latitude, total land area, and distance to the nearest waterway are obtained from the Central Intelligence Agency (2006), the World Bank (2006), and Gallup, Sachs and Mellinger (1999), respectively. Nordhaus (2006) provides disaggregated geospatial data at a 1-arc-minute resolution on surface undulation and elevation, from which we derive our country-level aggregate measures of terrain ruggedness and the mean and range of elevation. Finally, we obtain the country-level aggregate measures of the mean and range of agricultural land suitability directly from the data set of Michalopoulos (2012). See the data appendices of Ashraf and Galor (2013a,b) for further details.

[13]In addition to "soaking up" the possibility of omitted-variable bias from unobserved time-invariant characteristics at the continent level, the need to account for continent fixed effects is perhaps even more binding for observed non-geographical factors, given the potential for systematic measurement error at the continent level in covariates reflecting cultural and institutional characteristics.

[14]The country-level indicators for British and French legal origins are obtained directly from the data set of La Porta et al. (1999). The measures of historical and contemporary colonial rule, on the other hand, are constructed using a number of secondary sources, and we refer the reader to the data appendix of Ashraf and Galor (2013b) for further details.

[15]The prevalence of anocracy, occurring when the polity score is between -5 and 5, therefore serves as the omitted political regime category.

Ethnolinguistic Fragmentation We include two time-invariant proxies for the degree of ethnolinguistic fragmentation of a national population in our baseline set of covariates, in order to account for the influence of this particular channel on the potential for civil conflict. Our first proxy is the well-known ethnic fractionalization index of Alesina et al. (2003), reflecting the probability that two individuals, randomly selected from a country's population, will belong to different ethnic groups. Our second proxy for this channel – owing to its more appealing theoretical foundation and stronger predictive power in previous empirical analyses of civil conflict – is an index of ethnolinguistic polarization, obtained from the data set of Desmet, Ortuño-Ortín and Wacziarg (2012). At a conceptual level, this index measures the extent to which the ethnic composition of a country's population resembles a perfectly polarized configuration, in which the national population consists of two ethnic groups of equal size. Further, although Desmet, Ortuño-Ortín and Wacziarg (2012) provide measures of several such polarization indices, constructed at different levels of aggregation of linguistic groups in a country's population (based on hierarchical linguistic trees), the specific polarization measure we employ is the one corresponding to their most disaggregated level – namely, level 15. This measure therefore reflects the extent of polarization across subnational groups classified according to modern-day languages, thus making it conceptually comparable to other polarization indices used in the literature.[16]

Development Outcomes The empirical literature on civil conflict has proposed various causal mechanisms through which revenues collected from the extraction of natural resources, the size of a country's population, and the standard of living may each correlate – in a reduced-form sense – with conflict risk in a national population. In light of the fact that these variables are all codetermined with the level of economic development, and because genetic diversity has been shown to confer a hump-shaped influence on productivity at the country level by Ashraf and Galor (2013a), our baseline specifications include controls for the intensity of oil production, total population size, and GDP per capita, in order to estimate the impact of genetic diversity on conflict potential, independently from its indirect influence through the correlates of economic development. In particular, these control variables are derived from annual time-series data at the country level on the per-capita value of oil production from Ross (2013) and on population size and per-capita GDP from Maddison (2010). As is the case with our time-varying controls for institutional quality, these variables enter our cross-country regressions as their respective temporal means over the entirety of the relevant sample period, but in our regressions using repeated cross-country data, they enter as either their respective temporal means over the previous 5-year interval (when we exploit quinquennially repeated data) or simply their respective lagged values (when using annually repeated data). In all cases, however, we apply a log transformation to each of these variables before including them as covariates in our regressions.

[16] We prefer employing Desmet, Ortuño-Ortín and Wacziarg (2012) as the data source for our baseline control variable that captures the degree of ethnolinguistic polarization, primarily due to the more comprehensive geographical coverage of their data set, relative to other potential data sources such as Montalvo and Reynal-Querol (2005) and Esteban, Mayoral and Ray (2012). In addition, although our preferred polarization measure does not incorporate intergroup linguistic distances, in a robustness check, we control for an alternative measure of polarization from the data set of Esteban, Mayoral and Ray (2012) that surmounts this particular shortcoming.

It is worth noting at this stage that we expect many of our controls for institutional quality, ethnolinguistic fragmentation, and the correlates of economic development to be endogenous in an empirical model of civil conflict, and as such, their estimated coefficients in our regressions do not permit a causal interpretation. Nonetheless, as will become more evident, controlling for these factors is essential to minimize specification errors and assess the extent to which the reduced-form influence of genetic diversity on conflict potential can be attributed to more conventional explanations in the literature.

3.1.4 Summary Statistics

Tables B.1 and B.2 in Appendix B present the summary statistics of all the variables in the baseline samples exploited by our cross-country analyses of overall and ethnic civil conflict frequency, respectively. With regard to our dependent variables, the cross-country mean of our measure of the annual frequency of new PRIO25 civil conflict outbreaks is 0.020 onsets per year in the 1960–2008 time horizon, and its standard deviation is 0.030.[17] The corresponding statistics for our measure of the annual frequency of new WCM09 ethnic civil conflict outbreaks in the 1960–2005 time horizon are very similar, possessing a sample mean of 0.018 and a standard deviation of 0.033. In terms of our main explanatory variable, the mean level of genetic diversity is 0.728, which roughly corresponds to the diversity of the national populations in Central Asia (e.g., Kazakhstan and Uzbekistan), and its standard deviation is 0.027. Further, the 10th and the 90th percentiles of the cross-country distribution of genetic diversity are 0.688 and 0.752, corresponding approximately to the diversity of the national populations of the Republic of Korea and the Democratic Republic of the Congo, respectively.

3.2 Empirical Framework

Our empirical framework comprises different specifications, exploiting variations in either cross-country or repeated cross-country data, in order to investigate the explanatory power of genetic diversity for observed variations in three different dimensions of civil conflict in the post-1960 time period, namely (i) the annual frequency of new conflict outbreaks over the sample period; (ii) the likelihood of conflict prevalence in any given 5-year interval; and (iii) the likelihood of a conflict outbreak in any given year.

3.2.1 Analysis of Civil Conflict Frequency in Cross-Country Data

Our cross-country regressions attempt to explain the variation across countries in the annual frequency of new civil conflict onsets – i.e., the average number of new civil conflict eruptions

[17] Due to the fact that new civil conflict outbreaks are rare events in the data, our dependent variables both possess zero-inflated, positively skewed cross-country distributions. Thus, in an effort to improve the fit of our empirical models under OLS estimation, we apply a log transformation to each of our dependent variables after scaling them up by one unit. We address this issue more formally, however, in robustness checks that estimate negative-binomial and Poisson models for the count of new civil conflict outbreaks over the relevant sample period.

per year – over the sample period. Specifically, the baseline empirical model for our cross-country analysis is as follows.

$$CF_i = \beta_0 + \beta_1 \widehat{GD}_i + \beta_2' GEO_i + \beta_3' INS_i + \beta_4' ETH_i + \beta_5' DEV_i + \varepsilon_i, \tag{3}$$

where CF_i is the (log transformed) average number of new civil conflict outbreaks per year in country i; $\widehat{GD}_i$ is the ancestry-adjusted genetic diversity of the national population; GEO_i, INS_i, ETH_i, and DEV_i are the respective vectors of control variables for geographical characteristics (including continent fixed effects), institutional factors, ethnolinguistic fragmentation, and the correlates of economic development, as described in Section 3.1.3; and finally, ε_i is a country-specific disturbance term. As mentioned previously, all the time-varying controls for institutional factors and development outcomes enter the model as their respective temporal means over the relevant sample period – namely, 1960–2008 in the regressions explaining the annual frequency of new PRIO25 civil conflict outbreaks, and 1960–2005 in the regressions explaining the same outcome for WCM09 ethnic civil conflict onsets.

We start by performing our cross-country regressions using the OLS estimator in our global sample of 143 countries. Nevertheless, since the ancestry-adjusted measure of genetic diversity accounts for the impact of cross-country migrations in the post-1500 era on the diversity of contemporary national populations, and because these migrations may have been spurred by historically persistent spatial patterns of conflict, the estimated coefficient on genetic diversity in such a framework could be afflicted by endogeneity bias. We employ two alternative strategies to address this issue. The first strategy is to simply confine our analysis to exploiting variations across countries that only belong to the Old World, where as discussed previously, the genetic diversity of contemporary national populations overwhelmingly reflects the genetic diversity of populations that have been native to their current locations since well before the colonial era – a pattern that primarily arises from the fact that historical cross-country migrations in the Old World did not result in the admixture of populations that are genetically distant from one another. The second strategy is to exploit variations in our global sample of countries with the 2SLS estimator, employing the migratory distance from East Africa of the prehistorically native population in each country as an instrument for the country's contemporary genetic diversity. The identifying assumption therefore is that the migratory distance of a country's prehistorically native population from East Africa is exogenous to the risk of civil conflict in the post-1960 time period, plausibly satisfying the exclusion restriction, conditional on our substantial set of controls for geographical and institutional factors as well as the correlates of economic development.

3.2.2 Analysis of Civil Conflict Incidence in Repeated Cross-Country Data

The second dimension of civil conflict that we examine is its temporal prevalence. Specifically, exploiting the time structure of quinquennially repeated cross-country data, we investigate the predictive power of genetic diversity for the likelihood of observing the incidence of one or more active conflict episodes in a given 5-year interval during the post-1960 time horizon.[18] For our baseline analysis of conflict incidence, we estimate the following probit model using maximum-likelihood estimation.[19]

$$CP^*_{i,t} = \gamma_0 + \gamma_1 \widehat{GD}_i + \gamma_2' GEO_i + \gamma_3' INS_{i,t-1} + \gamma_4' ETH_i + \gamma_5' DEV_{i,t-1} + \gamma_6 C_{i,t-1} + \gamma_7' \delta_t + \eta_{i,t} \equiv \gamma' Z_{i,t} + \eta_{i,t}; \quad (4)$$

$$C_{i,t} = \mathbf{1}\left[CP^*_{i,t} \geq D^*\right]; \quad (5)$$

$$Pr\left(C_{i,t} = 1 | Z_{i,t}\right) = Pr\left(CP^*_{i,t} \geq D^* | Z_{i,t}\right) = \Phi\left(\gamma' Z_{i,t} - D^*\right), \quad (6)$$

where $CP^*_{i,t}$ is a latent variable measuring the potential for an active conflict episode in country i during any given 5-year interval, t, and it is modeled as a linear function of explanatory variables. Further, the time-invariant explanatory variables $\widehat{GD}_i$, GEO_i, and ETH_i are all as previously defined, but now, the time-varying covariates included in $INS_{i,t-1}$ and $DEV_{i,t-1}$ enter as their respective temporal means over the previous 5-year interval, δ_t is a vector of time-interval (5-year period) dummies, and $\eta_{i,t}$ is a country-period-specific disturbance term.[20] By specifying each of our time-varying controls to enter the model with a one-period lag, we aim to mitigate the concern that the use of contemporaneous measures of these covariates may exacerbate reverse-causality bias in their estimated coefficients.[21] Finally, we assume that contemporary conflict potential additionally depends on the lagged incidence of civil conflict, $C_{i,t-1}$, which accounts for the possibility that countries with a conflict experience in the immediate past may exhibit a higher conflict potential in the current period, mainly because of the intertemporal spillovers that are common to most conflict processes – e.g., the self-reinforcing nature of past casualties on either side of a conflict.[22] Because the continuous variable reflecting conflict potential, $CP^*_{i,t}$, is unobserved, its level can only be inferred from the binary incidence variable, $C_{i,t}$, indicating whether the latent conflict potential

[18] Following the norm in the literature for analyzing the prevalence of civil conflict, our outcome measure is coded to capture the "continuation" or temporal persistence of experiencing an active state of conflict, and as such, it does not distinguish between new conflicts and subsequent episodes of preexisting conflicts.

[19] In robustness checks, however, we confirm that our main findings with respect to both the incidence and the onset of civil conflict in repeated cross-country data remain fully intact under both the standard logit and the "rare events" logit alternatives for the modeling of incidence and onset events in the data.

[20] We confirm the robustness of our analysis of conflict incidence to exploiting variations in annually (rather than quinquennially) repeated cross-country data. Naturally, in those regressions, the time-dependent covariates enter as their lagged annual values (instead of their lagged 5-year temporal means) and time fixed effects are captured by a set of year dummies.

[21] An alternative method to address the reverse-causality problem, in the context of quinquennially repeated cross-country data, is to control for time-dependent covariates as measured in the initial year of each 5-year interval. Although this method would retain the first period-observation for each country, which is dropped under the current specification, it leaves open the possibility that the presence or absence of an active conflict in the first year of each period may still exert a direct influence on the time-varying controls.

[22] In adopting this strategy, our analysis of conflict incidence follows Esteban, Mayoral and Ray (2012). We also note here that because our measure of genetic diversity is time-invariant (as is indeed the case with all known measures of ethnolinguistic fragmentation, based on fractionalization or polarization indices), we are unable to account for country fixed effects in our model or exploit dynamic panel estimation methods, despite the time dimension in our repeated cross-country data. In all our regressions exploiting such data, however, the robust standard errors of the estimated coefficients are always clustered at the country level.

was sufficiently intense for the annual battle-related death threshold of a civil conflict episode to have been surpassed during a given 5-year interval. As is evident from equations (5)-(6), D^* is the corresponding threshold for unobserved conflict potential, and it appears as an intercept in $\Phi(.)$, the cumulative distribution function for the disturbance term, $\eta_{i,t}$.

To address the potential endogeneity of the measure of contemporary genetic diversity in our analysis of conflict incidence, we follow the same strategies as those employed for our cross-country regressions. Specifically, we first confine our probit analysis to exploiting variations in a repeated cross-country sample that comprises nations from only the Old World, and we then re-estimate our model of conflict incidence in a globally representative sample using an instrumental-variables maximum-likelihood probit (IV probit) estimator. In the latter case, the migratory distance of a country's prehistorically native ethnic groups from East Africa is employed as an excluded instrument for the ancestry-adjusted genetic diversity of its contemporary national population, in a first-stage regression estimated using OLS.

3.2.3 Analysis of Civil Conflict Onset in Repeated Cross-Country Data

The third dimension of conflict examined by our analysis is the onset of civil conflict. Unlike the model of conflict incidence, the onset model focuses solely on explaining the outbreak of conflict events, classifying the subsequent years into which a given conflict persists as nonevent years (akin to civil peace), unless they coincide with the eruption of another conflict. Conceptually, this model assesses the extent to which genetic diversity at the national level influences sociopolitical instability by *triggering* conflicts, rather than only contributing to their perpetuation over time. The probit model for our analysis of conflict onset is similar to the model of conflict incidence, described by equations (4)-(6), except that now, following the convention in the literature, (i) we exploit variations in annually repeated cross-country data, with our binary outcome variable assuming a value of 1 if a country-year observation coincides with the first year of a "new period" of conflict (as discussed below), and 0 otherwise; and (ii) a set of cubic splines of the number of preceding years of uninterrupted peace is included as a control, along with year dummies, in order to account for temporal or duration dependence (Beck, Katz and Tucker, 1998). Further, to mitigate issues of causal identification of the influence of genetic diversity on conflict onset, we implement the same two strategies followed by our analyses of conflict frequency and conflict incidence.

What constitutes a "new period" of conflict depends on the specific onset measure – introduced in Section 3.1.1 – that we employ in our analysis. Specifically, under the PRIO2 measure, a new period of conflict is initiated either by a new episode of a preexisting conflict or by the first episode of a new conflict, so long as either event is preceded by at least 2 years of uninterrupted civil peace. With the PRIO-NC measure, on the other hand, a new period of conflict is initiated whenever the country experiences the outbreak of a new conflict, regardless of the number of years of civil peace preceding this event.

4 Empirical Results

This section reveals our findings regarding the highly significant and robust reduced-form causal influence of genetic diversity on various intrastate conflict outcomes over the past half-century. We commence with the results of our cross-country regressions that explain the annual frequency of new civil conflict outbreaks in our sample period, first discussing the findings from our baseline analyses of both overall and ethnic civil conflict frequency and then presenting several robustness checks. We next discuss the results of our regressions explaining the incidence of either overall or ethnic civil conflict in repeated cross-country data, starting with our baseline findings for conflict incidence, followed by the results from robustness checks that account for alternative correlates of conflict incidence and consider alternative codings and types of conflict. After this, we reveal the findings from our analysis of the onset of either overall or ethnic civil conflict in repeated cross-country data, presenting both our baseline results from this analysis and some results establishing robustness to accounting for additional correlates of conflict onset.

We then substantively augment our body of evidence by establishing the reduced-form causal influence of genetic diversity on two less well-explored but nevertheless important dimensions of social conflict. Namely, we first examine the impact of genetic diversity on the intensive margin of conflict in society, exploiting variations in repeated cross-country data to establish genetic diversity as a contributing factor to the *severity* of social unrest, reflected by two alternative measures of the intensity of intrastate conflict. Second, consistently with priors based on the fact that our measure of diversity incorporates both intergroup and intragroup personal differences, we exploit cross-country variations to establish the contribution of genetic diversity to the incidence of *intragroup* factional conflicts in a national population, thereby demonstrating the importance of genetic diversity as a significant contributor not only to intergroup incompatibilities in society but also to diminished social cohesion within subnational groups.

The section concludes with an investigation of three mechanisms that can potentially mediate the reduced-form causal influence of genetic diversity on the various manifestations of intrastate conflict. Specifically, we provide some suggestive evidence in line with the hypothesis that our reduced-form results partly reflect (i) the contribution of genetic diversity to the degree of fractionalization and polarization across ethnic, linguistic, and religious groups in the national population; (ii) the adverse influence of genetic diversity on interpersonal trust and cooperation; and (iii) the association between genetic diversity and divergence in preferences for public goods and redistributive policies.

4.1 Analysis of Civil Conflict Frequency in Cross-Country Data

Before presenting the results from our baseline analysis of conflict frequency, we document by way of bivariate regressions in a sample of 154 countries, how the unconditioned influence of genetic diversity – i.e., on the annual frequency of new PRIO25 civil conflict outbreaks during the 1960–2008 time period – compares with the influence of each of other well-known diversity measures that

Table 1: Genetic Diversity vs. Other Diversity Measures in Explaining the Frequency of Overall Civil Conflict Onset across Countries – Bivariate Regressions

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
	Log number of new PRIO25 civil conflict onsets per year during the 1960–2008 time period					
Genetic diversity (ancestry adjusted)	0.212***					
	[0.076]					
Ethnic fractionalization (Alesina et al., 2003)		0.024***				
		[0.008]				
Linguistic fractionalization (Alesina et al., 2003)			0.032***			
			[0.009]			
Religious fractionalization (Alesina et al., 2003)				0.006		
				[0.009]		
Ethnolinguistic fractionalization (Desmet et al., 2012)					0.026***	
					[0.008]	
Ethnolinguistic polarization (Desmet et al., 2012)						0.007
						[0.009]
Effect of increasing diversity measure	0.014***	0.017***	0.025***	0.004	0.022***	0.005
from the 10th to the 90th percentile	[0.005]	[0.005]	[0.007]	[0.006]	[0.007]	[0.006]
Continent dummies	No	No	No	No	No	No
Sample	Global	Global	Global	Global	Global	Global
Observations	154	154	154	154	154	154
R^2	0.032	0.037	0.081	0.002	0.061	0.003

Notes: This table employs bivariate regressions to assess the unconditional importance of contemporary genetic diversity and each of other well-known diversity measures for explaining the cross-country variation in the annual frequency of new overall (PRIO25) civil conflict onsets during the 1960–2008 time period. In each regression, the estimated effect associated with increasing the corresponding diversity measure from the tenth to the ninetieth percentile of the measure's cross-country distribution is expressed in terms of the actual number of new conflict onsets per year. Robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

capture the degree of ethnolinguistic fragmentation of a national population.[23] As is evident from the results presented in Table 1, genetic diversity appears as a positive and statistically significant correlate of civil conflict frequency, although the cross-country variation in genetic diversity explains only about 3.2 percent of the cross-country variation in the temporal frequency of new civil conflict outbreaks. Specifically, the estimated coefficient suggests that a move from the 10th to the 90th percentile of the cross-country genetic diversity distribution is associated with an increase in conflict frequency by 0.014 new PRIO25 civil conflict outbreaks per year, a relationship that is statistically significant at the 1 percent level. Bearing in mind that the sample mean of the dependent variable is about 0.021 outbreaks per year, this association is also of sizable economic significance, reflecting 40.3 percent of a standard deviation across countries in the temporal frequency of new civil conflict onsets. In terms of the other diversity variables, the different measures of ethnic and linguistic fractionalization enter their respective bivariate regressions with positive and statistically significant coefficients, although in the absence of conditioning covariates, neither religious fractionalization nor ethnolinguistic polarization appear to be significantly associated with the temporal frequency of new civil conflict outbreaks.

In Table 2, we augment the preceding analysis by way of regressions that conduct explanatory "horse races" between genetic diversity, on the one hand, and various combinations of the measures of ethnolinguistic fragmentation, on the other. The first column simply reproduces the

[23]The sample employed by these regressions, as well as those presented in Table 2, is larger than our baseline sample of 143 countries due to the fact that the latter is conditioned on the availability of data on our baseline controls for geographical characteristics, institutional factors, and development outcomes.

Table 2: Genetic Diversity vs. Other Diversity Measures in Explaining the Frequency of Overall Civil Conflict Onset across Countries – "Horse race" Regressions

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS
	Log number of new PRIO25 civil conflict onsets per year during the 1960–2008 time period									
Genetic diversity (ancestry adjusted)	0.212***	0.178**	0.146**	0.209***	0.157**	0.207***	0.183**	0.158**	0.158**	0.160**
	[0.076]	[0.073]	[0.072]	[0.072]	[0.074]	[0.076]	[0.071]	[0.069]	[0.073]	[0.068]
Ethnic fractionalization (Alesina et al., 2003)		0.021***					0.021***			-0.004
		[0.007]					[0.007]			[0.014]
Linguistic fractionalization (Alesina et al., 2003)			0.029***					0.031***		0.029*
			[0.009]					[0.009]		[0.016]
Religious fractionalization (Alesina et al., 2003)				0.001			-0.003	-0.009		-0.010
				[0.009]			[0.009]	[0.009]		[0.010]
Ethnolinguistic fractionalization (Desmet et al., 2012)					0.023***				0.026**	0.007
					[0.008]				[0.011]	[0.023]
Ethnolinguistic polarization (Desmet et al., 2012)						0.005			-0.008	-0.006
						[0.008]			[0.012]	[0.015]
Effect of increasing genetic diversity	0.014***	0.011**	0.009**	0.013***	0.010**	0.013***	0.012***	0.010**	0.010**	0.010**
from the 10^{th} to the 90^{th} percentile	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.004]	[0.005]	[0.004]
Continent dummies	No	No	No	No	No	No	No	No	No	No
Sample	Global	Global	Global	Global	Global	Global	Global	Global	Global	Global
Observations	154	154	154	154	154	154	154	154	154	154
Adjusted R^2	0.026	0.046	0.084	0.019	0.066	0.021	0.041	0.082	0.063	0.066

Notes: This table employs regressions that run "horse races" between contemporary genetic diversity and other well-known diversity measures to assess their relative importance for explaining the cross-country variation in the annual frequency of new overall (PRIO25) civil conflict onsets during the 1960–2008 time period, establishing the robustness of genetic diversity over other diversity measures as a predictor of conflict frequency. The estimated effect associated with increasing genetic diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the actual number of new conflict onsets per year. Robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

unconditioned relationship between genetic diversity and conflict frequency, but in subsequent columns, we introduce the other diversity measures to the specification as controls; initially, one at a time, then in pairs, and finally, all at the same time. The results indicate that the positive and statistically significant association of genetic diversity with conflict frequency does not vanish even after its potential influence through the degree of ethnic fragmentation is accounted for by the analysis. Notably, the coefficient on genetic diversity remains qualitatively robust throughout this analysis, although its statistical significance drops to the 5 percent level – primarily due to a modest decrease in its point estimate – when conditioned on the measures of ethnic and linguistic fractionalization. This provides suggestive evidence that certain dimensions of ethnolinguistic fragmentation could be capturing an important though not the only proximate channel that potentially mediates the deeper influence of genetic diversity on the propensity for conflict in society.[24]

4.1.1 Baseline Analysis of PRIO25 Civil Conflict Frequency

Table 3 presents the results from our baseline cross-country analysis of the influence of genetic diversity on the annual frequency of new PRIO25 civil conflict onsets during the 1960–2008 time horizon. In Column 1, we replicate the previously revealed bivariate relationship, but we do so in our baseline sample of 143 countries, in order to provide an appropriate benchmark for the

[24] The results from a more systematic investigation of this and two other potential mechanisms are presented in Table 11 and will be discussed in Section 4.6.

subsequent regressions. In particular, beginning with Column 2, our analysis progressively includes an expanding set of covariates to the specification – first incorporating exogenous geographical conditions and then additionally accounting for semi-endogenous institutional factors, before including the more endogenous outcomes of economic development – until our full empirical model is attained in Column 8. In what follows, we reveal the stability characteristics of our coefficient of interest when it is subjected to a successively larger set of covariates, providing a discussion along the way of the theoretical basis underlying our choice of the various control variables.

Accounting for Geographical Characteristics The regression in Column 2 conditions the analysis on our baseline set of geographical covariates (excluding continent fixed effects), each of which may be correlated a priori with prehistoric migratory distance from East Africa and may also confer an influence on conflict propensity through mechanisms that have little to do with genetic diversity. Geographical characteristics like absolute latitude and distance to the nearest waterway, for instance, can exert an influence on economic development and, thus, on conflict potential through climatological, institutional, and trade-related mechanisms. In addition, the total land area of country may contribute to its propensity for civil conflict due to the potentially positive association between land area, on the one hand, and either the size or the diversity of the national population, on the other.[25]

A country's topographical features can also affect its proclivity for conflict, primarily because rugged terrains can provide safe havens for rebels and enable them to sustain continued resistance by protecting them from numerically and militarily superior government forces (Fearon and Laitin, 2003). Beyond that, by geographically isolating subgroups of a regional population, terrain ruggedness could have facilitated the forces of "cultural drift" and ethnic differentiation among these groups (Michalopoulos, 2012), thereby making the population more prone to conflict over time. Finally, in light of evidence that conditional on their respective country-level means, greater intracountry dispersion in agricultural land suitability and elevation can contribute to ethnolinguistic diversity (Michalopoulos, 2012), these natural attributes could also impart an indirect influence on conflict propensity through the ethnolinguistic fragmentation of the population.[26]

According to the regression in Column 2, accounting for the potentially confounding influence of all the aforementioned geographical conditions actually increases the magnitude of the coefficient on genetic diversity, relative to the estimate in Column 1. Indeed, the influence of genetic diversity continues to remain statistically significant at the 1 percent level, but now, its point estimate is more than twice as large as before. This sizable increase in the estimated influence of genetic diversity appears to be largely driven by the inclusion of absolute latitude and the range of agricultural land suitability as covariates to the model, given that both variables enter the

[25]Total land area can also account for any bias that might result from the possibility that our measure of genetic diversity, by virtue of being based on migratory distance from East Africa to the modern-day capital city of a country, may be less comparable across countries of different geographical size.

[26]Although we directly control for measures of ethnolinguistic fragmentation in our full empirical model, those measures are afflicted endogeneity bias, and beyond that, their exogenous geographical determinants may still explain some unobserved component of intrapopulation heterogeneity in ethnic and cultural traits, thereby exerting some residual influence on the potential for conflict in society.

Table 3: Genetic Diversity and the Frequency of Overall Civil Conflict Onset across Countries – The Baseline Analysis

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) 2SLS	(12) 2SLS
	Log number of new PRIO25 civil conflict onsets per year during the 1960–2008 time period											
Genetic diversity (ancestry adjusted)	0.182**	0.422***	0.322**	0.366**	0.350**	0.390**	0.377**	0.398**	0.639**	0.855**	0.599***	0.805***
	[0.077]	[0.123]	[0.150]	[0.171]	[0.168]	[0.179]	[0.184]	[0.183]	[0.257]	[0.333]	[0.231]	[0.275]
Ethnic fractionalization					0.011		0.006	0.007		0.012		-0.002
					[0.012]		[0.012]	[0.012]		[0.014]		[0.013]
Ethnolinguistic polarization						0.013	0.010	0.010		0.007		0.019
						[0.013]	[0.014]	[0.013]		[0.015]		[0.013]
Absolute latitude		-0.404***	-0.440*	-0.331	-0.225	-0.356	-0.292	0.149	-0.333	0.255	-0.529**	-0.116
		[0.119]	[0.255]	[0.263]	[0.320]	[0.256]	[0.307]	[0.287]	[0.301]	[0.356]	[0.243]	[0.296]
Land area		0.765	1.825	1.709	1.972	1.719	1.862	1.586	4.177	4.114	1.626	1.311
		[2.119]	[2.287]	[2.358]	[2.382]	[2.403]	[2.436]	[2.675]	[2.797]	[2.743]	[2.247]	[2.631]
Ruggedness		0.038	0.028	0.030	0.036	0.032	0.035	0.056	0.041	0.080	0.034	0.054
		[0.038]	[0.044]	[0.044]	[0.047]	[0.045]	[0.047]	[0.047]	[0.053]	[0.054]	[0.043]	[0.042]
Mean elevation		-0.016*	-0.015	-0.017*	-0.018*	-0.018*	-0.018*	-0.020**	-0.019	-0.025**	-0.016*	-0.023**
		[0.009]	[0.009]	[0.010]	[0.010]	[0.011]	[0.011]	[0.010]	[0.012]	[0.012]	[0.009]	[0.009]
Range of elevation		0.009**	0.009**	0.009**	0.008*	0.008*	0.008*	0.004	0.009*	0.003	0.010**	0.005
		[0.005]	[0.005]	[0.004]	[0.004]	[0.004]	[0.005]	[0.004]	[0.006]	[0.005]	[0.004]	[0.004]
Mean land suitability		0.013	0.018	0.016	0.019	0.019	0.020	0.006	0.018	0.006	0.018	0.003
		[0.012]	[0.013]	[0.015]	[0.014]	[0.015]	[0.014]	[0.016]	[0.015]	[0.020]	[0.013]	[0.015]
Range of land suitability		0.013	0.014	0.012	0.011	0.014	0.013	0.010	0.019	0.007	0.017	0.015
		[0.008]	[0.011]	[0.012]	[0.012]	[0.014]	[0.014]	[0.015]	[0.013]	[0.016]	[0.012]	[0.014]
Distance to nearest waterway		0.008	0.005	0.007	0.006	0.007	0.007	0.002	0.000	-0.001	0.004	0.001
		[0.009]	[0.010]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.011]	[0.013]	[0.009]	[0.010]
Executive constraints, 1960–2008 average				0.004	0.004	0.004	0.004	0.006*		0.005		0.008**
				[0.003]	[0.003]	[0.003]	[0.003]	[0.003]		[0.004]		[0.004]
Fraction of years under democracy, 1960–2008				-0.015	-0.014	-0.015	-0.014	-0.012		-0.002		-0.017
				[0.019]	[0.019]	[0.019]	[0.019]	[0.018]		[0.019]		[0.017]
Fraction of years under autocracy, 1960–2008				-0.006	-0.005	-0.005	-0.005	-0.008		-0.009		-0.007
				[0.017]	[0.016]	[0.017]	[0.016]	[0.016]		[0.017]		[0.015]
Log oil production per capita, 1960–2008 average								0.002**		0.002*		0.002*
								[0.001]		[0.001]		[0.001]
Log population, 1960–2008 average								0.003		0.004		0.003
								[0.003]		[0.003]		[0.003]
Log GDP per capita, 1960–2008 average								-0.015***		-0.016***		-0.016***
								[0.005]		[0.005]		[0.004]
Effect of increasing genetic diversity from the 10th to the 90th percentile	0.012**	0.027***	0.021**	0.024**	0.023**	0.025**	0.025**	0.026**	0.032**	0.042**	0.039***	0.052***
	[0.005]	[0.008]	[0.010]	[0.011]	[0.011]	[0.012]	[0.012]	[0.012]	[0.013]	[0.016]	[0.015]	[0.018]
Continent dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Colonial history dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Sample	Global	Global	Global	Global	Global	Global	Global	Global	Old World	Old World	Global	Global
Observations	143	143	143	143	143	143	143	143	119	119	143	143
Partial R^2 of genetic diversity	–	0.117	0.046	0.056	0.051	0.063	0.056	0.066	0.094	0.141	–	–
Partial R^2 sum of other diversity measures	–	–	–	–	0.006	0.010	0.007	0.007	–	0.009	–	–
First-stage adjusted R^2	–	–	–	–	–	–	–	–	–	–	0.755	0.759
First-stage partial R^2 of migratory distance	–	–	–	–	–	–	–	–	–	–	0.478	0.438
First-stage F statistic	–	–	–	–	–	–	–	–	–	–	211.910	103.087
Adjusted R^2	0.019	0.190	0.197	0.192	0.190	0.193	0.188	0.235	0.256	0.308	–	–

Notes: This table exploits cross-country variations to establish a significant positive reduced-form impact of contemporary genetic diversity on the annual frequency of new overall (PRIO25) civil conflict onsets during the 1960–2008 time period, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict. For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of colonial history dummies includes indicators for whether a country was ever a colony of the U.K., France, and any other major colonizing power. The 2SLS regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated effect associated with increasing genetic diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the actual number of new conflict onsets per year. Robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

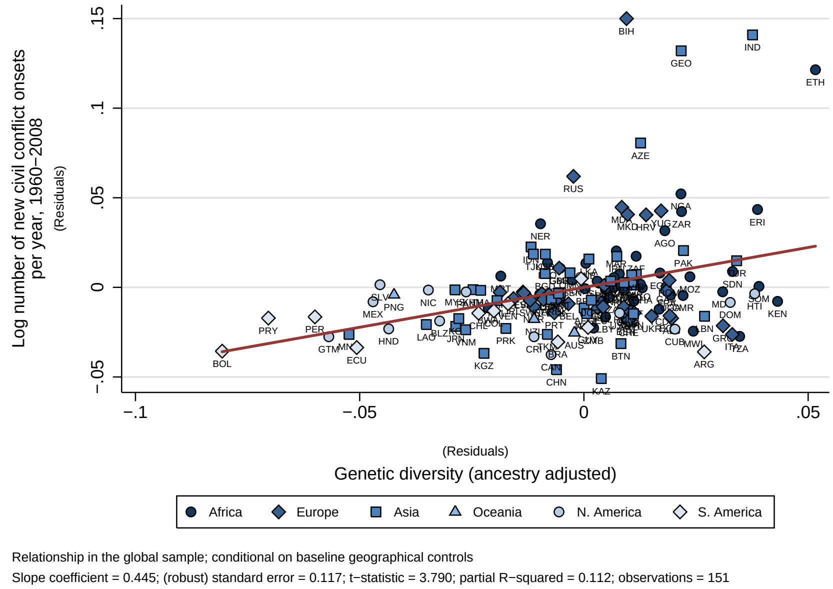
regression significantly and with their expected signs. Specifically, countries located farther from the Equator have seen fewer conflict outbreaks on average, while those with greater dispersion in their respective land endowments have experienced such outbreaks more frequently, a result that plausibly reflects the conflict-promoting role of ethnolinguistic fragmentation, following the rationale provided by the findings of Michalopoulos (2012). The scatter plots in Figure 1 depict the positive and statistically significant cross-country relationship, conditional on our baseline set of geographical covariates considered by the regression in Column 2 of Table 3, between genetic diversity and the annual frequency of new PRIO25 civil conflict onsets, both in our full sample of countries (Panel A) and in a sample that omits apparently influential outliers (Panel B).

In the course of the prehistoric demic expansion of humans from East Africa to the rest of the world, the occurrence of ancient population bottlenecks at specific intercontinental "crossings" – a classic example being the Bering Land Bridge – led to discrete spatial differences in observed genetic diversity across either side of these isthmuses. The global spatial distribution of genetic diversity therefore exhibits what population geneticists refer to as punctuated "clines," raising the possibility that the cross-continental variation in genetic diversity may well be correlated with cross-continental variation in unobserved – or observed but imprecisely measured – continent-specific characteristics. Indeed, as is apparent from the regression in Column 3, the point estimate of the influence of genetic diversity on civil conflict frequency becomes somewhat diminished in both magnitude and statistical precision, once we condition the specification to only exploit intracontinental variations by augmenting our baseline set of geographical covariates with a complete set of continent dummies. Nevertheless, our coefficient of interest continues to remain notably larger than its unconditioned estimate from Column 1, and it suggests that a move from the 10th to the 90th percentile of the cross-country genetic diversity distribution is associated with an increase in conflict frequency by 0.021 new PRIO25 civil conflict outbreaks per year (or, equivalently, 65.6 percent of a standard deviation in the cross-country conflict frequency distribution), a relationship that is statistically significant at the 5 percent level.

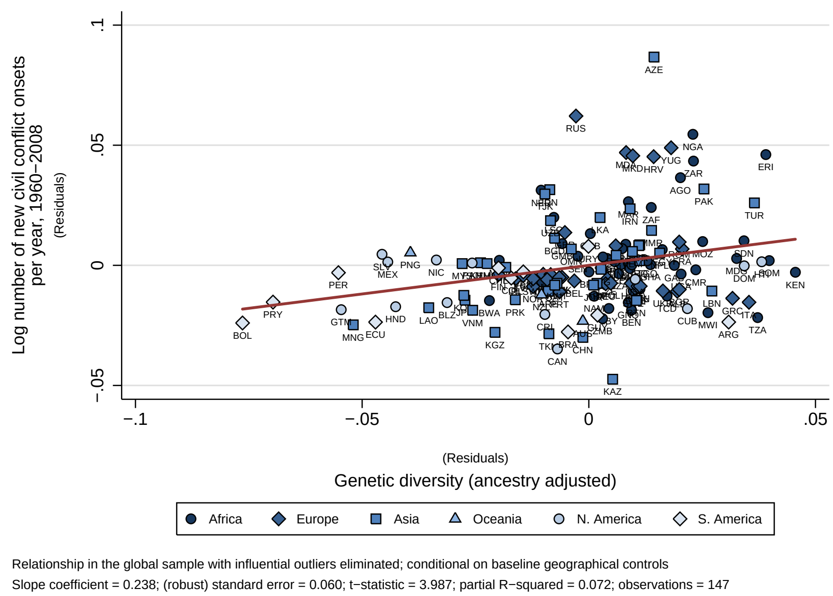
Accounting for Institutional Factors The specification examined in Column 4 additionally incorporates our baseline set of institutional covariates to account for the potentially confounding influence of various institutional factors. It is well-accepted, for instance, that colonial legacies may have significantly shaped the political economy of interethnic cleavages in newly independent states (Posner, 2003). More generally, the heritage of colonial rule and the identity of the former colonizer may have important ramifications for the nature and stability of contemporary political institutions at the national level, thereby influencing the potential for conflict in society (see, e.g., Michalopoulos and Papaioannou, 2013). We therefore condition our analysis in Column 4 on a set of indicators for the historical prevalence of colonial rule by the U.K., France, and any other major colonizing power, along with fixed effects for British and French legal origins that account for any residual influence of the legal code inherited by a country from the colonial period.[27]

In addition, although many studies find the prevalence of democratic institutions to be an insignificant predictor of the risk of civil conflict (e.g., Fearon and Laitin, 2003; Collier and Hoeffler, 2004), some point to a negative association between conflict potential and democracy (e.g., Esty et al., 1998; Gurr, 2000). There is also evidence that anocratic or hybrid political regimes can be even more susceptible to conflict than pure autocracies (Fearon and Laitin, 2003), consistently with the notion that weaknesses in political legitimacy – and the political grievances associated with

[27]Throughout the presentation of our results, in the interest of conserving space, we refrain from reporting the coefficients associated with our baseline controls for colonial history, legal origins, and continent fixed effects, and in our regressions based on repeated cross-country data, we additionally refrain from reporting the coefficients associated with time dummies or any other controls for duration dependence.



(A) Relationship in the full sample



(B) Relationship with outliers omitted from the sample

Figure 1: Genetic Diversity and the Frequency of Overall Civil Conflict Onset across Countries in the Global Sample

Notes: This figure depicts the global cross-country relationship between contemporary genetic diversity and the annual frequency of new overall (PRIO25) civil conflict onsets during the 1960–2008 time period in (i) an unrestricted sample [Panel A]; and (ii) a sample without influential outliers [Panel B], conditional on the baseline geographical correlates of conflict, as considered by the analysis in Column 2 of Table 3. Each panel presents an added-variable plot with a partial regression line. Given that the sample employed by the analysis in Panel A is not restricted by the availability of data on the additional covariates considered by the analysis in Table 3, the regression coefficient in Panel A is marginally different from that presented in Column 2 of Table 3. The influential outliers that are omitted from the sample in Panel B include Bosnia and Herzegovina (BIH), Ethiopia (ETH), Georgia (GEO), and India (IND).

them – can generate violent dissensions, particularly when the state apparatus is less repressive than in a fully fledged autocratic regime. To account for the possibility of such effects imparted by the type of political regime, the regression in Column 4 further incorporates controls for the temporal prevalence of both democratic and autocratic political regimes during the sample period, with the temporal prevalence of anocracy serving as the omitted category. The specification also includes the temporal mean of the degree of institutionalized constraints on the discretionary power of the chief executive as an additional control for the influence of the quality of contemporary national institutions on the risk of civil conflict.

As is evident from the results presented in Column 4, however, controlling for the potentially confounding impact of institutional factors does not significantly affect the stability of the coefficient on genetic diversity. In particular, the point estimate of the reduced-form influence of genetic diversity on conflict frequency experiences a very modest increase in magnitude, relative to the estimate presented in Column 3, and it remains statistically significant at the 5 percent level. This finding is especially reassuring in light of the fact that the coefficients associated with some of the institutional covariates are likely to be afflicted by endogeneity bias.

Accounting for Ethnolinguistic Fragmentation In Columns 5–7, we assess the extent to which the intergroup fragmentation of a national population, as reflected by the well-known indices of ethnic fractionalization and ethnolinguistic polarization, can explain away the reduced-form influence of genetic diversity on conflict frequency, conditional on the set of geographical and institutional covariates that are already considered by our analysis.

Whether and how the risk of civil conflict is related to ethnic fractionalization at the national level has been a topic of much debate in the literature. Remarkably, previous empirical findings regarding the role of ethnic fractionalization have generally been somewhat mixed, exhibiting substantial sensitivity to model specifications and conflict codings (Fearon and Laitin, 2003; Collier and Hoeffler, 2004; Hegre and Sambanis, 2006). Moreover, theoretical work on the link between the ethnic composition of a society and the risk of civil conflict suggests that ethnic fractionalization by itself may be insufficient to fully capture the conflict potential that can be attributed to broader ethnolinguistic configurations of the population (Esteban and Ray, 2008a, 2011b). In light of their well-grounded structural foundations, indices of polarization have gained popularity as a substitute for – or in addition to – the fractionalization measures commonly considered by empirical analyses of civil conflict.[28] Indeed, many empirical studies find that ethnic polarization possesses stronger explanatory power for the likelihood of civil conflict (e.g., Montalvo and Reynal-Querol, 2005; Esteban, Mayoral and Ray, 2012).[29]

[28] Esteban and Ray (1994) provide the first serious attempt to measure polarization, derive its theoretical properties, and highlight its role in contributing to the potential for social conflict. In addition, Esteban and Ray (1999) formally demonstrate that the likelihood of social conflict is maximized when preferred societal outcomes are characterized by a two-point symmetric distribution – a situation that corresponds to the highest degree of polarization.

[29] Although some earlier studies (e.g., Collier and Hoeffler, 1998; Elbadawi and Sambanis, 2002) documented a statistically significant nonmonotonic association between ethnic fractionalization and conflict potential, it is now well-known that fractionalization is also nonmonotonically related with polarization in the cross-country data, and as such, one cannot conclusively distinguish the nonmonotonic influence of fractionalization from the monotonic

Recent evidence uncovered by Ashraf and Galor (2013b) establishes genetic diversity as an underlying cause of various existing manifestations of ethnolinguistic fragmentation at the national level, demonstrating a positive influence of genetic diversity on several measures of ethnolinguistic fractionalization and polarization, including those reflecting more ancestral intergroup cleavages. These findings suggest that part of the reduced-form impact of genetic diversity on conflict potential in the modern era could be operating through its deeper role in shaping the broader ethnolinguistic configurations of a national population – namely, the influence of genetic diversity on the prehistoric formation and subsequent ethnic differentiation of coalitional groups at a given location in the distant past. Nevertheless, the residual variation in genetic diversity that is not manifested in measures of contemporary ethnolinguistic fragmentation could continue to play a role in explaining conflict potential through channels associated with interpersonal trust, heterogeneity in preferences for public goods and redistributive policies, and economic inequality.

The regressions in Columns 5–7 indicate that when additionally subjected to controls for ethnic fractionalization and ethnolinguistic polarization, either individually or jointly, the point estimate of the coefficient on genetic diversity, conditional on our baseline set of geographical and institutional covariates, continues to remain largely stable in both magnitude and statistical precision.[30] In contrast, neither ethnic fractionalization nor ethnolinguistic polarization appear to possess any significant explanatory power for the cross-country variation in the temporal frequency of civil conflict outbreaks, conditional on genetic diversity and our baseline set of geographical and institutional covariates. Specifically, the partial R^2 statistics associated with the regression in Column 7 suggest that while the residual cross-country variation in genetic diversity can explain 5.6 percent of the residual cross-country variation in conflict frequency, only 0.7 percent of the residual cross-country variation in conflict frequency is explained by the residual cross-country variations in the two measures of ethnolinguistic fragmentation.

Accounting for Development Outcomes The regression in Column 8 further augments the analysis by incorporating our controls for the potentially confounding influence of oil revenues, population size, and income per capita, thus attaining the specification corresponding to our full empirical model of the temporal frequency of new PRIO25 civil conflict outbreaks. Indeed, as argued by many scholars in the literature, higher revenues from natural resources – like oil, amongst others – can foster the risk of civil conflict through various mechanisms, including (i) the weakening of state institutions and the increased attractiveness of the state as a target for rebel groups (e.g., Fearon and Laitin, 2003; Fearon, 2005; Dube and Vargas, 2013); (ii) the provision of easier financing for rebel organizations (e.g., Collier and Hoeffler, 2004; Angrist and Kugler, 2008); (iii) the greater

influence of polarization on the risk of civil conflict. By restricting both fractionalization and polarization measures to enter our regressions linearly, our baseline approach follows Esteban, Mayoral and Ray (2012), but we nevertheless checked the robustness of our main finding to employing alternative specifications that allow for both a linear and a quadratic term in ethnic fractionalization, and we found qualitatively similar results (not reported).

[30]Table A.3 in Appendix A repeats our baseline analysis from Table 3, replacing the ethnic fractionalization index of Alesina et al. (2003) with the closely related linguistic fractionalization index from the same study, on the grounds that the latter variable also entered our bivariate and "horse race" regressions – presented in Tables 1 and 2 – with a statistically significant coefficient. Reassuringly, as is evident from the results presented in Table A.3, all our baseline findings remain qualitatively intact under this robustness check.

vulnerability of interest groups to terms-of-trade shocks (e.g., Humphreys, 2005); and (iv) the increased sovereignty value of resource-rich areas as perceived by separatist movements (e.g., Ross, 2006). We therefore condition our full specification in Column 8 on the temporal mean of the annual value of oil production per capita over the entirety of our sample period.

Given that the battle-related death threshold, implicit in the definition of a civil conflict event, is not corrected for total population size, most empirical studies of civil conflict account for the size of a country's population, on the grounds that the extensive margin of violence could be mechanically affected by total population size. In addition, a larger population may imply a larger recruitment pool for rebels (Fearon and Laitin, 2003), and to the extent that it is associated with greater intrapopulation heterogeneity, a larger population could also be associated with stronger motives for secessionist conflicts (Collier and Hoeffler, 1998; Alesina and Spolaore, 2003; Desmet et al., 2011). Our regression in Column 8 therefore additionally includes the temporal mean of annual population size over our sample period as a covariate.

Finally, motivated by several arguments proposed in the literature on civil conflict, we also incorporate the temporal mean of the annual level of GDP per capita as a covariate in our full empirical model in Column 8. One argument, due to Grossman (1991) and Hirshleifer (1995), is that higher per-capita incomes raise the opportunity cost for potential rebels to engage in insurrections, thus predicting an inverse empirical relationship between the level or growth of income, on the one hand, and the risk of civil conflict, on the other (Collier and Hoeffler, 1998, 2004; Miguel, Satyanath and Sergenti, 2004; Brückner and Ciccone, 2010). Another idea, due to Hirshleifer (1991) and Grossman (1999), is that by raising the return to predation, higher per-capita incomes can contribute to the risk of rapacious activities over society's resources, consistently with empirical findings from some of the aforementioned studies on the link between income from natural resources and conflict potential. Furthermore, to the extent that income per capita serves as a proxy for state capabilities (Fearon and Laitin, 2003), a higher level of per-capita income can reflect the notion of a state that is better able to prevent or defend itself against rebel insurgencies, an idea that has also found some recent empirical support (e.g., Bazzi and Blattman, 2014).

As noted by several authors, however, per-capita oil revenues, population size, and GDP per capita are all expected to be endogenous in an empirical model of civil conflict (e.g., Ross, 2006; Brückner, 2010; Blattman and Miguel, 2010), so the estimated coefficients associated with these covariates in our model do not reflect causal effects. Nevertheless, their inclusion in our model permits us to assess the extent to which the reduced-form influence of genetic diversity on the risk of civil conflict can be attributed to some of the aforementioned channels that are proxied – to a greater or lesser extent – by measures that are correlated with economic development. This is additionally important in light of the previously established robust and significant hump-shaped influence of genetic diversity on productivity Ashraf and Galor (2013a), because we are ultimately interested in uncovering the reduced-form impact of genetic diversity on conflict potential, independently of its indirect influence through the level of economic development.

The regression in Column 8 reassuringly indicates that the inclusion of controls for oil production per capita, population size, and GDP per capita to our specification hardly sways the point estimate of the coefficient on genetic diversity, which remains remarkably stable in both magnitude and statistical significance when compared to the estimates from previous columns. In particular, our coefficient of interest from this regression suggests that conditional on our complete set of controls for geographical characteristics, institutional factors, ethnolinguistic fragmentation, and the correlates of economic development, a move from the 10th to the 90th percentile of the cross-country genetic diversity distribution is associated with an increase in conflict frequency by 0.026 new PRIO25 civil conflict outbreaks per year (or 81.2 percent of a standard deviation in the cross-country conflict frequency distribution, which is comparable to a move from the 50th to the 90th percentile of this distribution). Moreover, the adjusted R^2 statistic of the regression suggests that our baseline empirical model explains 23.5 percent of the cross-country variation in conflict frequency, whereas the partial R^2 statistic associated with genetic diversity indicates that the residual cross-country variation in genetic diversity can explain 6.6 percent of the residual cross-country variation in conflict frequency.

Our results thus far demonstrate a significant and robust cross-country association between genetic diversity and the temporal frequency of civil conflict onsets over the last half-century, even after conditioning the analysis on a sizable set of controls for geographical characteristics, institutional factors, ethnolinguistic fragmentation, and development outcomes. Nevertheless, this association could be marred by endogeneity bias, in light of the possibility that the large-scale human migrations of the post-1500 era – incorporated into our ancestry-adjusted measure of genetic diversity for contemporary national populations – and the spatial pattern of conflicts in the modern era could be codetermined by common unobserved forces (e.g., the spatial pattern of *historical* conflicts) that may not be fully captured by our control variables. As discussed previously in Section 3.2, we exploit two alternative identification strategies to address this issue, but before proceeding to the results from those analyses, we first assess the extent to which our estimates thus far are likely to be biased by selection on unobservables.

Selection on Observables and Unobservables Following the method developed by Altonji, Elder and Taber (2005), we exploit the idea that the amount of selection on the unobserved variables in a model can be inferred from the amount of selection on the observed explanatory variables, thus permitting an assessment of how much larger the selectivity bias from unobserved heterogeneity needs to be, relative to the bias from selection on observables, in order to fully explain away the coefficient on our explanatory variable of interest.[31] Specifically, we compare the estimated coefficient, $\hat{\beta}_1^R$, on genetic diversity from any one of our restricted models in Columns 1–7 with

[31] Altonji, Elder and Taber (2005) develop this method for the case where the explanatory variable of interest is binary in nature, while Bellows and Miguel (2009) consider the case of a continuous explanatory variable. Roughly speaking, the assumption underlying assessments of this type is that the covariation of the outcome variable with observables, on the one hand, and its covariation with unobservables, on the other, are identically related to the explanatory variable of interest. Altonji, Elder and Taber (2005) provide some sufficient conditions for such an assumption to hold.

its estimated coefficient, $\hat{\beta}_1^F$, from our full empirical model in Column 8, examining the absolute magnitude of the ratio, $\hat{\beta}_1^F/(\hat{\beta}_1^R - \hat{\beta}_1^F)$. Intuitively, a higher absolute value for this ratio implies that the additional control variables included in the full model, relative to the restricted one, are not sufficient to explain away the estimated coefficient on genetic diversity in the full specification, and as such, this coefficient cannot be completely attributed to omitted-variable bias unless the amount of selection on unobservables is much larger than that on observables.

In the interest of brevity, we focus our analysis here on two different restricted models; the first one being the specification from Column 2 that only includes exogenous geographical covariates, and the second one being the specification from Column 3 that additionally accounts for continent fixed effects. When comparing the estimated coefficient on genetic diversity from each of these regressions with that from our full specification in Column 8, the resulting ratios of relevance are 16 and 5, respectively. These numbers suggest that selection on unobservables would have to be at least five times larger than selection on observables, and on average, over ten times larger, in order for our estimated coefficient of interest in Column 8 to be entirely attributable to selection on unobservables. The results from this analysis therefore support our view that it is rather unlikely for our baseline estimate of the influence of genetic diversity on conflict frequency to be explained away by unobserved heterogeneity.

Addressing Endogeneity We now present our findings that reveal the reduced-form causal influence of genetic diversity on the temporal frequency of civil conflict outbreaks in the post-1960 time horizon, exploiting our two alternative identification strategies for addressing the potential endogeneity of our ancestry-adjusted measure of genetic diversity in a globally representative sample of countries. For each of our two identification strategies, we estimate two different specifications; one corresponding to the model in Column 3 that conditions the analysis on only exogenous geographical covariates (including continent fixed effects), and the other corresponding to the model in Column 8 that conditions the analysis on our full set of controls for geographical conditions, institutional factors, ethnolinguistic fragmentation, and development outcomes.

In Columns 9 and 10, we implement our first approach to causal identification by simply restricting the OLS estimator to exploit variations in a subsample of countries that only belong to the Old World. This strategy exploits the fact that the great human migrations of the post-1500 era had systematically differential impacts on the genetic composition of national populations in the Old World versus the New World. Specifically, although post-1500 population flows had a dramatic effect on the genetic diversity of national populations in the Americas and Oceania, the diversity of resident populations in Africa, Europe, and Asia remained largely unaltered, primarily because native populations in the Old World were not subjected to substantial inflows of migrant settlers that were descended from genetically distant ancestral populations. As such, by confining our analysis to the Old World, the spatial variation in contemporary genetic diversity that we exploit effectively reflects the variation in genetic diversity across prehistorically indigenous populations, overwhelmingly determined by an ancient serial founder effect associated with the "out of Africa" migration process.

We implement our second approach to causal identification in Columns 11 and 12, exploiting variations in our globally representative sample of countries with a 2SLS estimator that employs the migratory distance of a country's prehistorically native population from East Africa as an instrument for the country's contemporary genetic diversity. This strategy exploits the fact that the mark of ancient population bottlenecks that occurred during the prehistoric "out of Africa" demic diffusion of humans across the globe continues to be seen on average in the worldwide pattern of genetic diversity across contemporary national populations – a fact reflected by the sizable correlation of 0.750 between the measures of precolonial and contemporary genetic diversity in our global sample of countries. In addition, this strategy rests on the identifying assumption that the migratory distance of a country's prehistorically indigenous population from East Africa is plausibly excludable from an empirical model of the risk of civil conflict faced by its modern national population, conditional on our large set of controls for the geographical and institutional determinants of conflict as well as the correlates of economic development.

As is apparent from the regressions in Columns 9–12, comparing specifications with the same set of covariates, the two alternative identification strategies yield remarkably similar results, with the point estimate of the coefficient on genetic diversity being noticeably larger in magnitude, relative to its less well-identified counterpart (in either Column 3 or Column 8), based on an OLS regression in our global sample of countries. Specifically, the findings from exploiting our first strategy in Columns 9 and 10 suggest that a move from the 10th to the 90th percentile of the cross-country genetic diversity distribution in the Old World leads to an increase in conflict frequency by 0.032 to 0.042 new PRIO25 civil conflict outbreaks per year (or 91.9 to 123 percent of a standard deviation in the cross-country conflict frequency distribution in the Old World), based on estimates that are statistically significant at the 5 percent level. Similarly, the findings from exploiting our second strategy in Columns 11 and 12 suggest that a move from the 10th to the 90th percentile of the global cross-country genetic diversity distribution leads to an increase in conflict frequency by 0.039 to 0.052 new PRIO25 civil conflict outbreaks per year (or 122 to 164 percent of a standard deviation in the global cross-country conflict frequency distribution), reflecting estimates that are statistically significant at the 1 percent level.

In our view, there are three distinct rationales – perhaps operating in tandem – for why our better-identified point estimates of the coefficient on genetic diversity are larger than their less well-identified counterparts. First, to the extent that the spatial pattern of social conflict exhibits long-run persistence for reasons other than genetic diversity, the emigrations and atrocities spurred by unobserved historical conflicts in the past 500 years may on average have had a homogenizing influence in historically conflict-prone populations (Fletcher and Iyigun, 2010), thereby leading to a downward bias in the estimated coefficient on genetic diversity in an OLS regression that explains the global variation in civil conflict potential in the modern era.

A second plausible explanation is that the pattern of conflict risk in the modern era, especially across populations in the New World that experienced a substantial increase in diversity from migrations in the post-1500 era, has been influenced not so much by the higher genetic diversity of the immigrants but more so by the unobserved (or observed but noisily measured) human capital that European settlers brought with them, the colonization strategies that they pursued, and the sociopolitical institutions that they established. To the extent that these unobserved factors associated with European settlers in the New World served, in one way or another, to reduce the risk of social conflict in the modern national populations of the Americas and Oceania, they could also introduce a negative bias in the OLS-estimated relationship between genetic diversity and conflict risk in a global sample of countries.

A third possible rationale is that in the end, the genetic diversity that really matters for the conflict propensity of a population is its prehistorically determined component that may have contributed to the formation and ethnic differentiation of *native* groups in a given location and, thus, to more deeply rooted interethnic divisions amongst these groups. As such, conditional on continent fixed effects that absorb any systematic differences in the pattern of post-1500 population flows into locations in the Old World versus the New World, our ancestry-adjusted measure of genetic diversity (that incorporates the diversity of both native and nonnative groups in a contemporary national population) may simply be operating as a noisy proxy for the "true" measure of prehistorically determined genetic diversity that matters for conflict potential, implying that the estimated influence of the ancestry-adjusted measure of genetic diversity ends up being attenuated in an OLS regression that exploits worldwide variations.

Given that both of our identification strategies ultimately exploit the variation in genetic diversity across populations that have been prehistorically indigenous to their current locations, either by omitting the modern national populations of the New World from the estimation sample or by instrumenting contemporary genetic diversity in a globally representative sample of countries with the prehistoric migratory distance of a country's geographical location from East Africa, our better-identified estimates mitigate all the aforementioned sources of negative bias in the influence of genetic diversity on conflict potential in contemporary national populations.

4.1.2 Baseline Analysis of WCM09 Ethnic Civil Conflict Frequency

As a phenomenon, ethnic civil conflict broadly refers to that type of intrastate conflict where state forces fight against armed opposition groups that represent ethnic or religious minorities with ethnonationalist demands and/or ethnically motivated concerns.[32] Although not all modern civil conflicts occur across interethnic divisions, a significant fraction are indeed considered to be ethnic civil conflicts (Horowitz, 1985; Brubaker and Laitin, 1998). Nonetheless, with some notable exceptions (e.g., Sambanis, 2001; Fearon and Laitin, 2003; Wimmer, Cederman and Min, 2009), the majority of empirical work on the causes of civil conflict – including studies that primarily focus on the role of interethnic divisions in society – do not explicitly distinguish between conflicts that

[32] Classic examples include some of the long-standing internal conflicts in Myanmar, fought between a state that is largely controlled by members of the Bamar ethnic majority group, on the one hand, and rebels representing either the Karen or the Kachin ethnic minority groups, on the other. Another well-known example is the Rwandan Civil War, fought between the Hutu-led government and the Rwandan Patriotic Front (RPF), a rebel group primarily composed of Tutsi refugees.

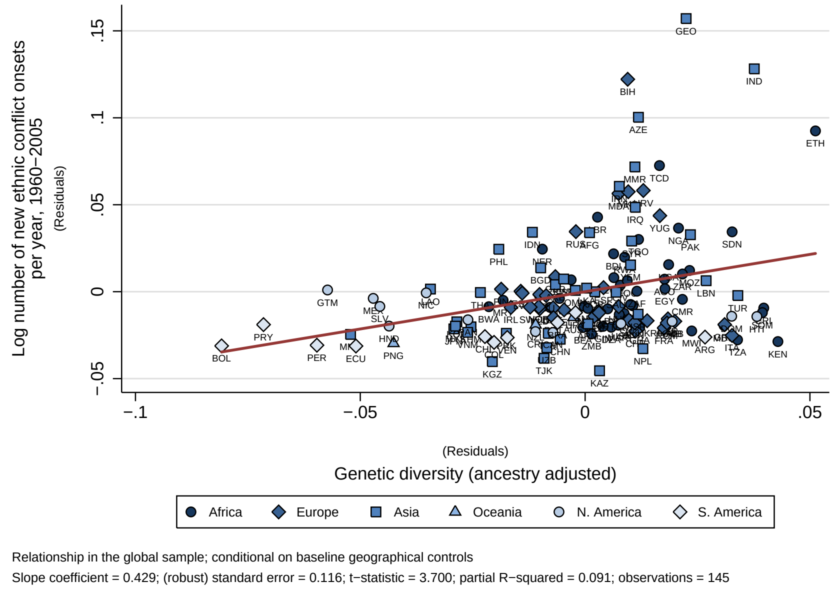
are ethnic versus nonethnic in nature. This asymmetry partly reflects the fact that in practice, it is somewhat difficult to draw a clear distinction between ethnic and nonethnic civil conflict events in the data, but as is well-known in the theoretical literature on civil conflict (e.g., Esteban and Ray, 2008b, 2011a; Caselli and Coleman, 2013), the conceptual difference between the two categories may indeed be crucial to understand the role of the ethnic configuration of a society's population.[33] Throughout much of our empirical analysis, we therefore complement our examination of overall civil conflict (that does not necessarily reflect interethnic cleavages) with an investigation of the influence of genetic diversity on outcomes associated with ethnic civil conflict, especially in light of evidence that genetic diversity may have contributed to various manifestations of ethnolinguistic fragmentation in contemporary national populations (Ashraf and Galor, 2013b).

Replicating our methodology from Table 3, Table 4 presents the results from our baseline cross-country analysis of the influence of genetic diversity on the annual frequency of new WCM09 ethnic civil conflict onsets during the 1960–2005 time horizon. In comparison to our preceding analysis of overall civil conflict frequency, a remarkably similar pattern with respect to the influence of genetic diversity on ethnic civil conflict frequency is evident across specifications in Table 4.[34] Indeed, the unconditioned estimate of the influence of genetic diversity in Column 1 becomes markedly stronger once it is conditioned on our baseline set of geographical covariates (excluding continent fixed effects) in Column 2. Maintaining symmetry with the reporting of our results in the preceding section, we depict the latter relationship between genetic diversity and the annual frequency of WCM09 ethnic civil conflict onsets – conditioned on our baseline set of geographical covariates – on the cross-country scatter plots in Figure 2, both in our full sample of countries (Panel A) and in a sample that omits apparently influential outliers (Panel B).

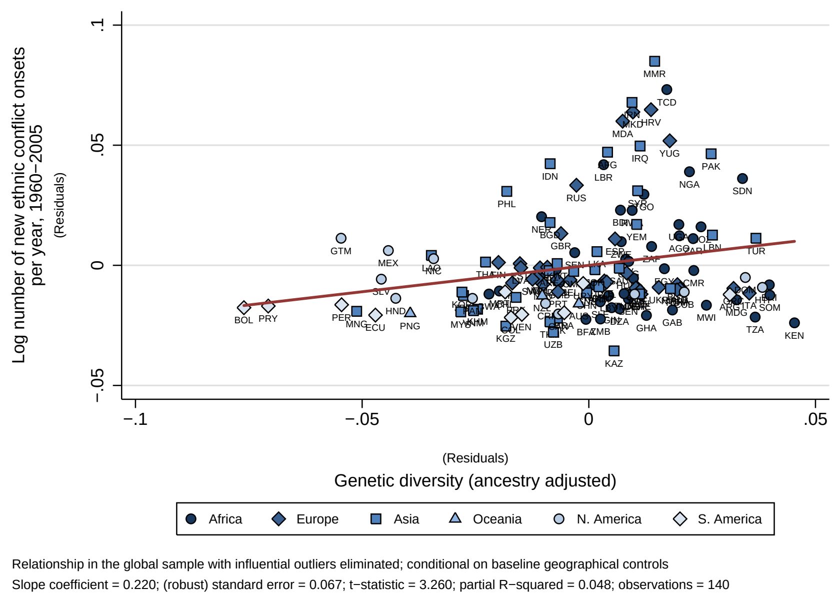
As with our earlier analysis of overall civil conflict frequency, the estimate of relevance from Column 2 becomes somewhat moderated once it is further conditioned on continent fixed effects, but it thereafter remains largely stable in both magnitude and statistical precision when subjected to additional controls for institutional factors, ethnolinguistic fragmentation, and development outcomes, until our full empirical model is attained by the specification in Column 8. The results from this regression suggest that conditional on our complete set of baseline controls, a move from the 10th to the 90th percentile of the cross-country genetic diversity distribution is associated with an increase in conflict frequency by 0.025 new WCM09 ethnic civil conflict outbreaks per year (or 73.0 percent of a standard deviation in the relevant conflict frequency distribution across countries). In addition, the adjusted R^2 statistic of the regression indicates that our baseline empirical model explains 20.7 percent of the cross-country variation in the temporal frequency of ethnic civil conflict onsets, whereas the partial R^2 statistic associated with genetic diversity suggests that the residual cross-country variation in genetic diversity can explain 5.2 percent of

[33] For instance, from a conceptual viewpoint, measures of ethnolinguistic fragmentation are unlikely candidates to explain an armed conflict between a drug cartel and the state.

[34] Similar to the relationship between Table A.3 and Table 3, Table A.4 in Appendix A verifies the qualitative robustness of the results of our analysis of ethnic civil conflict frequency from Table 4 to employing the linguistic fractionalization index of Alesina et al. (2003) in lieu of our baseline control for the ethnic fractionalization index from the same source.



(A) Relationship in the full sample



(B) Relationship with outliers omitted from the sample

Figure 2: Genetic Diversity and the Frequency of Ethnic Civil Conflict Onset across Countries in the Global Sample

Notes: This figure depicts the global cross-country relationship between contemporary genetic diversity and the annual frequency of new ethnic (WCM09) civil conflict onsets during the 1960–2005 time period in (i) an unrestricted sample [Panel A]; and (ii) a sample without influential outliers [Panel B], conditional on the baseline geographical correlates of conflict, as considered by the analysis in Column 2 of Table 4. Each panel presents an added-variable plot with a partial regression line. Given that the sample employed by the analysis in Panel A is not restricted by the availability of data on the additional covariates considered by the analysis in Table 4, the regression coefficient in Panel A is marginally different from that presented in Column 2 of Table 4. The influential outliers that are omitted from the sample in Panel B include Azerbaijan (AZE), Bosnia and Herzegovina (BIH), Ethiopia (ETH), Georgia (GEO), and India (IND).

Table 4: Genetic Diversity and the Frequency of Ethnic Civil Conflict Onset across Countries – The Baseline Analysis

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) 2SLS	(12) 2SLS
	Log number of new WCM09 ethnic civil conflict onsets per year during the 1960–2005 time period											
Genetic diversity (ancestry adjusted)	0.217***	0.418***	0.375**	0.385**	0.352**	0.408**	0.366*	0.391**	0.780***	0.904**	0.707***	0.795***
	[0.082]	[0.121]	[0.152]	[0.183]	[0.176]	[0.189]	[0.192]	[0.191]	[0.277]	[0.356]	[0.254]	[0.297]
Ethnic fractionalization					0.021		0.019	0.016		0.021		0.005
					[0.014]		[0.014]	[0.015]		[0.018]		[0.014]
Ethnolinguistic polarization						0.012	0.005	0.002		-0.008		0.011
						[0.015]	[0.016]	[0.015]		[0.017]		[0.014]
Absolute latitude		-0.366***	-0.564**	-0.419	-0.215	-0.445	-0.250	0.292	-0.537	0.298	-0.673***	0.017
		[0.134]	[0.274]	[0.293]	[0.346]	[0.283]	[0.326]	[0.316]	[0.355]	[0.405]	[0.244]	[0.296]
Land area		-0.917	0.193	0.528	1.018	0.551	0.969	1.365	2.072	2.920	-0.034	1.109
		[1.513]	[1.733]	[1.783]	[1.834]	[1.826]	[1.850]	[1.853]	[2.109]	[1.961]	[1.730]	[1.836]
Ruggedness		0.033	-0.002	-0.003	0.007	-0.001	0.007	0.025	0.023	0.062	0.004	0.022
		[0.043]	[0.047]	[0.046]	[0.049]	[0.047]	[0.049]	[0.047]	[0.055]	[0.056]	[0.045]	[0.042]
Mean elevation		-0.012	-0.006	-0.007	-0.007	-0.007	-0.007	-0.011	-0.014	-0.020	-0.008	-0.014
		[0.009]	[0.009]	[0.011]	[0.011]	[0.011]	[0.011]	[0.010]	[0.012]	[0.013]	[0.009]	[0.010]
Range of elevation		0.007	0.006	0.005	0.004	0.005	0.004	0.001	0.006	0.000	0.006	0.002
		[0.004]	[0.004]	[0.004]	[0.005]	[0.005]	[0.005]	[0.004]	[0.006]	[0.006]	[0.004]	[0.004]
Mean land suitability		0.013	0.017	0.013	0.019	0.016	0.019	0.005	0.017	0.001	0.017	0.001
		[0.014]	[0.014]	[0.017]	[0.016]	[0.017]	[0.016]	[0.017]	[0.017]	[0.021]	[0.013]	[0.015]
Range of land suitability		0.026***	0.030**	0.031**	0.029**	0.033**	0.030**	0.034*	0.033**	0.026	0.034**	0.038**
		[0.010]	[0.013]	[0.013]	[0.013]	[0.014]	[0.015]	[0.017]	[0.016]	[0.020]	[0.013]	[0.016]
Distance to nearest waterway		0.009	0.004	0.003	0.001	0.003	0.002	-0.005	0.001	-0.007	0.003	-0.007
		[0.009]	[0.010]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.011]	[0.012]	[0.010]	[0.010]
Executive constraints, 1960–2005 average				-0.000	-0.001	-0.000	-0.001	0.003		0.003		0.005
				[0.004]	[0.004]	[0.004]	[0.004]	[0.004]		[0.005]		[0.004]
Fraction of years under democracy, 1960–2005				-0.011	-0.007	-0.011	-0.007	-0.008		0.003		-0.013
				[0.025]	[0.024]	[0.025]	[0.024]	[0.023]		[0.024]		[0.021]
Fraction of years under autocracy, 1960–2005				-0.015	-0.013	-0.014	-0.013	-0.015		-0.011		-0.015
				[0.020]	[0.020]	[0.020]	[0.020]	[0.019]		[0.020]		[0.017]
Log oil production per capita, 1960–2005 average								0.003**		0.003*		0.003**
								[0.001]		[0.001]		[0.001]
Log population, 1960–2005 average								-0.000		0.002		-0.000
								[0.003]		[0.004]		[0.003]
Log GDP per capita, 1960–2005 average								-0.021***		-0.023***		-0.022***
								[0.005]		[0.005]		[0.005]
Effect of increasing genetic diversity	0.014***	0.027***	0.024**	0.025**	0.023**	0.027**	0.024*	0.025**	0.038***	0.045**	0.046***	0.052***
from the 10th to the 90th percentile	[0.005]	[0.008]	[0.010]	[0.012]	[0.011]	[0.012]	[0.012]	[0.012]	[0.014]	[0.018]	[0.016]	[0.019]
Continent dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Colonial history dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Sample	Global	Global	Global	Global	Global	Global	Global	Global	Old World	Old World	Global	Global
Observations	141	141	141	141	141	141	141	141	117	117	141	141
Partial R^2 of genetic diversity	–	0.092	0.050	0.049	0.042	0.054	0.042	0.052	0.108	0.123	–	–
Partial R^2 sum of other diversity measures	–	–	–	–	0.017	0.007	0.011	0.008	–	0.015	–	–
First-stage adjusted R^2	–	–	–	–	–	–	–	–	–	–	0.753	0.760
First-stage partial R^2 of migratory distance	–	–	–	–	–	–	–	–	–	–	0.475	0.438
First-stage F statistic	–	–	–	–	–	–	–	–	–	–	206.014	97.246
Adjusted R^2	0.024	0.127	0.165	0.139	0.146	0.138	0.140	0.207	0.194	0.233	–	–

Notes: This table exploits cross-country variations to establish a significant positive reduced-form impact of contemporary genetic diversity on the annual frequency of new ethnic (WCM09) civil conflict onsets during the 1960–2005 time period, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict. For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of colonial history dummies includes indicators for whether a country was ever a colony of the U.K., France, and any other major colonizing power. The 2SLS regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated effect associated with increasing genetic diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the actual number of new conflict onsets per year. Robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

the residual cross-country variation in the conflict outcome variable. Further, in line with our analysis in the preceding section, the better-identified estimates of the influence of genetic diversity on ethnic civil conflict frequency – presented in Columns 9–12 – are sizably larger than their less well-identified counterparts. Specifically, depending on the identification strategy and the set of covariates included in the specification, the results from the regressions in Columns 9–12 suggest that a move from the 10th to the 90th percentile of the cross-country genetic diversity distribution in the relevant sample leads to an increase in conflict frequency by between 0.038 and 0.052 new

WCM09 ethnic civil conflict outbreaks per year (or between 103 and 148 percent of a standard deviation in the relevant conflict frequency distribution across countries).

To provide an appropriate benchmark for making *quantitative* comparisons between our findings with respect to ethnic versus overall civil conflict frequency, Table A.7 in Appendix A replicates our baseline cross-country analysis of the temporal frequency of overall civil conflict onsets, using an outcome variable that reflects the PRIO25 civil conflict coding of WCM09 – i.e., our data source for ethnic civil conflict events, based on an earlier version of the UCDP/PRIO Armed Conflict Dataset. As is apparent from comparing the results presented in Table 4 versus Table A.7, the reduced-form impact of genetic diversity is indeed markedly stronger on the temporal frequency of overall (rather than only ethnic) civil conflict outbreaks, a finding that is consistent with our priors that the influence of genetic diversity on manifestations of intrastate conflict – more broadly defined – operates through mechanisms associated with social divisions that go well beyond only ethnopolitical incompatibilities. Nonetheless, beyond simply corroborating our flagship findings pertaining to overall civil conflict frequency, the findings from our analysis of the frequency of ethnic civil conflict onsets is consistent with our view that at least part of the reduced-form influence of genetic diversity on the potential for social conflict can be attributed to the deeper role of genetic diversity in facilitating endogenous coalitional group formation in prehistory and the subsequent emergence of ethnic markers of cultural differentiation across these coalitional groups over a long expanse of time.

4.1.3 Robustness Checks

We now present three important robustness checks for our cross-country analysis of the influence of genetic diversity on the temporal frequency of either overall or ethnic civil conflict outbreaks in the post-1960 time horizon.

Robustness to the Method of Estimation Given that our baseline cross-country regressions employ least-squares estimation, we apply a log transformation to each of our outcome variables in order to partly address the issue that their cross-country distributions are positively skewed with excess zeros, arising from the fact that new civil conflict onsets are generally rare events in cross-sectional data. An alternative approach to this issue, however, is to employ an estimation method that is tailored to the analysis of over-dispersed count data. Tables A.1 and A.2 in Appendix A therefore replicate our baseline cross-country analyses from Tables 3 and 4, estimating negative-binomial regressions that explain the cross-country variations in the total count of new conflict onsets during the relevant sample period, in relation to PRIO25 civil conflicts and WCM09 ethnic civil conflicts, respectively.[35]

As is evident from the results presented in the appendix tables, the estimated influence of genetic diversity on the total count of new civil conflict onsets is indeed qualitatively identical to

[35] In light of the over-dispersed nature of our count variables, both of which possess cross-country distributions with coefficients of variation larger than unity, an analysis that is based on the negative-binomial model is indeed most appropriate. Nevertheless, we also conducted regressions exploiting the Poisson and zero-inflated Poisson models, and we obtained qualitatively similar results (not reported).

our baseline findings with respect to the temporal frequency of such events. To interpret some of the most stringently conditioned results from our analyses of the conflict count data, the regressions in Column 8 of Tables A.1 and A.2 suggest that a move from the 10th to the 90th percentile of the cross-country genetic diversity distribution in the relevant sample is associated with an increase in the count of conflicts by 0.935 new PRIO25 civil conflict onsets and 0.906 new WCM09 ethnic civil conflict onsets, respectively. These responses in the count data correspond, respectively, to 68.8 percent and 68.5 percent of a standard deviation in the relevant cross-country distribution of the total count of new civil conflict outbreaks. Further, consistently with our priors, the better-identified counterparts of these estimates – presented in Column 10 of either table and obtained by restricting the regression to exploit variations across countries that only belong to the Old World – suggest noticeably larger effects of genetic diversity.[36]

Robustness to Accounting for Spatial Dependence As with any empirical analysis that exploits spatial variations in cross-sectional data, autocorrelation in disturbance terms across observations could be biasing our estimates of the standard errors in our baseline cross-country analyses of conflict frequency. In Tables A.5 and A.6 in Appendix A, we therefore replicate all our regressions from Tables 3 and 4, reporting standard errors that are corrected for cross-sectional spatial dependence, using the methodology proposed by Conley (1999). Reassuringly, depending on the specification examined, the corrected standard errors of the estimated coefficient on genetic diversity are either similar in magnitude or noticeably smaller when compared to their heteroskedasticity robust counterparts in our baseline analyses. This suggests that as far as our analysis of the influence of genetic diversity on conflict frequency is concerned, issues concerning spatial dependence do not pose a threat to identification.

Robustness to the Elimination of Regions from the Estimation Sample Following the norm in cross-country empirical studies of civil conflict, we also investigate whether our baseline findings – specifically, with respect to the influence of genetic diversity on either overall or ethnic civil conflict frequency in the relevant globally representative sample of countries – are driven by potentially influential observations in a given world region. Namely, for each of the two conflict frequency outcome variables that we consider, we first drop the observations belonging to a given world region from our full estimation sample and then re-estimate our baseline empirical model in the residual sample of countries.[37] The findings obtained under this robustness check for our analyses of overall and ethnic civil conflict frequency are collected, respectively, in Tables A.8 and A.9 in Appendix A, wherein the first five columns of each table report the results from OLS

[36] In our analyses of conflict count data, given the absence of an appropriate IV estimator, we are unable to implement our second identification strategy that is based on the instrumentation of genetic diversity in a globally representative sample of countries.

[37] The world regions that we consider for one-at-a-time elimination from the full estimation sample include Sub-Saharan Africa (SSA), Middle East and North Africa (MENA), East Asia and Pacific (EAP), and Latin America and the Caribbean (LAC). Further, given the lower degrees of freedom afforded by the estimation samples with eliminated regions, we now ignore continent fixed effects in each of the estimated regressions, in order to preserve as much of the cross-country variation in conflict frequency as possible, thus permitting the independent variables in the model to possess at least some explanatory power.

regressions and the remaining five present their 2SLS counterparts. Reassuringly, both tables reveal that our baseline findings are not qualitatively sensitive to the exclusion of any potentially influential world region from our full estimation samples, in the sense that genetic diversity retains its significant explanatory power for the temporal frequency of either overall or ethnic civil conflict onsets in all restricted samples.

4.2 Analysis of Civil Conflict Incidence in Repeated Cross-Country Data

4.2.1 Baseline Analysis of Civil Conflict Incidence

Exploiting variations in quinquennially repeated cross-country data, Table 5 presents the results from our baseline analysis of the influence of genetic diversity on the temporal prevalence of civil conflict – namely, the likelihood of observing the annual incidence of one or more active civil conflict episodes in a given 5-year interval during the post-1960 time horizon. The first four columns report our findings from regressions explaining the incidence of PRIO25 civil conflict episodes, whereas the remaining four collect our results from regressions explaining the incidence of WCM09 ethnic civil conflict episodes.

To keep the exposition concise, we focus our analysis on regressions that employ either one of our two identification strategies. Specifically, for each of our two conflict incidence outcome variables, the first two regressions represent probit models that exploit repeated cross-country variations in a sample of countries only belonging to the Old World, and the latter two represent IV probit models that exploit variations in a globally representative sample, employing the migratory distance of a country's prehistorically native population from East Africa as an instrument for the genetic diversity of its contemporary national population. In addition, for each outcome variable and for each identification strategy, we estimate two distinct specifications; one that partials out the influence of only exogenous geographical covariates (including continent fixed effects), and the other that conditions the analysis on the full set of controls in our baseline empirical model of conflict incidence. All our regressions, however, always account for the lagged observation of the outcome variable and a complete set of time-interval (5-year period) dummies, and wherever relevant, our time-varying controls for institutional factors and development outcomes enter the specification with a one-period lag, in order to mitigate issues of reverse-causality bias in their estimated coefficients.

Regardless of the identification strategy employed or the set of covariates included in the specification, the results collected in Table 5 establish genetic diversity as a robust and highly significant predictor of civil conflict incidence, with respect to both PRIO25 civil conflicts and WCM09 ethnic civil conflicts.[38] For instance, exploiting variations in our globally representative

[38] The results presented in Table A.11 in Appendix A additionally confirm that when employing logit rather than probit models of civil conflict incidence, our baseline findings from all specifications that exploit our first identification strategy of restricting attention to variations in the Old World remain qualitatively unchanged. The table also verifies the robustness of our baseline findings from these specifications to employing the so-called "rare events" logit model of King and Zeng (2001), which corrects for any bias in maximum-likelihood estimation that may arise when there are a small number of observed cases for the rarer outcome of a binary dependent variable. In the absence of readily

Table 5: Genetic Diversity and the Incidence of Civil Conflict in Quinquennially Repeated Cross-Country Data

	(1) Probit	(2) Probit	(3) IV Probit	(4) IV Probit	(5) Probit	(6) Probit	(7) IV Probit	(8) IV Probit
	PRIO25 civil conflict incidence				WCM09 ethnic civil conflict incidence			
Genetic diversity (ancestry adjusted)	11.883***	12.031**	12.540***	12.829***	20.171***	21.488***	15.379***	15.732***
	[4.502]	[4.685]	[4.215]	[4.808]	[5.830]	[6.174]	[5.198]	[5.897]
Ethnic fractionalization		-0.232		-0.297		0.179		-0.328
		[0.356]		[0.328]		[0.469]		[0.472]
Ethnolinguistic polarization		0.223		0.398		0.102		0.744
		[0.345]		[0.318]		[0.420]		[0.505]
Absolute latitude	-18.176***	-8.131	-22.414***	-17.520**	-30.736***	-20.536*	-34.580***	-34.652***
	[6.775]	[8.665]	[6.119]	[8.545]	[8.838]	[12.258]	[7.875]	[12.046]
Land area	4.931	-0.176	-28.616	-51.352	27.294	27.762	10.644	13.181
	[41.279]	[36.358]	[43.673]	[48.091]	[37.055]	[45.950]	[47.713]	[56.145]
Ruggedness	1.297	2.785**	0.832	1.561	1.416	2.585**	0.062	0.183
	[1.149]	[1.153]	[0.965]	[1.020]	[1.267]	[1.272]	[1.332]	[1.595]
Mean elevation	-0.517*	-0.702***	-0.372*	-0.503**	-0.553*	-0.678**	0.017	-0.027
	[0.282]	[0.262]	[0.215]	[0.214]	[0.296]	[0.267]	[0.318]	[0.343]
Range of elevation	0.106	-0.042	0.140**	0.043	0.000	-0.174	-0.089	-0.181
	[0.072]	[0.096]	[0.065]	[0.087]	[0.072]	[0.109]	[0.088]	[0.122]
Mean land suitability	0.149	-0.205	0.227	0.004	0.378	0.271	0.096	0.286
	[0.323]	[0.390]	[0.280]	[0.346]	[0.368]	[0.504]	[0.392]	[0.503]
Range of land suitability	0.844***	0.994***	0.689***	0.775**	1.336***	1.464***	1.406***	1.801***
	[0.290]	[0.343]	[0.256]	[0.316]	[0.419]	[0.494]	[0.433]	[0.504]
Distance to nearest waterway	0.315	0.475**	0.342*	0.503**	0.465**	0.609**	0.279	0.451*
	[0.209]	[0.220]	[0.184]	[0.211]	[0.213]	[0.243]	[0.212]	[0.232]
Average executive constraints, lagged		0.084		0.079		0.191***		0.130*
		[0.062]		[0.057]		[0.067]		[0.068]
Fraction of years under democracy, lagged		-0.290		-0.483*		-0.472		-0.609**
		[0.275]		[0.253]		[0.292]		[0.269]
Fraction of years under autocracy, lagged		-0.207		-0.295*		-0.111		-0.333
		[0.188]		[0.172]		[0.292]		[0.277]
Log average oil production per capita, lagged		0.043		0.023		0.042		0.017
		[0.031]		[0.029]		[0.034]		[0.034]
Log average population, lagged		0.078		0.093		0.106		0.033
		[0.072]		[0.069]		[0.110]		[0.100]
Log average GDP per capita, lagged		-0.380***		-0.233**		-0.350***		-0.154
		[0.116]		[0.103]		[0.122]		[0.158]
Conflict incidence, lagged	1.765***	1.660***	1.654***	1.570***	2.151***	2.048***	2.001***	1.912***
	[0.134]	[0.127]	[0.126]	[0.122]	[0.184]	[0.199]	[0.190]	[0.201]
Marginal effect	2.231***	2.137***	2.538***	2.490***	2.600***	2.596***	2.170***	2.107***
	[0.810]	[0.816]	[0.848]	[0.947]	[0.740]	[0.755]	[0.725]	[0.783]
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5-year period dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	Yes	No	Yes	No	Yes	No	Yes
Colonial history controls	No	Yes	No	Yes	No	Yes	No	Yes
Sample	Old World	Old World	Global	Global	Old World	Old World	Global	Global
Observations	944	944	1,154	1,154	927	927	1,039	1,039
Countries	119	119	141	141	117	117	129	129
Time horizon	1960–2008	1960–2008	1960–2008	1960–2008	1960–2005	1960–2005	1960–2005	1960–2005
Time frequency	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly
Pseudo R^2	0.423	0.457	–	–	0.516	0.549	–	–

Notes: This table exploits variations in a quinquennially repeated cross-section of countries to establish a significant positive reduced-form impact of contemporary genetic diversity on the likelihood of observing the incidence of (i) an overall (PRIO25) civil conflict in any given 5-year interval during the 1960–2008 time horizon; and (ii) an ethnic (WCM09) civil conflict in any given 5-year interval during the 1960–2005 time horizon, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict. To account for temporal dependence in conflict outcomes, all regressions control for the incidence of conflict in the previous 5-year interval, following Esteban, Mayoral and Ray (2012). For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of time-varying (lagged) colonial history controls includes variables that reflect the fraction of years from the previous 5-year interval that a country served as a colony of the U.K., France, and any other major colonizing power. The IV probit regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated marginal effect of a 1 percentage point increase in genetic diversity is the average marginal effect across the entire cross-section of observed diversity values, and it reflects the increase in the likelihood of a conflict incidence in any given 5-year interval, expressed in percentage points. Robust standard errors, clustered at the country level, are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

sample of countries, the IV probit regression presented in Column 4 suggests that conditional on our complete set of controls for geographical characteristics, institutional factors, ethnolinguistic fragmentation, and the correlates of economic development, a 1 percentage point increase in genetic diversity leads to an increase in the quinquennial likelihood of a PRIO25 civil conflict incidence by 2.49 percentage points, as reflected by an estimated average marginal effect that is statistically significant at the 1 percent level. Further, according to the similar IV probit regression presented in Column 8, a 1 percentage point increase in genetic diversity leads to an increase in the quinquennial likelihood of a WCM09 ethnic civil conflict incidence by 2.11 percentage points – an estimated average marginal effect that is also statistically significant at the 1 percent level.

Given that the aforementioned relationships only reflect the *average* marginal effects of genetic diversity in the relevant samples, the plots presented in Figure 3 illustrate precisely how the *predicted* likelihoods – associated with the incidence of either PRIO25 civil conflicts (Panel A) or WCM09 ethnic civil conflicts (Panel B) – vary as one moves along the global cross-country genetic diversity distribution, based on the IV probit regressions from Columns 4 and 8.[39] The economically significant influence of genetic diversity is clearly evident in these plots, which indicate that a move from the 10th to the 90th percentile of the cross-country genetic diversity distribution in the relevant estimation sample leads to an increase in the predicted quinquennial likelihood of civil conflict incidence from 19.2 percent to 34.9 percent for PRIO25 civil conflicts, and from 12.2 percent to 23.8 percent for WCM09 ethnic civil conflicts.

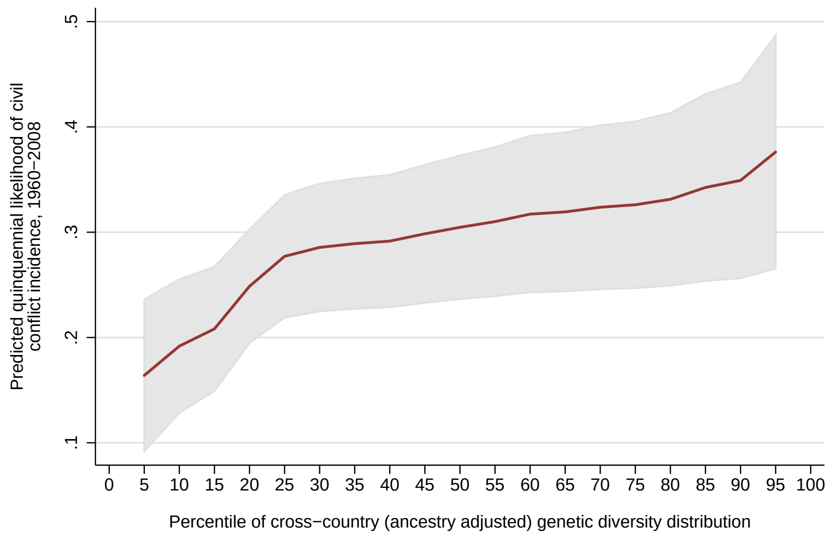
4.2.2 Robustness Checks

We now subject the findings from our baseline analysis of the influence of genetic diversity on the temporal prevalence of either overall or ethnic civil conflict in repeated cross-country data to several robustness checks.

Robustness to Alternative Correlates of Conflict Incidence We excluded a few potentially confounding control variables from our baseline set of covariates, in the interest of keeping our main specifications from becoming unwieldy and in order to maximize the number of observations in our estimation samples, given the more limited coverage by the data on some of these other control variables. The results presented in Table 6 confirm that our key findings with respect to the influence of genetic diversity on civil conflict incidence are indeed robust to accounting for the potentially confounding effects of these other correlates of conflict incidence. The table is organized into four quadrants. In the top panel of the table, we focus on specifications examining the incidence of PRIO25 civil conflicts, whereas the bottom panel presents the corresponding specifications for the incidence of WCM09 ethnic civil conflicts. Further, each panel examines the robustness of the main

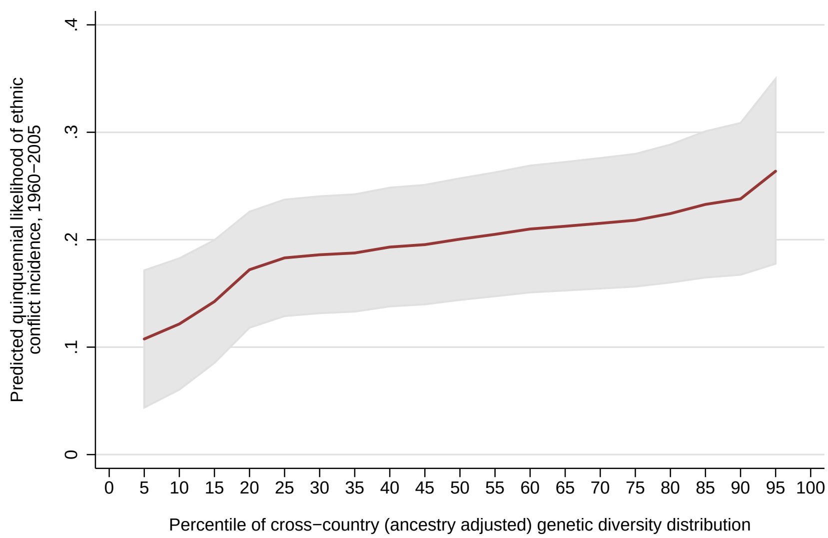
available IV counterparts of the logit and "rare events" logit regression models, we are unable to implement similar robustness checks for the specifications that exploit our second identification strategy, based on the instrumentation of genetic diversity in a globally representative sample of countries.

[39]Similar to Figure 3, the two panels of Figure A.1 in Appendix A depict the manner in which the predicted quinquennial likelihoods – associated with the incidence of either PRIO25 civil conflicts (Panel A) or WCM09 ethnic civil conflicts (Panel B) – respond as one moves along the cross-country genetic diversity distribution in the Old World, based on the probit regressions from Columns 2 and 6.



Predicted likelihoods based on an IV probit regression of conflict incidence on instrumented diversity; conditional on all baseline controls
Average marginal effect of a 0.01−increase in diversity = 2.490 percent; standard error = 0.947; p−value = 0.009

(A) Effect on overall civil conflict incidence



Predicted likelihoods based on an IV probit regression of conflict incidence on instrumented diversity; conditional on all baseline controls
Average marginal effect of a 0.01−increase in diversity = 2.107 percent; standard error = 0.783; p−value = 0.007

(B) Effect on ethnic civil conflict incidence

Figure 3: The Effect of Instrumented Genetic Diversity on the Quinquennial Likelihood of Civil Conflict Incidence in the Global Sample

Notes: This figure depicts the influence of contemporary genetic diversity at the country level on the *predicted* likelihood of observing the incidence of (i) an overall (PRIO25) civil conflict in any given 5-year interval during the 1960–2008 time horizon [Panel A]; and (ii) an ethnic (WCM09) civil conflict in any given 5-year interval during the 1960–2005 time horizon [Panel B], conditional on other well-known diversity measures, the proximate geographical, institutional, and development-related correlates of conflict, controls for temporal dependence in conflict outcomes, and continent and 5-year time-interval dummies. In each panel, the predicted likelihood of conflict incidence is illustrated as a function of the percentile of the cross-country genetic diversity distribution, and the prediction is based on the relevant IV probit regression from Table 5, exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity, and conducted using the global sample of countries and the full set of covariates considered by the analysis of the conflict outcome in question. The shaded area in each plot reflects the 95-percent confidence-interval region of the depicted relationship.

result to six individual sets of additional covariates, employing either probit regressions (presented in Columns 1–6) that restrict attention to variations in the Old World or their corresponding IV probit regressions (presented in Columns 6–12) that exploit global variations while instrumenting the genetic diversity of a country's contemporary national population with the migratory distance of its prehistorically indigenous settlements from East Africa. All regressions include our full set of baseline covariates (not reported in the table to conserve space), except in specifications where the set of additional covariates for the robustness check makes one or more of our baseline control variables conceptually redundant.

The first four specifications in each quadrant of Table 6 investigate the robustness of our main findings with respect to controls for additional well-known diversity measures, based on intergroup distributional indices of ethnolinguistic fragmentation. Specifically, in the first three specifications, we separately introduce controls for the ethnic fractionalization index of Fearon (2003) and the measures of linguistic and religious fractionalization from Alesina et al. (2003), each in lieu of our baseline control for ethnic fractionalization from the latter source. Moreover, consistently with the theory of intergroup conflict formulated by Esteban and Ray (2011b), the empirical study of Esteban, Mayoral and Ray (2012) finds that three conceptually distinct indices of ethnolinguistic fragmentation – namely, fractionalization, polarization, and a Greenberg-Gini index of ethnic difference – are all simultaneously important for explaining the potential for civil conflict in society. The fourth specification in each quadrant thus simultaneously introduces controls for these three indices of ethnolinguistic diversity from Esteban, Mayoral and Ray (2012), in lieu of our baseline controls for both ethnic fractionalization from Alesina et al. (2003) and ethnolinguistic polarization from Desmet, Ortuño-Ortín and Wacziarg (2012). As is apparent from the results, irrespective of the specific outcome variable examined or the identification strategy employed, genetic diversity retains its statistically and economically significant influence on civil conflict incidence when subjected to all the aforementioned alternative controls for the degree of ethnolinguistic fragmentation of a national population.[40] In addition, the fact that the ethnolinguistic polarization index of Esteban, Mayoral and Ray (2012) enters the regressions that exploit global variations with a significant coefficient suggests that genetic diversity and ethnolinguistic polarization may well be capturing complementary mechanisms for explaining the incidence of civil conflict.

[40] In additional robustness checks (results not shown), we (i) performed one-at-a-time substitutions of our baseline control for the ethnic fractionalization index of Alesina et al. (2003) with three different measures of ethnolinguistic fractionalization from Desmet, Ortuño-Ortín and Wacziarg (2012), reflecting fractionalization across subnational groups that are categorized by more ancestral linguistic divisions, occurring at lower levels of country-specific hierarchical linguistic trees – namely, levels 1, 5, and 10; (ii) performed one-at-a-time substitutions of our baseline control for the ethnolinguistic polarization index (at level 15) of Desmet, Ortuño-Ortín and Wacziarg (2012) with three of its "more ancestral" counterparts, measured at levels 1, 5, and 10; and (iii) augmented our set of baseline controls for ethnolinguistic fragmentation with either the Greenberg-Gini index or the peripheral heterogeneity index of linguistic diversity, both from Desmet, Ortuño-Ortín and Weber (2009). In all cases, our main finding that genetic diversity confers a statistically and economically significant influence on the incidence of both overall and ethnic civil conflict events was qualitatively unaltered.

Table 6: Genetic Diversity and the Incidence of Civil Conflict in Quinquennially Repeated Cross-Country Data – Robustness to Alternative Correlates of Conflict Incidence

	(1) Probit	(2) Probit	(3) Probit	(4) Probit	(5) Probit	(6) Probit	(7) IV Probit	(8) IV Probit	(9) IV Probit	(10) IV Probit	(11) IV Probit	(12) IV Probit
PANEL A	PRIO25 civil conflict incidence											
Genetic diversity (ancestry adjusted)	12.406***	12.674***	13.284***	10.884**	12.152**	11.372**	12.501***	13.033***	13.008***	11.867**	14.753***	14.275***
	[4.658]	[4.670]	[4.818]	[5.448]	[4.925]	[5.022]	[4.642]	[4.861]	[4.716]	[5.533]	[5.644]	[5.519]
Ethnic fractionalization (Fearon, 2003)	0.006						0.018					
	[0.306]						[0.292]					
Linguistic fractionalization (Alesina et al., 2003)		0.259						0.198				
		[0.307]						[0.278]				
Religious fractionalization (Alesina et al., 2003)			-0.551**						-0.646**			
			[0.276]						[0.262]			
Ethnolinguistic fractionalization (Esteban et al., 2012)				0.287						0.259		
				[0.368]						[0.345]		
Ethnolinguistic polarization (Esteban et al., 2012)				2.379						3.504**		
				[1.605]						[1.485]		
Gini index of ethnolinguistic diversity (Esteban et al., 2012)				-1.480						-1.440		
				[1.450]						[1.340]		
Log percentage mountainous terrain					0.124*						0.089	
					[0.072]						[0.074]	
Noncontiguous state dummy					0.367*						0.559***	
					[0.217]						[0.199]	
Disease richness					-0.007						-0.003	
					[0.011]						[0.009]	
Log years since Neolithic Revolution (ancestry adjusted)						0.086						-0.129
						[0.236]						[0.243]
Marginal effect	2.270***	2.265***	2.376***	1.954**	2.155**	2.035**	2.478***	2.509***	2.531***	2.317**	2.859**	2.790**
	[0.839]	[0.817]	[0.846]	[0.975]	[0.860]	[0.884]	[0.932]	[0.943]	[0.927]	[1.094]	[1.120]	[1.099]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Old World	Old World	Old World	Old World	Old World	Old World	Global	Global	Global	Global	Global	Global
Observations	938	935	944	883	928	936	1,148	1,117	1,154	1,083	1,138	1,146
Countries	118	118	119	107	117	118	140	137	141	128	139	140
Time horizon	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008
Time frequency	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly
Pseudo R^2	0.448	0.456	0.458	0.460	0.460	0.455	–	–	–	–	–	–
PANEL B	WCM09 ethnic civil conflict incidence											
Genetic diversity (ancestry adjusted)	20.306***	23.605***	20.602***	18.694***	24.700***	22.256***	14.105**	19.464***	14.843***	13.137**	18.037***	16.225**
	[6.022]	[6.183]	[5.959]	[6.535]	[6.395]	[6.747]	[5.836]	[5.939]	[5.718]	[6.689]	[6.430]	[6.973]
Ethnic fractionalization (Fearon, 2003)	0.669*						0.536					
	[0.354]						[0.346]					
Linguistic fractionalization (Alesina et al., 2003)		0.892**						0.707*				
		[0.412]						[0.374]				
Religious fractionalization (Alesina et al., 2003)			-0.007						-0.417			
			[0.374]						[0.399]			
Ethnolinguistic fractionalization (Esteban et al., 2012)				0.632						0.353		
				[0.461]						[0.415]		
Ethnolinguistic polarization (Esteban et al., 2012)				-0.087						5.397**		
				[1.864]						[2.486]		
Gini index of ethnolinguistic diversity (Esteban et al., 2012)				-4.652						-3.213		
				[3.269]						[2.864]		
Log percentage mountainous terrain					-0.002						0.021	
					[0.094]						[0.087]	
Noncontiguous state dummy					0.337						0.157	
					[0.234]						[0.252]	
Disease richness					-0.003						0.022	
					[0.015]						[0.017]	
Log years since Neolithic Revolution (ancestry adjusted)						-0.071						-0.041
						[0.278]						[0.352]
Marginal effect	2.450***	2.775***	2.476***	2.295***	2.970***	2.689***	1.858**	2.443***	1.968***	1.727*	2.424***	2.172**
	[0.732]	[0.732]	[0.727]	[0.803]	[0.790]	[0.823]	[0.758]	[0.750]	[0.755]	[0.910]	[0.873]	[0.931]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Old World	Old World	Old World	Old World	Old World	Old World	Global	Global	Global	Global	Global	Global
Observations	936	918	927	882	927	927	1,048	1,002	1,039	984	1,039	1,039
Countries	118	116	117	107	117	117	130	125	129	118	129	129
Time horizon	1960–2005	1960–2005	1960–2005	1960–2005	1960–2005	1960–2005	1960–2005	1960–2005	1960–2005	1960–2005	1960–2005	1960–2005
Time frequency	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly
Pseudo R^2	0.548	0.557	0.550	0.551	0.550	0.549	–	–	–	–	–	–

Notes: This table exploits variations in a quinquennially repeated cross-section of countries to establish that the significant positive reduced-form impact of contemporary genetic diversity on the likelihood of observing the incidence of (i) an overall (PRIO25) civil conflict in any given 5-year interval during the 1960–2008 time horizon [Panel A]; and (ii) an ethnic (WCM09) civil conflict in any given 5-year interval during the 1960–2005 time horizon [Panel B], conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict, is robust to accounting for alternative distributional indices of diversity and for additional geographical and historical correlates of conflict. All regressions additionally control for the entire set of covariates considered by the baseline analysis of conflict incidence in Table 5, with the exception that in Columns 1–3 and 7–9, each of the reported control variables is employed in lieu of the baseline control for ethnic fractionalization (Alesina et al., 2003), whereas in Columns 4 and 10, the set of reported control variables from Esteban, Mayoral and Ray (2012) replaces the baseline controls for both ethnic fractionalization (Alesina et al., 2003) and ethnolinguistic polarization (Desmet, Ortuño-Ortín and Wacziarg, 2012), in the interest of mitigating multicollinearity. The IV probit regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated marginal effect of a 1 percentage point increase in genetic diversity is the average marginal effect across the entire cross-section of observed diversity values, and it reflects the increase in the likelihood of a conflict incidence in any given 5-year interval, expressed in percentage points. Robust standard errors, clustered at the country level, are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

The fifth specification in each quadrant establishes the robustness of our key result with respect to controls for additional geographical factors. According to the "insurgency theory" of civil conflict (Fearon and Laitin, 2003), conditions that facilitate the emergence and perpetuation of armed rebel organizations are potentially more important than ethnopolitical grievances or deeper interethnic cleavages. This theory emphasizes several determinants that may be conducive to insurgent activities, amongst which are two geographical characteristics – namely, the prevalence of mountainous regions within a country and the presence of one or more territories that are noncontiguous with the region containing the state's geopolitical center. It has also been argued, however, that the extrinsic mortality risk associated with a country's disease environment can increase its potential for civil conflict, either by contributing to interethnic divisions (Letendre, Fincher and Thornhill, 2010) or by reducing the opportunity cost of individual selection into violent activities (Cervellati, Sunde and Valmori, 2011). To account for the potentially confounding influence of all these geographically based mechanisms, the specification in question simultaneously introduces controls for the percentage of mountainous terrain in a country and an indicator that reflects whether a country possesses any noncontiguous territories, both adopted from Fearon and Laitin (2003), in addition to an index of "disease richness" (i.e., the total number of different types of infectious diseases in a country), as reported by Fincher and Thornhill (2008).[41] The results presented in Columns 5 and 11 of either panel, however, indicate that while the geographical factors associated with the "insurgency theory" do enter the specification significantly in some regressions, our main finding regarding the impact of genetic diversity on the incidence of either overall or ethnic civil conflict remains qualitatively intact (and even increases in magnitude in some regressions, relative to their baseline counterparts from Table 5) when conditioned on the influence of all three additional geographical covariates.

By shaping the historical pattern of economic development across premodern societies (Ashraf and Galor, 2011), the timing of the prehistoric transition from hunting and gathering to sedentary agriculture during the Neolithic Revolution may well have influenced the unobserved spatial pattern of historical conflict potential. In addition, by contributing to both interethnic fissions and fusions in the long run, the advent of sedentary agriculture may also have partly governed the unobserved pattern of ethnolinguistic fragmentation across premodern societies. Specifically, although the Neolithic Revolution may have catalyzed ethnic differentiation amongst groups by spurring social stratification in early sedentary societies, it could also have served to homogenize ethnic markers across groups through the rise of institutionalized statehood (Ashraf and Galor, 2013b). As such, the pattern of conflict potential across modern national populations could be partly rooted in the differential timing of the Neolithic Revolution across their ancestral societies. To account for the potentially confounding influence of this particular channel, the sixth and final specification in each quadrant of Table 6 introduces a control variable, adopted from Putterman and Weil (2010) and Ashraf and Galor (2013a), that reflects the timing of the transition to sedentary agriculture, as experienced on average across all the precolonially native and nonnative groups

[41] Although our baseline findings do not indicate a systematic link between terrain ruggedness and civil conflict incidence, our control for terrain ruggedness by itself may not fully capture the potential influence of the percentage of mountainous terrain in a country. Not only is the latter variable more commonly used in the literature, it may also be a superior proxy for the availability of safe havens for rebel groups.

whose descendants comprise a contemporary national population. As is evident from the results presented in Columns 6 and 12 of both panels, the significant influence of genetic diversity on the incidence of either overall or ethnic civil conflict not only remains fully intact but also increases in magnitude in some regressions (relative to their baseline counterparts from Table 5). In contrast, the ancestry-adjusted measure of the timing of the transition to agriculture does not enter the specification significantly in any of the four regressions examined.

Robustness to Alternative Measures of Conflict Incidence It is well-known that empirical results in the civil conflict literature can be rather sensitive to alternative definitions and types of intrastate conflict (e.g., Sambanis, 2004; Hegre and Sambanis, 2006). Our analysis in Table 7 addresses this potential sensitivity issue with respect to the influence of genetic diversity on the incidence of civil conflict. Specifically, we explore three different outcome variables, each reflecting an alternative definition or type of conflict incidence, and similar to our baseline analysis of conflict incidence in Table 5, we focus our robustness analysis on regressions that utilize one of our two identification strategies – namely, either probit models that exploit variations restricted to the Old World or their IV probit counterparts that exploit global variations while employing the migratory distance of a country's prehistorically indigenous population from East Africa as an instrument for the genetic diversity of its modern-day national population. Like before, for each outcome variable and identification strategy, we estimate two alternative specifications, where the first one considers only exogenous geographical covariates and the second one partials out the influence of our full set of baseline control variables.

The regressions presented in Columns 1–8 – examining the first two alternative definitions of conflict incidence – extend our results regarding the significant reduced-form causal impact of genetic diversity on conflict incidence to large-scale civil conflicts or "civil wars." Specifically, corresponding to our baseline analysis of the incidence of PRIO25 civil conflicts, the first four regressions explain the temporal prevalence in a given 5-year interval of one or more annual conflict events that are identified as episodes of civil war by the PRIO1000 criterion, which unlike the PRIO25 coding, imposes a death threshold of 1,000 annual battle-related casualties, rather than 25. Likewise, corresponding to our baseline analysis of the incidence of WCM09 ethnic civil conflicts, the second four regressions examine the quinquennial prevalence of one or more annual conflict events that are classified as episodes of ethnic civil war by Wimmer and Min (2006, 2009) (henceforth referred to as WM0609).[42] As is apparent from the results in Columns 1–8, genetic diversity does indeed confer a qualitatively robust and highly statistically significant influence on the temporal incidence of either overall or ethnic civil war, regardless of the set of covariates considered by the specification or the identification strategy employed. To interpret the economic significance of its impact, the estimated average marginal effects associated with the IV probit regressions

[42] Unfortunately, our data source for WCM09 ethnic civil conflict events – namely, the data set of Wimmer, Cederman and Min (2009) – does not separately identify large-scale ethnic civil conflicts (i.e., ethnic civil wars). As such, our current robustness check is restricted to employing the WM0609 coding of ethnic civil wars, which is based on the Correlates of War (COW) Project, as opposed to the UCDP/PRIO Armed Conflict Dataset. This constraint additionally restricts our sample period for the analysis of WM0609 ethnic civil war incidence to the 1960–2001 time horizon, as opposed to the 1960–2005 time span for the incidence of WCM09 ethnic civil conflicts.

Table 7: Genetic Diversity and the Incidence of Civil Conflict in Quinquennially Repeated Cross-Country Data – Robustness to Alternative Measures of Conflict Incidence

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Probit	Probit	IV Probit	IV Probit	Probit	Probit	IV Probit	IV Probit	Probit	Probit	IV Probit	IV Probit
	PRIO1000 civil war incidence				WM0609 ethnic civil war incidence				UCDP nonstate conflict incidence			
Genetic diversity (ancestry adjusted)	14.759***	14.024***	16.021***	16.725***	28.098***	33.755***	22.620***	25.920***	26.808***	35.301***	24.972***	30.326***
	[4.836]	[5.142]	[4.587]	[5.509]	[7.664]	[8.438]	[6.580]	[7.575]	[6.592]	[7.870]	[5.391]	[6.180]
Ethnic fractionalization		-0.420		-0.596		-0.493		-0.822		-0.051		0.102
		[0.496]		[0.471]		[0.717]		[0.647]		[0.771]		[0.680]
Ethnolinguistic polarization		0.190		0.389		-0.322		0.486		0.570		1.200**
		[0.448]		[0.398]		[0.622]		[0.607]		[0.562]		[0.490]
Absolute latitude	-21.520***	-16.346	-31.480***	-30.945***	-32.945***	-29.850**	-34.920***	-45.114***	-45.017***	-37.184**	-47.818***	-43.001***
	[7.750]	[11.269]	[8.029]	[11.463]	[9.240]	[14.386]	[9.116]	[14.160]	[9.711]	[15.808]	[8.468]	[13.601]
Land area	-1.022	14.755	-19.607	-6.101	-2.413	-0.678	-169.444	-59.345	4.666	-2.057	3.963	-63.013
	[45.030]	[40.533]	[47.268]	[47.083]	[120.759]	[150.973]	[166.770]	[148.535]	[51.466]	[53.573]	[41.946]	[56.493]
Ruggedness	1.430	2.422	0.738	1.162	2.571	2.308	0.147	-1.137	3.345	7.391***	1.746	4.478**
	[1.278]	[1.511]	[1.153]	[1.357]	[2.137]	[2.159]	[1.962]	[2.178]	[2.066]	[2.347]	[1.716]	[1.882]
Mean elevation	-0.503*	-0.694**	-0.216	-0.348	-0.428	-0.454	0.032	0.160	-1.153**	-1.703***	-0.631*	-0.894**
	[0.289]	[0.295]	[0.244]	[0.263]	[0.386]	[0.397]	[0.343]	[0.367]	[0.450]	[0.484]	[0.365]	[0.402]
Range of elevation	0.196***	0.155*	0.182***	0.152	0.094	0.056	0.069	0.110	0.307***	0.037	0.250***	-0.016
	[0.058]	[0.085]	[0.071]	[0.093]	[0.124]	[0.153]	[0.120]	[0.158]	[0.096]	[0.147]	[0.096]	[0.142]
Mean land suitability	0.458	0.387	0.109	0.026	-0.356	-1.049	-0.307	-0.535	0.667	-0.358	0.207	-0.296
	[0.396]	[0.499]	[0.393]	[0.462]	[0.564]	[0.730]	[0.490]	[0.601]	[0.583]	[0.585]	[0.561]	[0.596]
Range of land suitability	0.544*	0.865*	0.701**	1.054**	0.631	0.897	0.961**	1.383**	1.222***	1.292**	1.494***	1.480***
	[0.328]	[0.460]	[0.337]	[0.453]	[0.526]	[0.751]	[0.488]	[0.617]	[0.416]	[0.583]	[0.399]	[0.488]
Distance to nearest waterway	0.403*	0.436*	0.324	0.326	0.629***	0.331	0.551**	0.334	0.385	0.457*	0.255	0.448*
	[0.223]	[0.253]	[0.206]	[0.244]	[0.231]	[0.382]	[0.237]	[0.317]	[0.254]	[0.260]	[0.219]	[0.259]
Average executive constraints, lagged		-0.020		-0.008		0.010		-0.039		-0.063		-0.002
		[0.082]		[0.078]		[0.110]		[0.094]		[0.105]		[0.097]
Fraction of years under democracy, lagged		-0.469		-0.388		-0.122		-0.501		-0.200		-0.162
		[0.319]		[0.325]		[0.670]		[0.541]		[0.387]		[0.344]
Fraction of years under autocracy, lagged		-0.364		-0.323		-0.041		-0.460		0.075		0.140
		[0.249]		[0.222]		[0.341]		[0.314]		[0.361]		[0.287]
Log average oil production per capita, lagged		0.015		0.004		-0.003		-0.068		-0.091*		-0.028
		[0.040]		[0.036]		[0.059]		[0.058]		[0.055]		[0.047]
Log average population, lagged		0.000		0.011		-0.042		-0.088		0.401***		0.398***
		[0.089]		[0.084]		[0.131]		[0.117]		[0.129]		[0.124]
Log average GDP per capita, lagged		-0.230		-0.133		-0.413*		-0.074		-0.356**		-0.289*
		[0.172]		[0.154]		[0.231]		[0.219]		[0.175]		[0.160]
War/conflict incidence, lagged	1.680***	1.630***	1.611***	1.580***	1.843***	1.811***	1.892***	1.872***	1.346***	1.121***	1.164***	0.940***
	[0.142]	[0.153]	[0.143]	[0.154]	[0.262]	[0.268]	[0.243]	[0.263]	[0.262]	[0.266]	[0.220]	[0.208]
Marginal effect	1.725***	1.615***	1.906***	1.952**	1.842***	2.149***	1.664***	1.869***	4.046***	4.799***	4.232***	4.893***
	[0.601]	[0.627]	[0.650]	[0.759]	[0.582]	[0.671]	[0.588]	[0.652]	[0.959]	[1.071]	[0.958]	[1.159]
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5-year period dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Colonial history controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Sample	Old World	Old World	Global	Global	Old World	Old World	Global	Global	Old World	Old World	Global	Global
Observations	944	928	1,154	1,138	764	764	860	860	447	447	535	535
Countries	119	119	141	141	114	114	126	126	118	118	140	140
Time horizon	1960–2008	1960–2008	1960–2008	1960–2008	1960–2001	1960–2001	1960–2001	1960–2001	1990–2008	1990–2008	1990–2008	1990–2008
Time frequency	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly
Pseudo R^2	0.387	0.408	–	–	0.456	0.472	–	–	0.432	0.488	–	–

Notes: This table exploits variations in a quinquennially repeated cross-section of countries to establish a significant positive reduced-form impact of contemporary genetic diversity on the likelihood of observing the incidence of (i) a high-intensity overall (PRIO1000) civil war in any given 5-year interval during the 1960–2008 time horizon; (ii) a high-intensity ethnic (WM0609) civil war in any given 5-year interval during the 1960–2001 time horizon; and (iii) a low-intensity conflict involving nonstate actors in any given 5-year interval during the 1990–2008 time horizon, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict. To account for temporal dependence in conflict outcomes, all regressions control for the incidence of conflict in the previous 5-year interval, following Esteban, Mayoral and Ray (2012). For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of time-varying (lagged) colonial history controls includes variables that reflect the fraction of years from the previous 5-year interval that a country served as a colony of the U.K., France, and any other major colonizing power. The IV probit regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated marginal effect of a 1 percentage point increase in genetic diversity is the average marginal effect across the entire cross-section of observed diversity values, and it reflects the increase in the likelihood of a conflict incidence in any given 5-year interval, expressed in percentage points. Robust standard errors, clustered at the country level, are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

in Columns 4 and 8 suggest that, exploiting variations in a globally representative sample of countries, and conditional on our full set of baseline controls for geographical characteristics, institutional factors, ethnolinguistic fragmentation, and the correlates of economic development, a 1 percentage point increase in genetic diversity leads to an increase in the quinquennial likelihood of civil war incidence by 1.95 percentage points for PRIO1000 civil wars, and by 1.87 percentage points for WM0609 ethnic civil wars. Although these marginal effects are apparently more modest in comparison to their counterparts with respect to civil conflict incidence in Table 5, they are nevertheless quantitatively sizable, given that the incidence of a civil war is on average a much less likely event than the incidence of a civil conflict.

The conflict events considered by our empirical investigation thus far are of the type where government or state forces, on the one side of a conflict, fight against one or more internal armed opposition groups, on the other. Although conflicts of this type – namely, civil conflicts – tend to be both more severe and more prevalent in the conflict data, not all intrastate armed conflicts, broadly defined, involve government or state forces on either side of a conflict. In light of our priors, however, that the genetic diversity of a national population may well contribute to deeply rooted intergroup grievances – of a sociocultural, political, or economic nature – amongst various subnational groups, one expects such grievances to be manifested not only as armed conflicts between state and nonstate actors but also as armed conflicts that purely involve nonstate actors on either side of a given incompatibility.

Motivated by the aforementioned priors, the regressions presented in Columns 9–12 examine the quinquennial prevalence of one or more annual conflict events over the 1990–2008 time period that are categorized by the UCDP Non-State Conflict Dataset, Version 2.3-2010 (Sundberg, Eck and Kreutz, 2012) as episodes of "nonstate" conflicts – namely, conflicts that involve only nonstate armed opposition groups on either side and in which the use of armed force resulted in at least 25 annual battle-related casualties.[43] The results indicate that irrespective of the identification strategy employed or the set of covariates included in the specification, genetic diversity confers a statistically and economically significant impact on the temporal incidence of conflicts involving only nonstate actors. For instance, based on an estimated average marginal effect in a globally representative sample of countries, the IV probit model in Column 12 suggests that conditional on our complete set of baseline control variables, the quinquennial likelihood of a nonstate conflict incidence increases by 4.89 percentage points in response to a 1 percentage point increase in the genetic diversity of a contemporary national population.

Robustness to Exploiting Variations in Annually Repeated Cross-Country Data In spite of the fact that it is common practice to exploit variations in quinquennially repeated cross-country data when explaining conflict incidence (e.g., Montalvo and Reynal-Querol, 2005; Esteban, Mayoral and Ray, 2012), mainly in an effort to mitigate concerns regarding serial correlation in unobservables and the endogeneity of time-varying observables, this strategy comes at the cost of suppressing potentially valuable information on the continuation of hostilities across years within any given 5-year interval. To address concerns that ignoring such information could lead to an upward bias in our estimates of the influence of genetic diversity, Table A.10 in Appendix A replicates our baseline analysis of civil conflict incidence from Table 5 but does so by exploiting variations in annually repeated cross-country data.[44]

[43] Although the opposition groups in nonstate conflicts typically reflect an interethnic cleavage (e.g., the conflict between the Dizi and Surma peoples of Ethiopia), in a few cases of such conflict, they represent organized crime syndicates (e.g., the Sinaloa Cartel versus the Gulf Cartel in Mexico).

[44] In each of the regressions in this analysis, all time-varying covariates enter the specification as their lagged annual values (as opposed to their temporal means over the previous 5-year interval) and time fixed effects are accounted for by way of year (as opposed to 5-year period) dummies.

The results presented in the appendix table demonstrate that the impact of genetic diversity on the temporal incidence of either overall or ethnic civil conflict remains qualitatively unaltered under this alternative approach to organizing the temporal dimension of the data. In particular, exploiting variations across country-year observations from a globally representative cross-section, the average marginal effects associated with the IV probit regressions in Columns 4 and 8 of the appendix table suggest that conditional on our full set of controls for geographical conditions, institutional factors, ethnolinguistic fragmentation, and development outcomes, a 1 percentage point increase in genetic diversity leads to an increase in the annual likelihood of civil conflict incidence by 1.18 percentage points for PRIO25 civil conflicts, and by 0.484 percentage points for WCM09 ethnic civil conflicts.[45]

4.3 Analysis of Civil Conflict Onset in Repeated Cross-Country Data

Table 8 collects the results from our baseline analyses of overall and ethnic civil conflict onset in annually repeated cross-country data, spanning the post-1960 time horizon. Specifically, we examine the influence of genetic diversity on the annual likelihood of observing the outbreak of a new period of conflict, instigated by either (i) the eruption of a new episode of a (potentially recurrent) PRIO25 civil conflict, following at least two years of civil peace, as captured by the PRIO2 onset measure; or (ii) the emergence of a new PRIO25 civil conflict, regardless of the number of preceding years of civil peace, as reflected by the PRIO-NC onset measure. Corresponding to the latter definition of conflict onset, we also examine the influence of genetic diversity on the annual likelihood of observing the eruption of a new WCM09 ethnic civil conflict, thus bringing the number of different outcome variables explored by our current analysis of civil conflict onset to three.

Akin to our preceding analysis of civil conflict incidence, in the interest of brevity, we focus the presentation of our results from examining each onset measure on our better-identified estimates, obtained from either (i) probit regressions that restrict attention to variations in the Old World; or (ii) their corresponding IV probit counterparts that exploit global variations while instrumenting the genetic diversity of a country's contemporary national population with the migratory distance of its prehistorically indigenous settlements from East Africa. Furthermore, for each outcome variable examined and identification strategy employed, we present our findings from estimating two different specifications – namely, one that conditions the regression on only exogenous geographical covariates (including continent fixed effects) and another that partials out the influence of the full set of covariates considered by our baseline empirical model of conflict onset. In the latter case, to surmount the issue of contemporaneous bidirectional causality, all of our time-varying controls for institutional factors and development outcomes enter the specification with a one-year lag. In all our specifications, however, we include a complete set of year dummies, along with cubic splines

[45] Consistently with priors, these marginal effects are quantitatively smaller in comparison to those obtained by our baseline analysis of conflict incidence in quinquennially repeated cross-country data, given that the likelihood of a civil conflict incidence in a given year is expected to be considerably smaller than the likelihood of such an event in *any* year of a given 5-year interval.

Table 8: Genetic Diversity and the Onset of Civil Conflict in Annually Repeated Cross-Country Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Probit	Probit	IV Probit	IV Probit	Probit	Probit	IV Probit	IV Probit	Probit	Probit	IV Probit	IV Probit
	PRIO25 civil conflict onset [PRIO2]				New PRIO25 civil conflict onset [PRIO-NC]				New WCM09 ethnic civil conflict onset			
Genetic diversity (ancestry adjusted)	9.030***	10.688***	8.755***	10.947***	9.093***	10.300***	10.056***	11.560***	13.252***	14.710***	10.121***	10.885***
	[2.837]	[2.741]	[2.537]	[2.834]	[2.929]	[3.130]	[2.699]	[3.303]	[3.532]	[3.801]	[3.235]	[4.077]
Ethnic fractionalization		-0.047		-0.311		0.229		-0.116		0.603		0.158
		[0.244]		[0.252]		[0.262]		[0.277]		[0.411]		[0.455]
Ethnolinguistic polarization		0.229		0.416*		0.224		0.329		-0.117		0.428
		[0.243]		[0.241]		[0.258]		[0.258]		[0.340]		[0.427]
Absolute latitude	-15.296***	-11.706**	-17.765***	-16.589***	-9.425**	-4.188	-16.683***	-15.100**	-16.941***	-3.863	-21.208***	-16.527*
	[3.917]	[5.809]	[3.989]	[6.327]	[4.312]	[5.881]	[5.157]	[7.138]	[6.017]	[9.220]	[5.914]	[9.849]
Land area	20.704	19.552	-3.156	-4.676	49.107*	51.031*	18.510	16.347	10.142	24.585	0.179	7.044
	[25.582]	[27.682]	[29.432]	[32.918]	[26.748]	[29.072]	[32.886]	[37.836]	[28.254]	[35.315]	[36.602]	[47.207]
Ruggedness	0.995	1.240*	0.704	0.788	0.676	0.843	0.587	0.564	1.108	1.328	0.035	-0.443
	[0.762]	[0.734]	[0.681]	[0.700]	[0.791]	[0.816]	[0.722]	[0.795]	[0.917]	[0.987]	[1.056]	[1.305]
Mean elevation	-0.549***	-0.634***	-0.449***	-0.541***	-0.596***	-0.612***	-0.496***	-0.527***	-0.464**	-0.549**	-0.019	0.074
	[0.193]	[0.192]	[0.166]	[0.173]	[0.190]	[0.191]	[0.164]	[0.171]	[0.218]	[0.217]	[0.274]	[0.317]
Range of elevation	0.165***	0.128*	0.183***	0.163**	0.190***	0.126**	0.205***	0.165**	0.088	-0.010	0.021	-0.076
	[0.048]	[0.070]	[0.051]	[0.069]	[0.044]	[0.056]	[0.054]	[0.069]	[0.056]	[0.070]	[0.079]	[0.100]
Mean land suitability	-0.108	-0.163	-0.088	-0.139	0.497**	0.389	0.229	0.113	0.405	0.470	0.178	0.354
	[0.229]	[0.285]	[0.225]	[0.270]	[0.217]	[0.271]	[0.234]	[0.287]	[0.306]	[0.406]	[0.358]	[0.451]
Range of land suitability	0.696***	0.798***	0.633***	0.864***	0.179	0.085	0.262	0.291	1.000***	0.999**	1.127***	1.299**
	[0.219]	[0.234]	[0.211]	[0.263]	[0.190]	[0.240]	[0.204]	[0.266]	[0.317]	[0.476]	[0.350]	[0.512]
Distance to nearest waterway	0.125	0.121	0.155	0.142	0.147	0.122	0.160	0.134	0.320*	0.264	0.188	0.138
	[0.145]	[0.164]	[0.134]	[0.160]	[0.149]	[0.176]	[0.132]	[0.164]	[0.185]	[0.214]	[0.193]	[0.212]
Executive constraints, lagged		0.095***		0.086**		0.079		0.072		0.129**		0.100*
		[0.036]		[0.036]		[0.049]		[0.046]		[0.057]		[0.056]
Democracy dummy, lagged		-0.297**		-0.427***		-0.302*		-0.381**		-0.409**		-0.487***
		[0.150]		[0.150]		[0.181]		[0.181]		[0.199]		[0.182]
Autocracy dummy, lagged		-0.120		-0.178		-0.261*		-0.283**		-0.109		-0.273*
		[0.132]		[0.123]		[0.143]		[0.135]		[0.145]		[0.157]
Log oil production per capita, lagged		0.038**		0.033*		0.022		0.018		0.034		0.030
		[0.018]		[0.019]		[0.021]		[0.022]		[0.025]		[0.026]
Log population, lagged		0.008		0.001		0.037		0.023		0.017		0.011
		[0.049]		[0.049]		[0.040]		[0.046]		[0.098]		[0.100]
Log GDP per capita, lagged		-0.252***		-0.222***		-0.205**		-0.166**		-0.395***		-0.262**
		[0.072]		[0.068]		[0.083]		[0.078]		[0.101]		[0.133]
Conflict incidence, lagged					-0.281	-0.297*	-0.177	-0.192	-0.318*	-0.352**	-0.179	-0.184
					[0.172]	[0.167]	[0.165]	[0.162]	[0.169]	[0.172]	[0.189]	[0.197]
Marginal effect	0.680***	0.794***	0.632***	0.784***	0.455***	0.508***	0.492***	0.560***	0.620***	0.669***	0.563***	0.677**
	[0.226]	[0.226]	[0.205]	[0.239]	[0.153]	[0.163]	[0.157]	[0.191]	[0.181]	[0.187]	[0.210]	[0.278]
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peace duration cubic splines	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Colonial history dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Sample	Old World	Old World	Global	Global	Old World	Old World	Global	Global	Old World	Old World	Global	Global
Observations	4,376	4,354	5,531	5,508	3,849	3,828	4,896	4,874	3,607	3,585	4,038	4,016
Countries	119	119	141	141	119	119	141	141	117	117	129	129
Time horizon	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2005	1960–2005	1960–2005	1960–2005
Time frequency	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual
Pseudo R^2	0.152	0.171	–	–	0.120	0.141	–	–	0.146	0.180	–	–

Notes: This table exploits variations in an annually repeated cross-section of countries to establish a significant positive reduced-form impact of contemporary genetic diversity on the likelihood of observing the onset of (i) a new or recurring episode of an overall (PRIO2) civil conflict, following two or more years of uninterrupted peace, in any given year during the 1960–2008 time horizon; (ii) a new overall (PRIO-NC) civil conflict in any given year during the 1960–2008 time horizon; and (iii) a new ethnic (WCM09) civil conflict in any given year during the 1960–2005 time horizon, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict. To account for duration and temporal dependence in conflict outcomes, all regressions control for a set of cubic splines of the number of peace years, following Beck, Katz and Tucker (1998). In addition, with the exception of regressions explaining PRIO2 onset, for which a mechanical correlation with conflict incidence in the previous year would follow by definition, all regressions control for the lagged incidence of conflict, following Esteban, Mayoral and Ray (2012). For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of time-varying (lagged) colonial history dummies includes indicators for whether a country was a colony of the U.K., France, and any other major colonizing power in the previous year. The IV probit regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated marginal effect of a 1 percentage point increase in genetic diversity is the average marginal effect across the entire cross-section of observed diversity values, and it reflects the increase in the likelihood of a conflict onset in any given year, expressed in percentage points. Robust standard errors, clustered at the country level, are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

of the number of preceding years of civil peace, in order to account for temporal and duration dependence in conflict processes.[46]

[46] As in our analysis of conflict incidence, depending on the outcome variable, we include the lagged incidence of either overall or ethnic civil conflict as a standard control variable in all our onset regressions, with the exception of those that examine the PRIO2 onset measure, because in this particular case, *by definition*, the absence of an ongoing

The results presented in Table 8 demonstrate that irrespective of the specific measure of conflict onset examined, the identification strategy employed, or the set of covariates considered by the specification, genetic diversity confers a highly statistically significant and qualitatively robust positive influence on the annual likelihood of civil conflict outbreaks.[47] To elucidate the economic significance of this impact in a globally representative sample of countries, the estimated average marginal effects associated with the IV probit regressions in Columns 4, 8, and 12 suggest that, accounting for the influence of geographical conditions, institutional factors, ethnolinguistic fragmentation, and development outcomes, a 1 percentage point increase in genetic diversity leads to an increase in the annual likelihood of a PRIO25 civil conflict outbreak by 0.784 and 0.560 percentage points, as reflected by the PRIO2 and PRIO-NC onset measures, respectively, and it leads to an increase in the annual likelihood of a new WCM09 ethnic civil conflict eruption by 0.677 percentage points.[48]

The economically significant role of genetic diversity as a contributor to the outbreak of civil conflict is also evident in the plots presented in Figure 4. Specifically, based on the IV probit regressions from Columns 4, 8, and 12, the figure illustrates how the *predicted* likelihood associated with each of the three conflict onset measures responds as one moves along the global cross-country genetic diversity distribution in the relevant estimation sample.[49] According to these plots, in response to a move from the 10th to the 90th percentile of the cross-country genetic diversity distribution, the predicted annual likelihood of a PRIO2 onset event rises from 1.77 percent to 6.72 percent (Panel A), that of a PRIO-NC onset event rises from 0.834 percent to 4.23 percent (Panel B), and that of a new WCM09 ethnic civil conflict outbreak rises from 1.45 percent to 4.93 percent (Panel C).

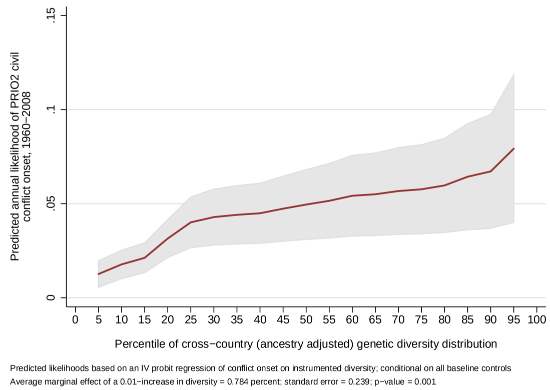
Robustness to Additional Correlates of Conflict Onset As in our baseline analysis of conflict incidence, our main specifications for examining the onset of civil conflict ignored a few potentially important covariates, reflecting our objective to maximize the number of observations in our baseline estimation samples, which would otherwise have been constrained by the more limited availability of data on these additional control variables. The results presented in Table 9, however, confirm that our key findings regarding the influence of genetic diversity on conflict onset remain unaltered when the most stringent specifications from our baseline analysis in Table 8 are augmented to additionally account for the influence of these other control variables, deemed by

conflict in the previous year is a necessary (but not sufficient) condition for observing the outbreak of conflict in the current year.

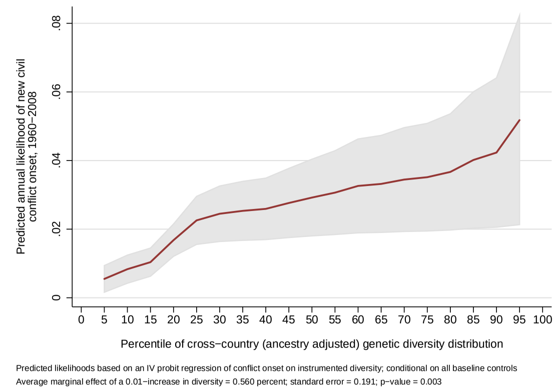
[47] As confirmed by the results presented in Table A.12 in Appendix A, all of our main findings from the probit regressions in Table 8 – that identify the impact of genetic diversity on civil conflict onset by restricting attention to variations in the Old World – remain qualitatively unaltered when employing either classical or "rare events" logit models of conflict onset.

[48] Not surprisingly, in light of the fact that the likelihood of observing a civil conflict outbreak in any given year will be smaller than the likelihood of observing either a conflict outbreak or an ongoing conflict in that year, these marginal effects are noticeably smaller in magnitude when compared to those obtained by our robustness analysis of conflict incidence in annually repeated cross-country data, as reported in Table A.10 in Appendix A.

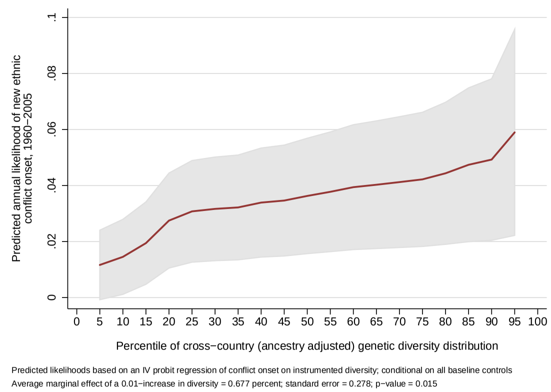
[49] Similarly, based on the probit regressions from Columns 2, 6, and 10, the plots presented in Figure A.2 in Appendix A depict how the predicted annual likelihoods associated with the three conflict onset measures respond as one moves along the cross-country genetic diversity distribution in the Old World.



(A) Effect on PRIO2 overall civil conflict onset



(B) Effect on new overall civil conflict onset



(C) Effect on new ethnic civil conflict onset

FIGURE 4: The Effect of Instrumented Genetic Diversity on the Annual Likelihood of Civil Conflict Onset in the Global Sample

Notes: This figure depicts the influence of contemporary genetic diversity at the country level on the *predicted* likelihood of observing the onset of (i) a new or recurring episode of an overall (PRIO2) civil conflict, following two or more years of uninterrupted peace, in any given year during the 1960–2008 time horizon [Panel A]; (ii) a new overall (PRIO-NC) civil conflict in any given year during the 1960–2008 time horizon [Panel B]; and (iii) a new ethnic (WCM09) civil conflict in any given year during the 1960–2005 time horizon [Panel C], conditional on other well-known diversity measures, the proximate geographical, institutional, and development-related correlates of conflict, controls for temporal dependence in conflict outcomes, and continent and year dummies. In each panel, the predicted likelihood of conflict onset is illustrated as a function of the percentile of the cross-country genetic diversity distribution, and the prediction is based on the relevant IV probit regression from Table 8, exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity, and conducted using the global sample of countries and the full set of covariates considered by the analysis of the conflict outcome in question. The shaded area in each plot reflects the 95-percent confidence-interval region of the depicted relationship.

Table 9: Genetic Diversity and the Onset of Civil Conflict in Annually Repeated Cross-Country Data – Robustness to Additional Correlates of Conflict Onset

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Probit	Probit	IV Probit	IV Probit	Probit	Probit	IV Probit	IV Probit	Probit	Probit	IV Probit	IV Probit
	PRIO25 civil conflict onset [PRIO2]				New PRIO25 civil conflict onset [PRIO-NC]				New WCM09 ethnic civil conflict onset			
Genetic diversity (ancestry adjusted)	7.879**	8.003**	8.266**	8.274**	8.280**	7.816**	10.516**	9.831**	13.203***	15.209***	10.322**	11.447**
	[3.548]	[3.723]	[3.615]	[3.803]	[3.871]	[3.923]	[4.146]	[4.216]	[4.192]	[5.000]	[4.375]	[4.880]
Ethnic dominance		-0.005		0.014		0.091		0.108		-0.205		0.013
		[0.118]		[0.104]		[0.134]		[0.120]		[0.144]		[0.158]
Political instability, lagged		0.112		0.124		0.203		0.138		0.254*		0.122
		[0.113]		[0.096]		[0.127]		[0.121]		[0.146]		[0.138]
New state dummy, lagged		0.263		0.008		0.194		-0.137		1.182**		0.962**
		[0.544]		[0.512]		[0.538]		[0.524]		[0.510]		[0.477]
Marginal effect	0.655**	0.665**	0.642**	0.642**	0.490**	0.461*	0.597**	0.551**	0.643***	0.730***	0.681**	0.743**
	[0.311]	[0.326]	[0.307]	[0.323]	[0.239]	[0.240]	[0.274]	[0.271]	[0.221]	[0.262]	[0.310]	[0.333]
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Old World	Old World	Global	Global	Old World	Old World	Global	Global	Old World	Old World	Global	Global
Observations	2,991	2,991	3,903	3,903	2,561	2,561	3,368	3,368	2,676	2,676	3,007	3,007
Countries	95	95	116	116	95	95	116	116	95	95	106	106
Time horizon	1960–1999	1960–1999	1960–1999	1960–1999	1960–1999	1960–1999	1960–1999	1960–1999	1960–1999	1960–1999	1960–1999	1960–1999
Time frequency	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual
Pseudo R^2	0.169	0.170	–	–	0.141	0.144	–	–	0.185	0.197	–	–

Notes: This table exploits variations in an annually repeated cross-section of countries to establish that the significant positive reduced-form impact of contemporary genetic diversity on the likelihood of observing the onset of (i) a new or recurring episode of an overall (PRIO2) civil conflict, following two or more years of uninterrupted peace, in any given year during the 1960–1999 time horizon; (ii) a new overall (PRIO-NC) civil conflict in any given year during the 1960–1999 time horizon; and (iii) a new ethnic (WCM09) civil conflict in any given year during the 1960–1999 time horizon, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict, is robust to accounting for an additional distributional index of diversity and for additional institutional correlates of conflict. All regressions control for the entire set of covariates considered by the baseline analysis of conflict onset in Table 8, with each of the odd-numbered columns providing the relevant baseline for the robustness check in the subsequent even-numbered column, given that both regressions are conducted using the same sample, restricted by the availability of data on the additional control variables from the robustness exercise. The IV probit regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated marginal effect of a 1 percentage point increase in genetic diversity is the average marginal effect across the entire cross-section of observed diversity values, and it reflects the increase in the likelihood of a conflict onset in any given year, expressed in percentage points. Robust standard errors, clustered at the country level, are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

previous studies (e.g., Hegre and Sambanis, 2006) to be statistically robust correlates of conflict onset. In particular, motivated by priors that the influence of these other correlates may not be fully accounted for by our baseline controls for ethnolinguistic fragmentation and institutional factors, the additional covariates considered by our robustness analysis in Table 9 comprise (i) a time-invariant indicator of "ethnic dominance" from the study of Collier and Hoeffler (2004), reflecting whether the majority ethnolinguistic group of a country comprises between 45 percent and 90 percent of its contemporary national population; and (ii) two time-varying indicators of institutional volatility from the study of Fearon and Laitin (2003), reflecting (a) whether a country is a newly independent state and (b) whether it is politically unstable, as captured by any observed change in its polity score over the preceding three years.[50] According to the civil conflict literature, because these covariates either proxy for a higher risk of ethnopolitical grievances amongst minority groups in the national population, as is the case for the "ethnic dominance" measure, or reflect the susceptibility of a weak state apparatus to violent insurgencies, as is the case for the indicators of institutional instability, all three covariates are expected to contribute to the temporal hazard of civil conflict outbreaks in society.

The results from estimating our augmented specifications – i.e., one for each onset measure examined and identification strategy considered by our baseline analysis in Table 8 – are revealed

[50]Following our convention for mitigating the issue of endogeneity with respect to contemporaneous time-varying covariates, the two indicators of institutional volatility from Fearon and Laitin (2003) enter our robustness specifications in Table 9 with a one-year lag.

in the even-numbered columns of Table 9. To be sure, all regressions include our full set of baseline covariates (not reported in the table to conserve space). In addition, given that the introduction of the additional control variables to the augmented specifications leads to a reduction in the number of observations, to permit fair assessments of the robustness of our coefficient of interest, the odd-numbered columns of the table present the results from estimating the corresponding baseline specifications – i.e., holding fixed the size of the relevant estimation sample. As is apparent from the results, regardless of the specific outcome variable examined or the identification strategy employed, the statistically and economically significant influence of genetic diversity on civil conflict onset remains largely intact (and even increases in magnitude in some cases) when additionally subjected to the aforementioned three covariates. In contrast, the indicator of "ethnic dominance" does not enter any of the augmented specifications with a significant coefficient, whereas the indicators of institutional instability only appear as statistically significant predictors of conflict onset in the case of new WCM09 ethnic civil conflict outbreaks.

4.4 Analysis of Intrastate Conflict Severity in Repeated Cross-Country Data

Our findings thus far establish that the genetic diversity of a contemporary national population is a robust and significant reduced-form contributor to the risk of civil conflict in society, as manifested by the frequency, prevalence, and emergence of both overall and ethnic civil conflict events over the past half-century. Broadly speaking, these results reflect the influence of genetic diversity on the *extensive* margin of conflict, given that the outcome variables employed by our analysis until now have all been based on binary measures that are subject to a predefined threshold of annual battle-related casualties being surpassed for the identification of civil conflict events. Although we have already shown that our results are not qualitatively sensitive to the adoption of alternative definitions of this extensive margin of conflict (e.g., the incidence of PRIO25 versus PRIO1000 civil conflict events in quinquennially repeated cross-country data), our analysis in this section employs both ordinal and continuous measures that capture the "severity" of intrastate conflicts, in order to establish the influence of genetic diversity on the *intensive* margin of conflict in quinquennially repeated cross-country data.

The first measure of conflict intensity that we examine exploits information on the apparent "magnitude scores" associated with "major episodes" of intrastate armed conflict, as reported by the Major Episodes of Political Violence (MEPV) data set (Marshall, 2010).[51] According to this data set, a "major episode" of armed conflict involves both (i) a minimum of 500 directly related fatalities in total; and (ii) systematic violence at a sustained rate of at least 100 directly related casualties per year. Importantly, for each such episode of conflict, the MEPV data set provides a "magnitude score" – namely, an ordinal measure on a scale of 1 to 10 of the episode's destructive impact on the directly affected society, incorporating information on multiple dimensions of conflict severity,

[51] The specific version of the MEPV data set that we employ provides annual information for a total of 175 countries over the 1946–2008 time period. See http://www.systemicpeace.org/warlist.htm for further details on our measure of conflict intensity from the MEPV data set.

including the capabilities of the state, the interactive intensity (means and goals) of the oppositional actors, the area and scope of death and destruction, the extent of population displacement, and the duration of the episode. The specific outcome variable from the MEPV data set that we employ reflects the aggregated magnitude score across all conflict episodes that are classified as one of four types of intrastate conflict – namely, civil war, civil violence, ethnic war, and ethnic violence.[52] In particular, this variable is reported by the MEPV data set at the country-year level, with nonevent years for a country being coded as 0. Moreover, given that our analysis of conflict severity follows Esteban, Mayoral and Ray (2012) in terms of exploiting variations in quinquennially repeated cross-country data, for each country in our sample, we collapse the annual data on conflict intensity from the MEPV data set to a quinquennial time series, by assigning to any given 5-year interval in our post-1960 sample period, the maximum level of conflict intensity reflected by the measure across all years in that 5-year interval.

Our second measure of conflict intensity is based on annual time-series data on a continuous index of social conflict at the country level, as reported by the Cross-National Time-Series (CNTS) Data Archive (Banks, 2010). Rather than adopting an ad hoc fatality-related threshold for the identification of conflict events, this index provides an aggregate summary of the general level of social discordance in any given country-year, by way of presenting a weighted average, following the methodology of Rummel (1963), across all observed occurrences of eight different types of sociopolitical unrest, including assassinations, general strikes, guerrilla warfare, major government crises, political purges, riots, revolutions, and anti-government demonstrations.[53] As with our measure of conflict severity from the MEPV data set, we convert the annual time series of the continuous index of social conflict for a given country to its quinquennial counterpart, by selecting the maximum annual value attained by the index for that country in any given 5-year interval in our sample period.

Table 10 reveals the results from our analysis of the influence of genetic diversity on intrastate conflict severity – as reflected by either the MEPV aggregate magnitude score of conflict intensity (Columns 1–4) or the CNTS index of social conflict (Columns 5–8) – in quinquennially repeated cross-country data. We mimic our previous analyses of the temporal incidence and onset of civil conflict by presenting our better-identified estimates. Specifically, for each outcome variable, the first two columns collect our results from OLS regressions that focus attention to variations in a sample composed of only countries belonging to the Old World, whereas the latter two columns

[52] Specifically, all episodes of intrastate conflict in the MEPV data set are categorized along two dimensions. With respect to the first dimension, an episode may be considered either (i) one of "civil" conflict, involving rival political groups; or (ii) one of "ethnic" conflict, involving the state agent and a distinct ethnic group. In terms of the second dimension, however, an episode may be either (i) one of "violence," involving the use of instrumental force, without necessarily possessing any exclusive goals; or (ii) one of "war," involving violent activities between distinct groups, with the intent to impose a unilateral result to the contention.

[53] The specific weights (reported in parentheses) assigned to the different types of sociopolitical unrest considered by the index are as follows: assassinations (25), general strikes (20), guerrilla warfare (100), major government crises (20), political purges (20), riots (25), revolutions (150), and anti-government demonstrations (10). For further details, the reader is referred to the codebook of the CNTS data archive, available at http://www.databanksinternational.com/32.html.

Table 10: Genetic Diversity and the Severity of Intrastate Conflict in Quinquennially Repeated Cross-Country Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	OLS	2SLS	2SLS	OLS	OLS	2SLS	2SLS
	MEPV civil conflict severity				CNTS index of social conflict			
Genetic diversity (ancestry adjusted)	3.897**	5.102**	3.741**	4.516**	6.182**	9.126***	5.644**	7.611***
	[1.854]	[2.037]	[1.769]	[2.019]	[2.912]	[2.815]	[2.760]	[2.912]
Ethnic fractionalization		0.017		-0.066		-0.186		-0.204
		[0.112]		[0.110]		[0.267]		[0.216]
Ethnolinguistic polarization		-0.109		0.033		-0.012		0.209
		[0.107]		[0.119]		[0.181]		[0.180]
Absolute latitude	-7.614***	-6.573**	-8.482***	-8.117***	-7.100	-3.802	-8.525**	-6.432
	[2.872]	[3.149]	[2.093]	[2.566]	[4.961]	[7.057]	[3.626]	[5.148]
Land area	9.232	14.294	-3.037	-8.511	13.922	15.241	15.128	-3.905
	[11.867]	[10.647]	[9.966]	[12.850]	[17.501]	[14.116]	[14.566]	[17.104]
Ruggedness	-0.063	0.211	-0.080	-0.015	0.335	1.084*	-0.028	0.304
	[0.281]	[0.296]	[0.251]	[0.276]	[0.508]	[0.550]	[0.469]	[0.503]
Mean elevation	-0.150	-0.157*	-0.117	-0.112	-0.328**	-0.424***	-0.174	-0.228**
	[0.095]	[0.087]	[0.075]	[0.074]	[0.137]	[0.123]	[0.115]	[0.107]
Range of elevation	0.105***	0.067**	0.098***	0.074***	0.144***	0.054	0.144***	0.089*
	[0.031]	[0.030]	[0.025]	[0.028]	[0.052]	[0.049]	[0.038]	[0.048]
Mean land suitability	0.136	-0.035	0.165	0.029	0.470**	-0.013	0.395**	0.038
	[0.119]	[0.114]	[0.103]	[0.113]	[0.194]	[0.177]	[0.171]	[0.175]
Range of land suitability	0.159	0.038	0.122	0.052	0.348*	0.140	0.300*	0.164
	[0.117]	[0.135]	[0.100]	[0.127]	[0.201]	[0.240]	[0.170]	[0.213]
Distance to nearest waterway	0.066	0.049	0.096	0.104	0.025	-0.010	-0.039	-0.022
	[0.075]	[0.077]	[0.060]	[0.073]	[0.101]	[0.100]	[0.079]	[0.096]
Average executive constraints, lagged		0.008		0.012		0.000		0.005
		[0.027]		[0.026]		[0.048]		[0.041]
Fraction of years under democracy, lagged		0.105		-0.005		-0.204		-0.386**
		[0.098]		[0.097]		[0.203]		[0.175]
Fraction of years under autocracy, lagged		-0.076		-0.107		-0.220*		-0.361***
		[0.087]		[0.082]		[0.113]		[0.110]
Log average oil production per capita, lagged		0.002		-0.002		-0.031*		-0.020
		[0.011]		[0.010]		[0.016]		[0.014]
Log average population, lagged		0.040*		0.036		0.117***		0.109***
		[0.023]		[0.022]		[0.033]		[0.033]
Log average GDP per capita, lagged		-0.078*		-0.058		-0.151*		-0.090
		[0.042]		[0.039]		[0.079]		[0.062]
Conflict severity, lagged	0.672***	0.661***	0.682***	0.670***	0.259***	0.212**	0.314***	0.276***
	[0.035]	[0.038]	[0.030]	[0.031]	[0.089]	[0.084]	[0.077]	[0.073]
Effect of increasing genetic diversity	0.209**	0.274**	0.252**	0.304**	0.332**	0.490***	0.379**	0.512***
from the 10th to the 90th percentile	[0.100]	[0.109]	[0.119]	[0.136]	[0.156]	[0.151]	[0.186]	[0.196]
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5-year period dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	Yes	No	Yes	No	Yes	No	Yes
Colonial history controls	No	Yes	No	Yes	No	Yes	No	Yes
Sample	Old World	Old World	Global	Global	Old World	Old World	Global	Global
Observations	944	944	1,173	1,173	942	942	1,171	1,171
Countries	119	119	143	143	119	119	143	143
Time horizon	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008
Time frequency	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly
Partial R^2 of genetic diversity	0.005	0.008	–	–	0.007	0.012	–	–
Partial R^2 sum of other diversity measures	–	0.001	–	–	–	0.001	–	–
First-stage adjusted R^2	–	–	0.769	0.791	–	–	0.769	0.791
First-stage partial R^2 of migratory distance	–	–	0.449	0.416	–	–	0.446	0.413
First-stage F statistic	–	–	180.390	113.647	–	–	170.303	105.199
Adjusted R^2	0.597	0.598	–	–	0.217	0.233	–	–

Notes: This table exploits variations in a quinquennially repeated cross-section of countries to establish a significant positive reduced-form impact of contemporary genetic diversity on the severity of conflict, as reflected by (i) the maximum value of an annual ordinal index of conflict intensity (from the MEPV data set) across all years in any given 5-year interval during the 1960-2008 time period; and (ii) the maximum value of an annual continuous index of the degree of social unrest (from the CNTS data set) across all years in any given 5-year interval during the 1960-2008 time period, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict. Given that both measures of conflict severity are expressed in units that have no natural interpretation, their intertemporal cross-country distributions are standardized prior to conducting the regression analysis. To account for temporal dependence in conflict outcomes, all regressions control for the severity of conflict in the previous 5-year interval, following Esteban, Mayoral and Ray (2012). For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of time-varying (lagged) colonial history controls includes variables that reflect the fraction of years from the previous 5-year interval that a country served as a colony of the U.K., France, and any other major colonizing power. The 2SLS regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated effect associated with increasing genetic diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of standard deviations of the intertemporal cross-country distribution of conflict severity. Robust standard errors, clustered at the country level, are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

present our findings from 2SLS regressions that exploit variations in a globally representative sample of countries while employing the migratory distance of a country's prehistorically indigenous settlements from East Africa as a plausibly exogenous source of variation for the genetic diversity of its contemporary national population.[54] As before, for each of our identification strategies and for each proxy for conflict intensity examined, we estimate two alternative specifications – namely, one that conditions the regression on a set of only exogenous geographical covariates (including continent fixed effects), and another that partials out the influence of our complete set of baseline covariates, capturing geographical conditions, institutional factors, ethnolinguistic fragmentation, and development outcomes. In all regressions, we account for temporal dependence in conflict severity by allowing both the lagged observation of the outcome variable and a full set of time-interval (5-year period) dummies to enter the specification, and as always, whenever time-varying covariates are allowed to enter the specification, they do so with a one-period lag. Finally, in light of the fact that the units in which either of our proxies for conflict intensity are measured in the data have no natural interpretation, we standardize both outcome variables prior to conducting our regression analyses.

Notwithstanding the measure for conflict intensity examined, the identification strategy exploited, or the set of covariates considered by the specification, the results from our analysis of conflict severity in Table 10 establish genetic diversity as a qualitatively robust and statistically significant reduced-form contributor to the intensive margin of intrastate conflict. In terms of the economic significance of its estimated influence, depending on the identification strategy employed, the regressions presented in the even-numbered columns suggest that conditional on our full set of controls for geographical characteristics, institutional factors, ethnolinguistic fragmentation, and the correlates of economic development, a move from the 10th to the 90th percentile of the cross-country genetic diversity distribution in the relevant sample leads to an increase in conflict severity by between 27.4 percent and 30.4 percent of a standard deviation from the observed distribution of the MEPV magnitude score of conflict intensity, and by 49 percent to 51.2 percent of a standard deviation from the observed distribution of the CNTS index of social conflict.

4.5 Analysis of Intragroup Factional Conflict Incidence in Cross-Country Data

One crucial dimension in which our measure of intrapopulation diversity at the national level adds value beyond all known indices of ethnolinguistic fragmentation, which necessarily impose *intragroup* homogeneity for all ethnolinguistic groups in a country's population, is that genetic diversity incorporates information on interpersonal heterogeneity not only across group boundaries but within such boundaries as well. As such, from a conceptual viewpoint alone, and in contrast to

[54] Despite the fact that our measure of conflict intensity from the MEPV data set is ordinal rather than continuous in nature, we choose to pursue least-squares (as opposed to maximum-likelihood) estimation methods when examining this particular outcome variable, primarily because this permits us to conveniently exploit both of our identification strategies. Specifically, although we are able to qualitatively replicate our key findings from Columns 1–2 using ordered probit rather than OLS regressions (results not shown), the absence (to our knowledge) of a readily available IV counterpart of the ordered probit regression model precludes conducting a similar robustness check on our key findings from Columns 3–4.

measures that capture the degree of ethnolinguistic fragmentation of a national population, to the extent that interpersonal heterogeneity can be expected to give rise to social, political, and economic grievances that culminate to violent contentions even across ethnically or linguistically homogenous subgroups, our measure is naturally better-suited to empirically link intrapopulation diversity with the incidence of such forms of conflict in society. Our analysis in this section elucidates precisely this virtue of our measure, by exploiting cross-country variations to establish genetic diversity as a statistically and economically significant predictor of the likelihood of observing the incidence of one or more *intragroup* factional conflict events during the 1990–1999 time period.

The primary source of our data on the incidence of intragroup factional conflict events across the globe is the Minorities at Risk (MAR), Phase IV data set (Minorities at Risk Project, 2009), which provides for each country with a national population of at least half a million, information on each subnational (i.e., nonstate communal) group that is considered a "minority at risk" – namely, an ethnopolitical group that (i) collectively suffers or benefits from systematic discriminatory treatment vis-à-vis other groups in the national population; and/or (ii) collectively mobilizes resources in defense or promotion of its self-defined interests. Specifically, for each such subnational group, the MAR data set furnishes an indicator for whether the group experienced any intragroup factional conflict event during the 1990–1999 time horizon. For our purposes, we simply aggregate this information to the country level, by coding a binary variable that reflects whether any of the MAR groups within a given country had an experience with intragroup factional conflict over this time span. To be sure, because the MAR data set does not provide information on the specific timing of intragroup factional conflict events, beyond the fact that they occurred at some point in the 1990–1999 time interval, we are restricted by the data to conduct our analysis in a cross-country framework, rather than in a repeated cross-country sample.

The results from our cross-country analysis of the influence of genetic diversity on the likelihood of observing the incidence of one or more intragroup factional conflict events during the 1990–1999 time period are collected in Table 11. As always, in the interest of keeping our exposition succinct, we concentrate our analysis on regressions that yield our better-identified estimates – namely, either (i) probit regressions (Columns 1–3) that exploit variations in a sample comprised of countries from the Old World; or (ii) their corresponding IV probit counterparts (Columns 4–6) that exploit worldwide variations across countries while instrumenting the genetic diversity of a country's modern-day national population with the migratory distance of its prehistorically native settlements from East Africa. For each of our two identification strategies, however, we now present the results from estimating three alternative specifications. The first two of these specifications follow from our expositional methodology in previous sections, in that one conditions the analysis on only exogenous geographical covariates (including continent fixed effects), whereas the other partials out the influence of our full set of baseline controls for geographical characteristics, institutional factors, ethnolinguistic fragmentation, and development outcomes. Our current analysis, however, introduces a third specification that augments our full baseline empirical model with additional controls for the total number and total share of all MAR groups in the national population. This

Table 11: Genetic Diversity and the Incidence of Intragroup Factional Conflict across Countries

	(1) Probit	(2) Probit	(3) Probit	(4) IV Probit	(5) IV Probit	(6) IV Probit
	MAR intragroup conflict incidence in the 1990–1999 time period					
Genetic diversity (ancestry adjusted)	25.761**	35.605**	48.598**	27.034***	41.653***	48.786***
	[11.897]	[16.944]	[19.653]	[9.051]	[12.430]	[12.094]
Ethnic fractionalization		-1.608	-2.998**		-0.891	-1.817*
		[1.284]	[1.402]		[1.008]	[1.000]
Ethnolinguistic polarization		2.098*	2.189*		3.347***	3.307***
		[1.113]	[1.161]		[0.875]	[0.855]
Absolute latitude	-48.592***	-73.897*	-106.026**	-49.501***	-81.253***	-94.303***
	[17.109]	[38.928]	[44.147]	[13.877]	[25.045]	[22.905]
Land area	143.999*	75.698	-78.668	125.611	-9.352	-102.355
	[81.548]	[95.597]	[135.361]	[117.771]	[118.627]	[136.216]
Ruggedness	-2.600	-1.762	-2.175	-3.080	-2.693	-2.905
	[2.430]	[2.464]	[2.599]	[2.100]	[2.204]	[2.151]
Mean elevation	0.547	0.417	0.067	0.687	0.773	0.515
	[0.608]	[0.586]	[0.673]	[0.503]	[0.518]	[0.543]
Range of elevation	0.286	-0.006	0.220	0.214	-0.018	0.077
	[0.186]	[0.268]	[0.360]	[0.177]	[0.233]	[0.265]
Mean land suitability	0.955	-0.721	-1.526	0.699	0.512	-0.081
	[0.935]	[1.173]	[1.400]	[0.826]	[1.068]	[1.092]
Range of land suitability	1.068	2.358**	1.540	1.180	1.634*	0.892
	[0.918]	[1.088]	[1.158]	[0.746]	[0.853]	[0.867]
Distance to nearest waterway	-0.150	0.312	0.599	0.006	0.635	0.823**
	[0.453]	[0.466]	[0.467]	[0.395]	[0.442]	[0.412]
Executive constraints, 1990–1999 average		-0.096	-0.075		0.065	0.113
		[0.250]	[0.260]		[0.232]	[0.232]
Fraction of years under democracy, 1990–1999		1.272	1.316		0.765	0.694
		[0.881]	[0.917]		[0.868]	[0.916]
Fraction of years under autocracy, 1990–1999		-0.956	-1.474*		-0.614	-0.840
		[0.703]	[0.773]		[0.601]	[0.601]
Log oil production per capita, 1990–1999 average		-0.050	-0.126		0.036	-0.014
		[0.115]	[0.116]		[0.086]	[0.078]
Log population, 1990–1999 average		0.325	0.476*		0.373*	0.448**
		[0.253]	[0.281]		[0.207]	[0.216]
Log GDP per capita, 1990–1999 average		-0.415	-0.330		-0.156	-0.104
		[0.331]	[0.359]		[0.251]	[0.244]
Number of minority groups			0.275			0.265
			[0.227]			[0.188]
Population share of minority groups			1.012			0.758
			[1.112]			[0.858]
Marginal effect	7.853**	8.368**	10.910***	8.130***	10.877***	12.288***
	[3.322]	[3.744]	[4.091]	[2.517]	[3.271]	[3.013]
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	Yes	Yes	No	Yes	Yes
Colonial history dummies	No	Yes	Yes	No	Yes	Yes
Sample	Old World	Old World	Old World	Global	Global	Global
Observations	84	84	84	103	103	103
Pseudo R^2	0.226	0.397	0.429	–	–	–

Notes: This table exploits cross-country variations to establish a significant positive reduced-form impact of contemporary genetic diversity on the likelihood of observing one or more factional conflicts *within* the "minorities at risk" (MAR) groups of a country's population in the 1990–1999 time period, conditional on other well-known diversity measures, the proximate geographical, institutional, and development-related correlates of conflict, and measures capturing the distribution of MAR groups in the national population. For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of colonial history dummies includes indicators for whether a country was ever a colony of the U.K., France, and any other major colonizing power. The IV probit regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated marginal effect of a 1 percentage point increase in genetic diversity is the average marginal effect across the entire cross-section of observed diversity values, and it reflects the increase in the likelihood of an intragroup factional conflict incidence in the 10-year interval, 1990–1999, expressed in percentage points. Robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

specification attempts to address potential concerns that because the MAR groups in a given country may not be representative of all of its subnational groups, if higher genetic diversity in a national population happens to be associated with a higher prevalence of MAR groups, and if

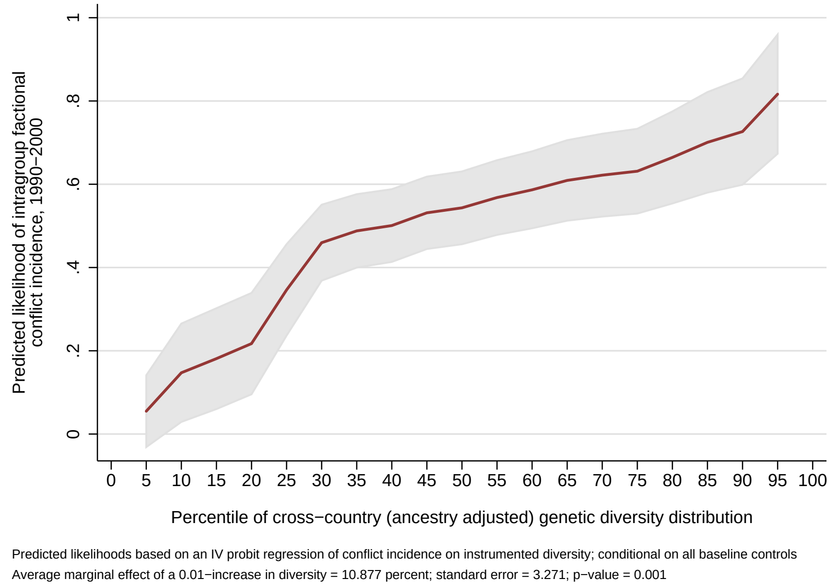


FIGURE 5: The Effect of Instrumented Genetic Diversity on the Likelihood of Intragroup Factional Conflict Incidence in the Global Sample

Notes: This figure depicts the influence of contemporary genetic diversity at the country level on the *predicted* likelihood of observing one or more factional conflicts *within* the "minorities at risk" (MAR) groups of a country's population in the 1990–1999 time period, conditional on other well-known diversity measures, the proximate geographical, institutional, and development-related correlates of conflict, and continent dummies. The predicted likelihood of observing one or more intragroup factional conflicts is illustrated as a function of the percentile of the cross-country genetic diversity distribution, and the prediction is based on the relevant IV probit regression from Table 11, exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity, and conducted using the global sample of countries and the baseline set of geographical, institutional, and development-related covariates. The shaded area reflects the 95-percent confidence-interval region of the depicted relationship.

MAR groups also happen to face a higher risk (relative to non-MAR groups) of intragroup factional conflict, then any observed positive influence of genetic diversity on the incidence of such conflict could be spurious. Finally, given that our analysis of intragroup factional conflict incidence exploits a standard cross-country framework, wherever relevant, our time-varying controls for institutional factors and development outcomes enter the specification as their respective temporal means over the 1990–1999 time interval.

Turning to our findings in Table 11, the results obtained across all specifications and identification strategies invariably indicate that genetic diversity contributes substantially to the risk of intragroup factional conflict events in society, imparting an influence that is not only highly statistically significant but considerable in terms of economic significance as well. For instance, exploiting variations in our globally representative sample of countries, the IV probit regression presented in Column 5 suggests that conditional on our complete set of baseline controls for geographical characteristics, institutional factors, ethnolinguistic fragmentation, and the correlates of economic development, a 1 percentage point increase in genetic diversity leads to an increase in the likelihood of observing the incidence of one or more intragroup factional conflict events in the 10-year interval between 1990 and 1999 by almost 10.9 percentage points, as reflected by an estimated average marginal effect that is statistically significant at the 1 percent level. Based on this regression, Figure 5 illustrates the manner in which the *predicted* likelihood associated with the outcome variable responds as one moves along the global cross-country genetic diversity distribution.[55] According to this figure, in response to a move from the 10th to the 90th percentile of the cross-country genetic diversity distribution, the predicted likelihood of observing one or more factional conflicts within the MAR groups of a country during the 1990–1999 time span increases from 14.7 percent to 72.6 percent, reflecting a change in the decadal hazard of experiencing such conflicts that is unarguably appreciable by all standards.

4.6 An Investigation of Some Potential Mediating Channels

We conclude the presentation of our results with a discussion of our findings from an exploratory analysis, examining some of our hypothesized proximate mechanisms that can potentially mediate the positive reduced-form cross-country relationship between genetic diversity and the risk of intrastate conflict, as reflected by the annual frequency of new PRIO25 civil conflict outbreaks during the 1960–2008 time period. Specifically, consistently with priors, our analysis in this section provides evidence suggesting that our main cross-country empirical finding is partly an expression of (i) the contribution of genetic diversity to the degree of ethnolinguistic fragmentation at the country level, as reflected by data from Fearon (2003) on the total number of ethnic groups in a national population;[56] (ii) the adverse influence of genetic diversity on social capital, based on data from the World Values Survey (2006, 2009) (henceforth referred to as WVS) on the prevalence of generalized interpersonal trust in a country's population;[57] and (iii) the association between genetic diversity and heterogeneity in preferences for public goods and redistributive policies at the national level, as captured by the intracountry dispersion in self-reported individual political positions on a politically "left"–"right" categorical scale, based on data from the WVS.[58]

[55] Similar to Figure 5, Figure A.3 in Appendix A depicts how the predicted likelihood associated with the incidence of one or more intragroup factional conflict events during the 1990–1999 time horizon responds as one moves along the cross-country genetic diversity distribution in the Old World, based on the probit regression presented in Column 2 of Table 11.

[56] Unlike measures of ethnolinguistic fragmentation that are based on fractionalization or polarization indices, the number of ethnic groups in the national population is potentially less endogenous in an empirical model of the risk of civil conflict, in light of the fact that this measure is not additionally tainted by the incorporation of information on the endogenous shares of the different subnational groups.

[57] In particular, this well-known measure of social capital reflects the proportion in a given country of all respondents (from across five different waves of the WVS, conducted over the 1981–2009 time horizon) that opted for the answer "Most people can be trusted" (as opposed to "Can't be too careful") when responding to the survey question "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?"

[58] Specifically, this country-level measure of heterogeneity in political attitudes reflects the intracountry standard deviation across all respondents (sampled over five different waves of the WVS during the 1981–2009 time horizon) of their self-reported positions on a categorical scale from 1 (politically "left") to 10 (politically "right") when answering the survey question "In political matters, people talk of 'the left' and 'the right.' How would you place your views on this scale, generally speaking?" Given that the unit of measurement of this particular variable does not possess any natural interpretation, we standardize the cross-country distribution of this variable prior to conducting our regressions.

Table 12 reveals the findings from our empirical examination of the aforementioned three potential mechanisms through which genetic diversity can partly contribute to the risk of intrastate conflict in society. For each posited channel, we present the results from estimating three different OLS regressions, exploiting worldwide variations in a common sample of countries, conditioned primarily by the availability of data on the mediating variable in question. In addition, throughout our analysis, we restrict our specifications to partialling out the influence of only our baseline set of geographical covariates (including continent or regional fixed effects), in order to prevent the inferential value of our findings regarding the mediating role of the proximate factors from being tainted by the presence of potentially endogenous control variables, many of which (like GDP per capita) may well be afflicted by reverse causality from the temporal frequency of civil conflict onsets and may also be determined in part by both genetic diversity and the mediating variable.

For our analysis of each mechanism, we proceed by first regressing the mediating variable on genetic diversity, documenting a highly statistically significant relationship that is qualitatively consistent with priors. In particular, based on coefficients that are all statistically significant at the 1 percent level, the regressions presented in Columns 1, 4, and 7 suggest that conditional on exogenous geographical factors, a move from the 10th to the 90th percentile of the cross-country genetic diversity distribution in the relevant sample is associated with (i) an increase by 2.236 in the total number of ethnic groups in a national population; (ii) a decrease in the prevalence of generalized interpersonal trust at the country level by 12.4 percent; and (iii) an increase in the intracountry dispersion in individual political attitudes by 97.4 percent of a standard deviation from the cross-country distribution of this particular measure.[59]

Notably, however, the latter two regressions in our analysis of each hypothesized channel confirm priors by establishing that the quantitative importance of genetic diversity as a predictor of the risk of civil conflict in society does indeed become diminished in both magnitude and explanatory power once the reduced-form influence of genetic diversity on the temporal frequency of civil conflict outbreaks is conditioned on the mediating variable of interest. Specifically, a comparison of the regressions in Columns 2 versus 3 indicates that when conditioned on the total number of ethnic groups in the national population, the influence of genetic diversity on conflict frequency, in terms of the response associated with a move from the 10th to the 90th percentile of the cross-country genetic diversity distribution, experiences a 20.5 percent decline in magnitude (from 0.022 to 0.018 new PRIO25 civil conflict onsets per year), whereas the explanatory power of genetic diversity for conflict frequency, as reflected by the partial R^2 statistic, diminishes by 37.3 percent (with the residual cross-country variation in genetic diversity explaining 2.5 percent as opposed to 4.1 percent of the residual cross-country variation in conflict frequency). The corresponding results obtained for each of the other two posited mechanisms are qualitatively similar but somewhat

[59] The three scatter plots presented in Figure A.4 in Appendix A depict these statistically significant cross-country relationships, conditional on our baseline set of geographical covariates (including continent or regional fixed effects), between genetic diversity and (i) the total number of ethnic groups in a national population (Panel A); (ii) the prevalence of generalized interpersonal trust at the country level (Panel B); and (iii) the intracountry dispersion in political attitudes (Panel C).

Table 12: Genetic Diversity and the Frequency of Overall Civil Conflict Onset across Countries – Mediating Channels

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
	Cultural-fragmentation channel			Trust channel			Preference-heterogeneity channel		
	Log number of ethnic groups	Frequency of new PRIO25 civil conflict onsets, 1960–2008		Prevalence of interpersonal trust	Frequency of new PRIO25 civil conflict onsets, 1960–2008		Variation in political attitudes	Frequency of new PRIO25 civil conflict onsets, 1960–2008	
Genetic diversity (ancestry adjusted)	5.431***	0.326**	0.259**	-2.151***	0.643**	0.583**	16.963***	0.548**	0.488*
	[1.884]	[0.141]	[0.124]	[0.807]	[0.258]	[0.277]	[6.371]	[0.246]	[0.290]
Log number of ethnic groups			0.012**						
			[0.006]						
Prevalence of interpersonal trust						-0.028			
						[0.033]			
Variation in political attitudes									0.004
									[0.007]
Absolute latitude	-16.323***	-0.525**	-0.324	3.363**	-0.664***	-0.571**	-35.588***	-0.553**	-0.426
	[4.557]	[0.253]	[0.269]	[1.378]	[0.216]	[0.264]	[9.132]	[0.222]	[0.276]
Land area	-30.686	1.853	2.231	10.087	-0.467	-0.189	-65.121	1.513	1.746
	[22.605]	[2.165]	[2.012]	[6.134]	[2.328]	[2.342]	[50.205]	[2.622]	[2.618]
Ruggedness	-0.377	0.026	0.031	0.053	0.074	0.075	-0.850	0.047	0.050
	[0.552]	[0.043]	[0.043]	[0.173]	[0.066]	[0.066]	[1.135]	[0.063]	[0.063]
Mean elevation	-0.057	-0.014	-0.013	-0.036	-0.035*	-0.036**	0.072	-0.023	-0.023
	[0.146]	[0.009]	[0.009]	[0.046]	[0.018]	[0.017]	[0.400]	[0.016]	[0.016]
Range of elevation	0.045	0.010**	0.009**	0.014	0.016**	0.017**	-0.053	0.018***	0.018***
	[0.046]	[0.004]	[0.004]	[0.013]	[0.007]	[0.006]	[0.179]	[0.007]	[0.006]
Mean land suitability	-0.393**	0.021	0.026*	-0.189***	0.037*	0.031	0.029	0.043**	0.043**
	[0.196]	[0.014]	[0.014]	[0.065]	[0.019]	[0.021]	[0.466]	[0.019]	[0.018]
Range of land suitability	0.601***	0.007	-0.001	-0.048	-0.001	-0.003	-0.399	-0.013	-0.012
	[0.172]	[0.012]	[0.012]	[0.064]	[0.016]	[0.016]	[0.459]	[0.020]	[0.020]
Distance to nearest waterway	0.165*	0.005	0.003	-0.104**	0.018	0.016	0.537	0.007	0.005
	[0.099]	[0.009]	[0.009]	[0.045]	[0.019]	[0.019]	[0.338]	[0.019]	[0.019]
Effect of increasing genetic diversity from the 10^{th} to the 90^{th} percentile	2.236***	0.022**	0.018**	-0.124***	0.038**	0.034**	0.974***	0.032**	0.029*
	[0.776]	[0.010]	[0.008]	[0.046]	[0.015]	[0.016]	[0.366]	[0.014]	[0.017]
Continent/region dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Global	Global	Global	Global	Global	Global	Global	Global	Global
Observations	147	147	147	84	84	84	81	81	81
Partial R^2 of genetic diversity	0.054	0.041	0.025	0.105	0.084	0.063	0.111	0.059	0.042
Adjusted R^2	0.350	0.136	0.157	0.432	0.167	0.161	0.398	0.190	0.183

Notes: This table exploits cross-country variations to demonstrate that the significant positive reduced-form influence of contemporary genetic diversity on the annual frequency of new overall (PRIO25) civil conflict onsets during the 1960–2008 time period, conditional on the baseline geographical correlates of conflict, is at least partly mediated by each of three potentially conflict-augmenting proximate channels that capture the contribution of genetic diversity to (i) the degree of cultural fragmentation, as reflected by the number of ethnic groups in the national population; (ii) the diminished prevalence of generalized interpersonal trust at the country level; and (iii) the extent of heterogeneity in preferences for redistribution and public-goods provision, as reflected by the intracountry dispersion in individual political attitudes on a politically "left"–"right" categorical scale. For each of the three mediating channels examined, the first regression documents the impact of genetic diversity on the proximate variable in the channel, the second presents the reduced-form influence of genetic diversity on conflict, and the third runs a "horse race" between genetic diversity and the proximate variable to establish reductions in the magnitude and explanatory power of the reduced-form influence of genetic diversity on conflict. All three regressions for each channel are conducted using a common cross-country sample, conditioned by the availability of data on the relevant variables employed by the analysis of the channel in question. The regressions for the "cultural fragmentation" channel control for the full set of continent dummies (i.e., indicators for Africa, Asia, North America, South America, and Oceania, with Europe being treated as the omitted category), whereas for the "trust" and "preference heterogeneity" channels, given the smaller degrees of freedom afforded by the more limited sample of countries, the regressions control for a more modest set of region dummies, including indicators for Sub-Saharan Africa and for Latin America and the Caribbean. Given that the unit of measurement for the variable reflecting the degree of intracountry dispersion in political attitudes has no natural interpretation, its cross-country distribution is standardized prior to conducting the relevant regressions. The estimated effect associated with increasing genetic diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of (i) the actual number of ethnic groups in the national population in Column 1; (ii) the fraction of individuals in a country who "think that most people can be trusted" in Column 4; (iii) the number of standard deviations of the cross-country distribution of the national-level dispersion in political attitudes in Column 7; and (iv) the actual number of new conflict onsets per year in all the remaining columns. Robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

more muted, possibly due to greater measurement error in the relevant mediating variable. In particular, when conditioned on either the prevalence of generalized interpersonal trust in the national population or the intracountry dispersion in political attitudes, the magnitude of the response in conflict frequency that is associated with a move from the 10th to the 90th percentile of the cross-country genetic diversity distribution decreases by either 9.25 percent (Columns 5 versus 6) or 11.1 percent (Columns 8 versus 9), while the explanatory power of genetic diversity for conflict frequency decays respectively by either 24.2 percent or 28.0 percent, as reflected by the observed reduction in the partial R^2 statistic. Interestingly, unlike the influence of the number of ethnic groups on temporal frequency of civil conflict onsets in Column 3, the regressions in Columns 6 and 9 indicate that neither the prevalence of interpersonal trust nor the dispersion in political attitudes confers a statistically significant influence on conflict frequency, conditional on genetic diversity and exogenous geographical factors. This finding, however, is consistent with attenuation bias afflicting the coefficients associated with the latter two mediating variables, in line with the aforementioned assertion regarding a potentially larger amount of white noise in the measurement of these variables.

One important caveat regarding the interpretation of our findings in Table 12 is that the mediating variables considered by our analysis may themselves be potentially endogenous in an empirical model of the risk of civil conflict. To be clear, as corroborated by evidence from recent studies (e.g., Fletcher and Iyigun, 2010; Rohner, Thoenig and Zilibotti, 2013; Besley and Reynal-Querol, 2014), the unobserved historical cross-regional pattern of conflict risk may not only have persisted to the modern era through various mechanisms, but by triggering the movement of ethnic groups across space and reinforcing extant interethnic cleavages (along with the social, political, and economic grievances associated with such divisions), it may also have partly contributed to the contemporary variations observed across countries in the degree of ethnolinguistic fragmentation, the prevalence of interpersonal trust, and the intracountry dispersion in revealed political preferences. Thus, by potentially introducing endogeneity bias to the estimated coefficients associated with the proximate determinants of conflict risk, this issue calls for some caution with respect to interpreting our findings as being reflective of the actual role of these factors as mediators of the reduced-form contribution of genetic diversity to the potential for violent dissensions in contemporary national populations. In order to assess our hypothesized mechanisms more conclusively, one would need to exploit an independent exogenous source of variation for each of these proximate factors, a task that we leave open for future exploration.

5 Concluding Remarks

This paper documents a novel but salient empirical fact – namely, that the risk of intrastate conflict faced by societies in the modern world partly reflects the long shadow of prehistory. Specifically, exploiting variations across contemporary national populations, we establish that genetic diversity, overwhelmingly determined during the course of the prehistoric demic diffusion of humans "out of Africa" to the rest of the globe, has contributed significantly to the temporal frequency, incidence, and onset of both overall and ethnic civil conflict events over the last half-century, accounting for the potentially confounding influence of a large set of geographical characteristics, institutional factors, measures of ethnolinguistic fragmentation, and outcomes of economic development. Our analysis additionally demonstrates that genetic diversity possesses significant explanatory power for not only the intensity of social unrest but also the incidence of intragroup factional conflicts in contemporary national populations. Importantly, the reduced-form causal influence of genetic diversity on the risk of intrastate conflict in the modern era remains qualitatively unchanged under a comprehensive range of robustness checks.

Our key finding in this paper arguably reflects the contribution of genetic diversity to the ethnolinguistic fragmentation of a national population, the adverse influence of genetic diversity on social capital, the contribution of genetic diversity to heterogeneity in preferences for public goods and redistributive policies, and possibly, the potential impact of genetic diversity on economic inequality within a society. Consistently with this assertion, our analysis documents that the quantitative importance of genetic diversity does indeed become diminished in both magnitude and explanatory power once its reduced-form influence on the risk of intrastate conflict is conditioned on either the total number of ethnic groups in a country's population, the prevalence of generalized interpersonal trust at the national level, or the intracountry dispersion in revealed individual political preferences. In light of the possible endogeneity of these proximate determinants, however, deriving stronger conclusions with respect to the mechanisms through which genetic diversity confers its reduced-form effect on the risk of conflict in society necessitates an exogenous source of variation for each of the aforementioned mediating variables, the exploration of which is left as an important task for future research.

References

Alesina, Alberto, and Eliana La Ferrara. 2005. "Ethnic Diversity and Economic Performance." *Journal of Economic Literature*, 43(3): 762–800.

Alesina, Alberto, and Enrico Spolaore. 2003. *The Size of Nations.* Cambridge, MA:MIT Press.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth*, 8(2): 155–194.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics*, 114(4): 1243–1284.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy*, 113(1): 151–184.

Angrist, Joshua D., and Adriana D. Kugler. 2008. "Rural Windfall or a New Resource Curse? Coca, Income, and Civil Conflict in Colombia." *Review of Economics and Statistics*, 90(2): 191–215.

Ashraf, Quamrul, and Oded Galor. 2011. "Dynamics and Stagnation in the Malthusian Epoch." *American Economic Review*, 101(5): 2003–2041.

Ashraf, Quamrul, and Oded Galor. 2013a. "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review*, 103(1): 1–48.

Ashraf, Quamrul, and Oded Galor. 2013b. "Genetic Diversity and the Origins of Cultural Fragmentation." *American Economic Review: Papers & Proceedings*, 103(3): 528–533.

Atkinson, Quentin D. 2011. "Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa." *Science*, 332(6027): 346–349.

Banks, Arthur S. 2010. "Cross-National Time-Series Data Archive." Databanks International, Jerusalem, Israel. http://www.databanksinternational.com/ (accessed March 23, 2012).

Barth, Fredrik, ed. 1969. *Ethnic Groups and Boundaries: The Social Organization of Culture Difference.* Olso, Norway:Universitetsforlaget.

Bates, Robert H. 1983. "Modernization, Ethnic Competition, and the Rationality of Politics in Contemporary Africa." In *State versus Ethnic Claims: African Policy Dilemmas.* , ed. Donald Rothchild and Victor A. Olorunsola, 152–171. Boulder, CO:Westview Press.

Bazzi, Samuel, and Christopher Blattman. 2014. "Economic Shocks and Conflict: Evidence from Commodity Prices." *American Economic Journal: Macroeconomics*, 6(4): 1–38.

Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series–Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science*, 42(4): 1260–1288.

Bellows, John, and Edward Miguel. 2009. "War and Local Collective Action in Sierra Leone." *Journal of Public Economics*, 93(11-12): 1144–1157.

Besley, Timothy, and Marta Reynal-Querol. 2014. “The Legacy of Historical Conflict: Evidence from Africa.” *American Political Science Review*, 108(2): 319–336.

Betti, Lia, François Balloux, William Amos, Tsunehiko Hanihara, and Andrea Manica. 2009. “Distance from Africa, Not Climate, Explains Within-Population Phenotypic Diversity in Humans.” *Proceedings of the Royal Society B: Biological Sciences*, 276(1658): 809–814.

Betti, Lia, Noreen von Cramon-Taubadel, Andrea Manica, and Stephen J. Lycett. 2013. “Global Geometric Morphometric Analyses of the Human Pelvis Reveal Substantial Neutral Population History Effects, Even across Sexes.” *PLoS ONE*, 8(2): e55909.

Blattman, Christopher, and Edward Miguel. 2010. “Civil War.” *Journal of Economic Literature*, 48(1): 3–57.

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

Brass, Paul R. 1997. *Theft of an Idol: Text and Context in the Representation of Collective Violence.* Princeton, NJ:Princeton University Press.

Brewer, Marilynn B. 1979. “In-Group Bias in the Minimal Intergroup Situation: A Cognitive-Motivational Analysis.” *Psychological Bulletin*, 86(2): 307–324.

Brewer, Marilynn B. 1991. “The Social Self: On Being the Same and Different at the Same Time.” *Personality and Social Psychology Bulletin*, 17(5): 475–482.

Brewer, Marilynn B. 1997. “The Social Psychology of Intergroup Relations: Can Research Inform Practice?” *Journal of Social Issues*, 53(1): 197–211.

Brubaker, Rogers. 2004. *Ethnicity without Groups.* Cambridge, MA:Harvard University Press.

Brubaker, Rogers, and David D. Laitin. 1998. “Ethnic and Nationalist Violence.” *Annual Review of Sociology*, 24: 423–452.

Brückner, Markus. 2010. “Population Size and Civil Conflict Risk: Is there a Causal Link?” *The Economic Journal*, 120(544): 535–550.

Brückner, Markus, and Antonio Ciccone. 2010. “International Commodity Prices, Growth and the Outbreak of Civil War in Sub-Saharan Africa.” *The Economic Journal*, 120(544): 519–534.

Cann, Howard M., Claudia de Toma, Lucien Cazes, Marie-Fernande Legrand, Valerie Morel, Laurence Piouffre, Julia Bodmer, et al. 2002. “A Human Genome Diversity Cell Line Panel.” *Science*, 296(5566): 261–262.

Caselli, Francesco, and Wilbur John Coleman, II. 2013. “On the Theory of Ethnic Conflict.” *Journal of the European Economic Association*, 11(S1): 161–192.

Cavalli-Sforza, L. Luca. 2005. “The Human Genome Diversity Project: Past, Present and Future.” *Nature Reviews Genetics*, 6(4): 333–340.

Cavalli-Sforza, L. Luca, and Marcus W. Feldman. 1981. *Cultural Transmission and Evolution: A Quantitative Approach.* Princeton, NJ:Princeton University Press.

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

Central Intelligence Agency, The. 2006. "The World Factbook 2006." The Central Intelligence Agency, Washington, DC. https://www.cia.gov/library/publications/the-world-factbook/ (accessed March 18, 2007).

Cervellati, Matteo, Uwe Sunde, and Simona Valmori. 2011. "Disease Environment and Civil Conflicts." IZA Discussion Paper No. 5614, Institute for the Study of Labor.

Collier, Paul, and Anke Hoeffler. 1998. "On Economic Causes of Civil War." *Oxford Economic Papers*, 50(4): 563–573.

Collier, Paul, and Anke Hoeffler. 2004. "Greed and Grievance in Civil War." *Oxford Economic Papers*, 56(4): 563–595.

Collier, Paul, and Anke Hoeffler. 2007. "Civil War." In *Handbook of Defense Economics, Vol. 2: Defense in a Globalized World.* , ed. Todd Sandler and Keith Hartley, 711–740. Amsterdam, The Netherlands:Elsevier, North-Holland.

Conley, T. G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics*, 92(1): 1–45.

Connor, Walker. 1994. *Ethnonationalism: The Quest for Understanding.* Princeton, NJ:Princeton University Press.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Romain Wacziarg. 2012. "The Political Economy of Linguistic Cleavages." *Journal of Development Economics*, 97(2): 322–338.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Shlomo Weber. 2009. "Linguistic Diversity and Redistribution." *Journal of the European Economic Association*, 7(6): 1291–1318.

Desmet, Klaus, Michel Le Breton, Ignacio Ortuño-Ortín, and Shlomo Weber. 2011. "The Stability and Breakup of Nations: A Quantitative Analysis." *Journal of Economic Growth*, 16(3): 183–213.

Dube, Oeindrila, and Juan F. Vargas. 2013. "Commodity Price Shocks and Civil Conflict: Evidence from Colombia." *Review of Economic Studies*, 80(4): 1384–1421.

Durham, William H. 1991. *Coevolution: Genes, Culture, and Human Diversity.* Stanford, CA:Stanford University Press.

Easterly, William, and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics*, 112(4): 1203–1250.

Eifert, Benn, Edward Miguel, and Daniel N. Posner. 2010. "Political Competition and Ethnic Identification in Africa." *American Journal of Political Science*, 54(2): 494–510.

Elbadawi, Ibrahim, and Nicholas Sambanis. 2002. "How Much War Will We See? Explaining the Prevalence of Civil War." *Journal of Conflict Resolution*, 46(3): 307–334.

Esteban, Joan, and Debraj Ray. 1999. "Conflict and Distribution." *Journal of Economic Theory*, 87(2): 379–415.

Esteban, Joan, and Debraj Ray. 2008a. "Polarization, Fractionalization and Conflict." *Journal of Peace Research*, 45(2): 163–182.

Esteban, Joan, and Debraj Ray. 2008b. "On the Salience of Ethnic Conflict." *American Economic Review*, 98(5): 2185–2202.

Esteban, Joan, and Debraj Ray. 2011a. "A Model of Ethnic Conflict." *Journal of the European Economic Association*, 9(3): 496–521.

Esteban, Joan, and Debraj Ray. 2011b. "Linking Conflict to Inequality and Polarization." *American Economic Review*, 101(4): 1345–1374.

Esteban, Joan, Laura Mayoral, and Debraj Ray. 2012. "Ethnicity and Conflict: An Empirical Study." *American Economic Review*, 102(4): 1310–1342.

Esteban, Joan-María, and Debraj Ray. 1994. "On the Measurement of Polarization." *Econometrica*, 62(4): 819–851.

Esty, Daniel C., Jack A. Goldstone, Ted Robert Gurr, Barbara Harff, Marc Levy, Geoffrey D. Dabelko, Pamela T. Surko, et al. 1998. "State Failure Task Force Report: Phase II Findings." Science Applications International Corporation, McLean, VA.

Fearon, James D. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth*, 8(2): 195–222.

Fearon, James D. 2005. "Primary Commodity Exports and Civil War." *Journal of Conflict Resolution*, 49(4): 483–507.

Fearon, James D., and David D. Laitin. 2000. "Violence and the Social Construction of Ethnic Identity." *International Organization*, 54(4): 845–877.

Fearon, James D., and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review*, 97(1): 75–90.

Fincher, Corey L., and Randy Thornhill. 2008. "Assortative Sociality, Limited Dispersal, Infectious Disease and the Genesis of the Global Pattern of Religion Diversity." *Proceedings of the Royal Society B: Biological Sciences*, 275(1651): 2587–2594.

Fletcher, Erin, and Murat Iyigun. 2010. "The Clash of Civilizations: A Cliometric Investigation." http://www.colorado.edu/economics/courses/iyigun/fractionalization013109.pdf.

Gallup, John Luke, Jeffrey D. Sachs, and Andrew D. Mellinger. 1999. "Geography and Economic Development." *International Regional Science Review*, 22(2): 179–232.

Galor, Oded. 2011. *Unified Growth Theory.* Princeton, NJ:Princeton University Press.

Geertz, Clifford. 1973. *The Interpretation of Cultures.* New York, NY:Basic Books.

Gellner, Ernest. 1983. *Nations and Nationalism.* Ithaca, NY:Cornell University Press.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand. 2002. "Armed Conflict 1946-2001: A New Dataset." *Journal of Peace Research*, 39(5): 615–637.

Griffiths, Anthony J. F., Jeffrey H. Miller, David T. Suzuki, Richard C. Lewontin, and William M. Gelbart. 2000. *An Introduction to Genetic Analysis.* . 7th ed., New York, NY:W. H. Freeman & Co.

Grossman, Herschel I. 1991. "A General Equilibrium Model of Insurrections." *American Economic Review*, 81(4): 912–921.

Grossman, Herschel I. 1999. "Kleptocracy and Revolutions." *Oxford Economic Papers*, 51(2): 267–283.

Gurr, Ted Robert. 2000. "Ethnic Warfare on the Wane." *Foreign Affairs*, 79(3): 52–64.

Hanihara, Tsunehiko. 2008. "Morphological Variation of Major Human Populations Based on Nonmetric Dental Traits." *American Journal of Physical Anthropology*, 136(2): 169–182.

Hardin, Russell. 1995. *One for All: The Logic of Group Conflict.* Princeton, NJ:Princeton University Press.

Hartl, Daniel L., and Andrew G. Clark. 2007. *Principles of Population Genetics.* . 4th ed., Sunderland, MA:Sinauer Associates, Inc.

Hegre, Håvard, and Nicholas Sambanis. 2006. "Sensitivity Analysis of Empirical Results on Civil War Onset." *Journal of Conflict Resolution*, 50(4): 508–535.

Heinrich, Joseph, and Richard McElreath. 2003. "The Evolution of Cultural Evolution." *Evolutionary Anthropology*, 12(3): 123–135.

Hirshleifer, Jack. 1991. "The Technology of Conflict as an Economic Activity." *American Economic Review: Papers & Proceedings*, 81(2): 130–134.

Hirshleifer, Jack. 1995. "Anarchy and its Breakdown." *Journal of Political Economy*, 103(1): 26–52.

Horowitz, Donald L. 1985. *Ethnic Groups in Conflict.* Berkeley, CA:University of California Press.

Horowitz, Donald L. 1999. "Structure and Strategy in Ethnic Conflict: A Few Steps Toward Synthesis." In *Annual World Bank Conference on Development Economics 1998.* , ed. Boris Pleskovic and Joseph E. Stiglitz, 345–370. Washington, DC:The World Bank.

Humphreys, Macartan. 2005. "Natural Resources, Conflict, and Conflict Resolution: Uncovering the Mechanisms." *Journal of Conflict Resolution*, 49(4): 508–537.

King, Gary, and Langche Zeng. 2001. "Logistic Regression in Rare Events Data." *Political Analysis*, 9(2): 137–163.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 1999. "The Quality of Government." *Journal of Law, Economics, and Organization*, 15(1): 222–279.

Letendre, Kenneth, Corey L. Fincher, and Randy Thornhill. 2010. “Does Infectious Disease Cause Global Variation in the Frequency of Intrastate Armed Conflict and Civil War?” *Biological Reviews*, 85(3): 669–683.

Maddison, Angus. 2010. “Statistics on World Population, GDP and Per Capita GDP, 1–2008 AD.” Groningen Growth and Development Centre, Groningen, The Netherlands. http://www.ggdc.net/maddison/oriindex.htm (accessed January 2, 2011).

Manica, Andrea, William Amos, François Balloux, and Tsunehiko Hanihara. 2007. “The Effect of Ancient Population Bottlenecks on Human Phenotypic Variation.” *Nature*, 448(7151): 346–348.

Marshall, Monty G. 2010. “Major Episodes of Political Violence (MEPV) and Conflict Regions, 1946–2008.” Center for Systemic Peace, Vienna, VA. http://www.systemicpeace.org/inscrdata.html (accessed February 10, 2011).

Marshall, Monty G., Ted Robert Gurr, and Keith Jaggers. 2009. “Polity IV Project: Political Regime Characteristics and Transitions, 1800–2009.” Center for Systemic Peace, Vienna, VA. http://www.systemicpeace.org/inscrdata.html (accessed January 2, 2011).

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

Michalopoulos, Stelios, and Elias Papaioannou. 2013. “The Long-Run Effects of the Scramble for Africa.” NBER Working Paper No. 17620, National Bureau of Economic Research.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti. 2004. “Economic Shocks and Civil Conflict: An Instrumental Variables Approach.” *Journal of Political Economy*, 112(4): 725–753.

Minorities at Risk Project, The. 2009. “Minorities at Risk Dataset.” Center for International Development and Conflict Management, College Park, MD. http://www.cidcm.umd.edu/mar/mar_data.asp (accessed August 24, 2012).

Montalvo, José G., and Marta Reynal-Querol. 2005. “Ethnic Polarization, Potential Conflict, and Civil Wars.” *American Economic Review*, 95(3): 796–816.

Nordhaus, William D. 2006. “Geography and Macroeconomics: New Data and New Findings.” *Proceedings of the National Academy of Sciences*, 103(10): 3510–3517.

Nunn, Nathan. 2014. “Historical Development.” In *Handbook of Economic Growth, Vol. 2A.* , ed. Philippe Aghion and Steven N. Durlauf, 347–402. Amsterdam, The Netherlands:Elsevier, North-Holland.

Posner, Daniel N. 2003. “The Colonial Origins of Ethnic Cleavages: The Case of Linguistic Divisions in Zambia.” *Comparative Politics*, 35(2): 127–146.

Prugnolle, Franck, Andrea Manica, and François Balloux. 2005. “Geography Predicts Neutral Genetic Diversity of Human Populations.” *Current Biology*, 15(5): R159–R160.

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

Ramachandran, Sohini, Omkar Deshpande, Charles C. Roseman, Noah A. Rosenberg, Marcus W. Feldman, and L. Luca Cavalli-Sforza. 2005. "Support from the Relationship of Genetic and Geographic Distance in Human Populations for a Serial Founder Effect Originating in Africa." *Proceedings of the National Academy of Sciences*, 102(44): 15942–15947.

Rohner, Dominic, Mathias Thoenig, and Fabrizio Zilibotti. 2013. "Seeds of Distrust: Conflict in Uganda." *Journal of Economic Growth*, 18(3): 217–252.

Ross, Michael. 2006. "A Closer Look at Oil, Diamonds, and Civil War." *Annual Review of Political Science*, 9: 265–300.

Ross, Michael L. 2013. "Oil and Gas Data, 1932–2011." http://dvn.iq.harvard.edu/dvn/dv/mlross/faces/study/StudyPage.xhtml?studyId=87422&versionNumber=2 (accessed December 26, 2013).

Rummel, Rudolph J. 1963. "Dimensions of Conflict Behavior Within and Between Nations." *General Systems Yearbook*, 8: 1–50.

Sambanis, Nicholas. 2001. "Do Ethnic and Nonethnic Civil Wars Have the Same Causes? A Theoretical and Empirical Inquiry (Part 1)." *Journal of Conflict Resolution*, 45(3): 259–282.

Sambanis, Nicholas. 2002. "A Review of Recent Advances and Future Directions in the Quantitative Literature on Civil War." *Defence and Peace Economics*, 13(3): 215–243.

Sambanis, Nicholas. 2004. "What Is Civil War? Conceptual and Empirical Complexities of an Operational Definition." *Journal of Conflict Resolution*, 48(6): 814–858.

Shils, Edward. 1957. "Primordial, Personal, Sacred and Civil Ties: Some Particular Observations on the Relationships of Sociological Research and Theory." *British Journal of Sociology*, 8(2): 130–145.

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

Spolaore, Enrico, and Romain Wacziarg. 2013a. "War and Relatedness." CAGE Online Working Paper No. 140, Centre for Competitive Advantage in the Global Economy.

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

Spolaore, Enrico, and Romain Wacziarg. 2014. "Long-Term Barriers to Economic Development." In *Handbook of Economic Growth, Vol. 2A.* , ed. Philippe Aghion and Steven N. Durlauf, 121–176. Amsterdam, The Netherlands:Elsevier, North-Holland.

Sundberg, Ralph, Kristine Eck, and Joakim Kreutz. 2012. "Introducing the UCDP Non-State Conflict Dataset." *Journal of Peace Research*, 49(2): 351–362.

Themnér, Lotta, and Peter Wallensteen. 2012. "Armed Conflicts, 1946–2011." *Journal of Peace Research*, 49(4): 565–575.

Themnér, Lotta, and Peter Wallensteen. 2014. "Armed Conflicts, 1946–2013." *Journal of Peace Research*, 51(4): 541–554.

Van den Berghe, Pierre L. 1981. *The Ethnic Phenomenon.* New York, NY:Elsevier, North-Holland.

Van den Berghe, Pierre L. 1995. "Does Race Matter?" *Nations and Nationalism*, 1(3): 357–368.

von Cramon-Taubadel, Noreen, and Stephen J. Lycett. 2008. "Brief Communication: Human Cranial Variation Fits Iterative Founder Effect Model with African Origin." *American Journal of Physical Anthropology*, 136(1): 108–113.

Wimmer, Andreas. 2002. *Nationalist Exclusion and Ethnic Conflict: Shadows of Modernity.* Cambridge, UK:Cambridge University Press.

Wimmer, Andreas, and Brian Min. 2006. "From Empire to Nation-State: Explaining Wars in the Modern World, 1816–2001." *American Sociological Review*, 71(6): 867–897.

Wimmer, Andreas, and Brian Min. 2009. "The Location and Purpose of Wars Around the World: A New Global Dataset, 1816–2001." *International Interactions*, 35(4): 390–417.

Wimmer, Andreas, Lars-Erik Cederman, and Brian Min. 2009. "Ethnic Politics and Armed Conflict: A Configurational Analysis of a New Global Data Set." *American Sociological Review*, 74(2): 316–337.

World Bank, The. 2006. "The World Development Indicators 2006." The World Bank, Washington, DC. http://data.worldbank.org/data-catalog/world-development-indicators (accessed February 15, 2007).

World Values Survey, The. 2006. "European and World Values Surveys, Four-Wave Integrated Data File, 1981–2004, version 20060423." The World Values Survey Association, Stockholm, Sweden. http://www.worldvaluessurvey.org/ (accessed August 24, 2010).

World Values Survey, The. 2009. "World Values Survey, 1981–2008 Official Aggregate, version 20090914." The World Values Survey Association, Stockholm, Sweden. http://www.worldvaluessurvey.org/ (accessed August 24, 2010).

Appendices

A Supplementary Results

Table A.1: Genetic Diversity and the Count of Overall Civil Conflict Onset across Countries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Negative Binomial	Negative Binomial	Negative Binomial	Negative Binomial	Negative Binomial	Negative Binomial	Negative Binomial	Negative Binomial	Negative Binomial	Negative Binomial
	Total count of new PRIO25 civil conflict onsets over the 1960–2008 time period									
Genetic diversity (ancestry adjusted)	9.090*	20.620***	17.048***	16.922***	16.809***	18.319***	18.296***	17.913***	24.833***	27.135***
	[4.842]	[4.391]	[6.242]	[6.012]	[6.008]	[5.996]	[6.013]	[5.496]	[7.298]	[5.848]
Ethnic fractionalization					0.251		0.038	0.250		0.560
					[0.518]		[0.540]	[0.537]		[0.581]
Ethnolinguistic polarization						0.647	0.638	0.853*		1.194**
						[0.487]	[0.502]	[0.471]		[0.474]
Absolute latitude		-34.697***	-33.666***	-24.749***	-22.968**	-29.119***	-28.799**	-3.924	-31.375***	1.313
		[6.498]	[8.831]	[9.148]	[10.216]	[10.014]	[11.351]	[11.644]	[9.533]	[12.287]
Land area		75.810	117.419**	101.690**	107.824**	104.576**	105.535**	43.343	158.988***	90.008
		[49.825]	[49.630]	[51.751]	[51.178]	[52.312]	[52.607]	[54.547]	[50.192]	[54.974]
Ruggedness		2.671*	2.579*	2.678*	2.744*	2.844*	2.852*	3.819**	2.869*	4.683***
		[1.406]	[1.561]	[1.552]	[1.568]	[1.544]	[1.555]	[1.579]	[1.695]	[1.653]
Mean elevation		-0.946***	-0.952***	-0.989***	-0.983***	-1.020***	-1.019***	-1.062***	-1.066***	-1.156***
		[0.301]	[0.337]	[0.309]	[0.308]	[0.317]	[0.316]	[0.300]	[0.385]	[0.356]
Range of elevation		0.438***	0.406***	0.362***	0.355***	0.357***	0.356***	0.098	0.380***	-0.011
		[0.100]	[0.085]	[0.083]	[0.086]	[0.086]	[0.089]	[0.116]	[0.083]	[0.136]
Mean land suitability		0.394	0.572	0.569	0.663	0.710	0.723	0.083	0.503	0.067
		[0.455]	[0.468]	[0.477]	[0.507]	[0.479]	[0.510]	[0.548]	[0.476]	[0.588]
Range of land suitability		0.733*	0.785*	0.870*	0.808*	0.938**	0.928**	0.467	1.047**	0.474
		[0.426]	[0.456]	[0.449]	[0.443]	[0.443]	[0.443]	[0.465]	[0.486]	[0.446]
Distance to nearest waterway		0.296	0.141	0.339	0.322	0.372	0.369	0.324	-0.013	0.156
		[0.253]	[0.253]	[0.316]	[0.320]	[0.323]	[0.328]	[0.301]	[0.284]	[0.334]
Executive constraints, 1960–2008 average				0.169	0.165	0.163	0.162	0.299**		0.237
				[0.148]	[0.151]	[0.148]	[0.149]	[0.135]		[0.161]
Fraction of years under democracy, 1960–2008				-0.735	-0.688	-0.706	-0.700	-0.984*		-0.904
				[0.577]	[0.581]	[0.582]	[0.582]	[0.571]		[0.672]
Fraction of years under autocracy, 1960–2008				0.128	0.152	0.132	0.136	-0.044		-0.188
				[0.606]	[0.601]	[0.605]	[0.597]	[0.561]		[0.619]
Log oil production per capita, 1960–2008 average								0.104**		0.110***
								[0.041]		[0.039]
Log population, 1960–2008 average								0.265***		0.339***
								[0.100]		[0.102]
Log GDP per capita, 1960–2008 average								-0.684***		-0.722***
								[0.171]		[0.195]
Effect of increasing genetic diversity from the 10^{th} to the 90^{th} percentile	0.473**	1.106***	0.885**	0.878**	0.872**	0.959**	0.958**	0.935***	1.170**	1.304***
	[0.241]	[0.322]	[0.369]	[0.359]	[0.358]	[0.375]	[0.376]	[0.338]	[0.472]	[0.420]
Continent dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Colonial history dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Sample	Global	Global	Global	Global	Global	Global	Global	Global	Old World	Old World
Observations	143	143	143	143	143	143	143	143	119	119
Pseudo R^2	0.012	0.171	0.197	0.216	0.217	0.220	0.220	0.257	0.217	0.292

Notes: This table exploits cross-country variations to establish a significant positive reduced-form impact of contemporary genetic diversity on the total count of new overall (PRIO25) civil conflict onsets during the 1960–2008 time period, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict. For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of colonial history dummies includes indicators for whether a country was ever a colony of the U.K., France, and any other major colonizing power. Given the absence of a negative binomial estimator that permits instrumentation, the current analysis is unable to employ the strategy of exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated effect associated with increasing genetic diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the total number of new conflict onsets. Robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A.2: Genetic Diversity and the Count of Ethnic Civil Conflict Onset across Countries

	(1) Negative Binomial	(2) Negative Binomial	(3) Negative Binomial	(4) Negative Binomial	(5) Negative Binomial	(6) Negative Binomial	(7) Negative Binomial	(8) Negative Binomial	(9) Negative Binomial	(10) Negative Binomial
	Total count of new WCM09 ethnic civil conflict onsets over the 1960–2005 time period									
Genetic diversity (ancestry adjusted)	11.835**	21.687***	22.008**	18.459**	17.260**	19.192**	17.757**	22.372***	33.674***	36.134***
	[5.825]	[5.860]	[9.224]	[8.502]	[8.560]	[8.091]	[8.233]	[7.879]	[8.146]	[7.438]
Ethnic fractionalization					1.436		1.307	1.461*		1.895*
					[0.921]		[0.955]	[0.882]		[0.997]
Ethnolinguistic polarization						0.714	0.319	0.397		0.287
						[0.767]	[0.794]	[0.688]		[0.696]
Absolute latitude		-37.659***	-57.138***	-40.054**	-29.618	-43.889***	-32.354*	4.220	-58.965***	5.322
		[9.746]	[12.716]	[16.099]	[18.215]	[16.502]	[19.001]	[20.265]	[13.311]	[22.768]
Land area		-35.909	54.576	50.451	94.239	52.818	91.234	32.986	97.864	41.717
		[73.944]	[72.486]	[76.144]	[79.594]	[79.110]	[82.581]	[77.014]	[66.202]	[82.511]
Ruggedness		3.133	0.601	1.332	1.639	1.382	1.637	3.223	2.067	5.388**
		[1.983]	[2.042]	[2.224]	[2.196]	[2.218]	[2.199]	[2.159]	[1.987]	[2.227]
Mean elevation		-0.914**	-0.391	-0.545	-0.479	-0.534	-0.484	-0.876	-0.767	-1.399**
		[0.457]	[0.476]	[0.462]	[0.469]	[0.477]	[0.477]	[0.558]	[0.495]	[0.576]
Range of elevation		0.346***	0.172	0.169	0.086	0.154	0.089	-0.220	0.195	-0.247
		[0.123]	[0.120]	[0.149]	[0.158]	[0.154]	[0.160]	[0.182]	[0.119]	[0.196]
Mean land suitability		0.461	0.338	0.815	1.487	1.001	1.510	0.439	0.343	0.473
		[0.769]	[0.723]	[0.803]	[0.957]	[0.855]	[0.971]	[0.984]	[0.739]	[1.017]
Range of land suitability		2.088***	2.411***	2.811***	2.545***	2.913***	2.606***	2.236**	2.219***	1.704
		[0.695]	[0.745]	[0.771]	[0.745]	[0.757]	[0.721]	[1.081]	[0.785]	[1.139]
Distance to nearest waterway		0.795*	0.495	0.562	0.503	0.614	0.533	0.515	0.427	0.596
		[0.452]	[0.410]	[0.488]	[0.516]	[0.524]	[0.549]	[0.442]	[0.415]	[0.459]
Executive constraints, 1960–2005 average				-0.057	-0.105	-0.085	-0.110	0.319		0.385
				[0.247]	[0.257]	[0.238]	[0.250]	[0.270]		[0.271]
Fraction of years under democracy, 1960–2005				-0.936	-0.543	-0.858	-0.551	-1.273		-1.229
				[1.076]	[1.029]	[1.050]	[1.030]	[1.090]		[1.137]
Fraction of years under autocracy, 1960–2005				-0.335	-0.258	-0.378	-0.275	0.087		0.609
				[0.891]	[0.871]	[0.891]	[0.865]	[0.774]		[0.754]
Log oil production per capita, 1960–2005 average								0.120*		0.122*
								[0.063]		[0.062]
Log population, 1960–2005 average								0.235		0.280
								[0.209]		[0.222]
Log GDP per capita, 1960–2005 average								-1.216***		-1.321***
								[0.280]		[0.286]
Effect of increasing genetic diversity	0.477**	0.893***	0.901**	0.761**	0.717*	0.799**	0.740**	0.906**	1.560**	1.781**
from the 10th to the 90th percentile	[0.212]	[0.287]	[0.430]	[0.384]	[0.381]	[0.375]	[0.371]	[0.380]	[0.626]	[0.734]
Continent dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Colonial history dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Sample	Global	Global	Global	Global	Global	Global	Global	Global	Old World	Old World
Observations	141	141	141	141	141	141	141	141	117	117
Pseudo R^2	0.011	0.121	0.191	0.218	0.226	0.221	0.226	0.280	0.173	0.274

Notes: This table exploits cross-country variations to establish a significant positive reduced-form impact of contemporary genetic diversity on the total count of new ethnic (WCM09) civil conflict onsets during the 1960–2005 time period, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict. For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of colonial history dummies includes indicators for whether a country was ever a colony of the U.K., France, and any other major colonizing power. Given the absence of a negative binomial estimator that permits instrumentation, the current analysis is unable to employ the strategy of exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated effect associated with increasing genetic diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the total number of new conflict onsets. Robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A.3: Genetic Diversity and the Frequency of Overall Civil Conflict Onset across Countries – The Analysis under Linguistic Fractionalization

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	2SLS	2SLS
	Log number of new PRIO25 civil conflict onsets per year during the 1960–2008 time period											
Genetic diversity (ancestry adjusted)	0.186**	0.448***	0.352**	0.397**	0.407**	0.420**	0.417**	0.449**	0.622**	0.889***	0.593**	0.838***
	[0.080]	[0.130]	[0.168]	[0.190]	[0.189]	[0.197]	[0.196]	[0.197]	[0.259]	[0.327]	[0.236]	[0.270]
Linguistic fractionalization					0.017*		0.015	0.016		0.023**		0.015
					[0.009]		[0.010]	[0.010]		[0.012]		[0.010]
Ethnolinguistic polarization						0.013	0.006	0.006		0.002		0.011
						[0.013]	[0.014]	[0.014]		[0.015]		[0.013]
Absolute latitude		-0.424***	-0.410	-0.332	-0.205	-0.352	-0.232	0.216	-0.294	0.399	-0.481**	0.065
		[0.119]	[0.253]	[0.262]	[0.271]	[0.256]	[0.255]	[0.265]	[0.297]	[0.340]	[0.244]	[0.260]
Land area		0.729	1.720	1.643	2.046	1.643	1.994	1.565	4.211	4.540*	1.487	1.160
		[2.129]	[2.374]	[2.480]	[2.510]	[2.529]	[2.543]	[2.833]	[2.809]	[2.720]	[2.304]	[2.751]
Ruggedness		0.039	0.025	0.026	0.030	0.028	0.031	0.056	0.041	0.085	0.032	0.062
		[0.039]	[0.045]	[0.045]	[0.045]	[0.046]	[0.046]	[0.045]	[0.052]	[0.053]	[0.044]	[0.040]
Mean elevation		-0.017*	-0.015	-0.017	-0.017	-0.018	-0.017	-0.020*	-0.020	-0.025**	-0.017*	-0.024**
		[0.009]	[0.010]	[0.011]	[0.011]	[0.011]	[0.011]	[0.010]	[0.013]	[0.012]	[0.010]	[0.010]
Range of elevation		0.010**	0.009**	0.009**	0.009**	0.008*	0.008*	0.005	0.009*	0.003	0.010**	0.005
		[0.005]	[0.005]	[0.004]	[0.004]	[0.005]	[0.004]	[0.004]	[0.006]	[0.005]	[0.004]	[0.004]
Mean land suitability		0.015	0.018	0.015	0.016	0.018	0.017	0.004	0.018	0.003	0.019	0.003
		[0.013]	[0.013]	[0.015]	[0.015]	[0.015]	[0.015]	[0.016]	[0.016]	[0.019]	[0.013]	[0.015]
Range of land suitability		0.012	0.016	0.013	0.010	0.016	0.012	0.008	0.019	0.001	0.018	0.010
		[0.009]	[0.011]	[0.013]	[0.012]	[0.015]	[0.015]	[0.017]	[0.013]	[0.017]	[0.011]	[0.015]
Distance to nearest waterway		0.008	0.004	0.007	0.003	0.007	0.003	-0.000	0.000	-0.005	0.004	-0.002
		[0.009]	[0.010]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.011]	[0.013]	[0.009]	[0.011]
Executive constraints, 1960–2008 average				0.004	0.003	0.004	0.003	0.005		0.004		0.006*
				[0.003]	[0.003]	[0.003]	[0.003]	[0.003]		[0.004]		[0.003]
Fraction of years under democracy, 1960–2008				-0.012	-0.010	-0.012	-0.011	-0.009		0.000		-0.011
				[0.019]	[0.019]	[0.019]	[0.019]	[0.019]		[0.020]		[0.017]
Fraction of years under autocracy, 1960–2008				-0.004	-0.005	-0.003	-0.004	-0.007		-0.010		-0.003
				[0.017]	[0.016]	[0.016]	[0.016]	[0.016]		[0.017]		[0.014]
Log oil production per capita, 1960–2008 average								0.002**		0.002*		0.002**
								[0.001]		[0.001]		[0.001]
Log population, 1960–2008 average								0.003		0.005		0.003
								[0.003]		[0.003]		[0.003]
Log GDP per capita, 1960–2008 average								-0.015***		-0.015***		-0.016***
								[0.005]		[0.005]		[0.004]
Effect of increasing genetic diversity	0.012**	0.029***	0.023**	0.026**	0.027**	0.027**	0.027**	0.029**	0.030**	0.042***	0.039**	0.055***
from the 10^{th} to the 90^{th} percentile	[0.005]	[0.008]	[0.011]	[0.012]	[0.012]	[0.013]	[0.013]	[0.013]	[0.012]	[0.016]	[0.015]	[0.018]
Continent dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Colonial history dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Sample	Global	Global	Global	Global	Global	Global	Global	Global	Old World	Old World	Global	Global
Observations	139	139	139	139	139	139	139	139	118	118	139	139
Partial R^2 of genetic diversity	–	0.127	0.050	0.060	0.064	0.066	0.066	0.080	0.091	0.159	–	–
Partial R^2 sum of other diversity measures	–	–	–	–	0.022	0.010	0.016	0.018	–	0.034	–	–
First-stage adjusted R^2	–	–	–	–	–	–	–	–	–	–	0.788	0.781
First-stage partial R^2 of migratory distance	–	–	–	–	–	–	–	–	–	–	0.517	0.475
First-stage F statistic	–	–	–	–	–	–	–	–	–	–	220.899	112.736
Adjusted R^2	0.019	0.204	0.208	0.203	0.213	0.204	0.208	0.253	0.260	0.333	–	–

Notes: This table exploits cross-country variations to establish that the significant positive reduced-form impact of contemporary genetic diversity on the annual frequency of new overall (PRIO25) civil conflict onsets during the 1960–2008 time period, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict, is robust to accounting for the influence of linguistic rather than ethnic fractionalization. For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of colonial history dummies includes indicators for whether a country was ever a colony of the U.K., France, and any other major colonizing power. The 2SLS regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated effect associated with increasing genetic diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the actual number of new conflict onsets per year. Robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A.4: Genetic Diversity and the Frequency of Ethnic Civil Conflict Onset across Countries – The Analysis under Linguistic Fractionalization

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) 2SLS	(12) 2SLS
	Log number of new WCM09 ethnic civil conflict onsets per year during the 1960–2005 time period											
Genetic diversity (ancestry adjusted)	0.205**	0.440***	0.408**	0.416**	0.434**	0.439**	0.431**	0.458**	0.757***	0.953***	0.699***	0.863***
	[0.085]	[0.129]	[0.169]	[0.198]	[0.191]	[0.204]	[0.201]	[0.197]	[0.278]	[0.343]	[0.258]	[0.283]
Linguistic fractionalization					0.034***		0.035***	0.033***		0.039***		0.032***
					[0.011]		[0.012]	[0.011]		[0.013]		[0.011]
Ethnolinguistic polarization						0.013	-0.002	-0.005		-0.014		0.000
						[0.016]	[0.016]	[0.015]		[0.016]		[0.014]
Absolute latitude		-0.383***	-0.511*	-0.393	-0.140	-0.414	-0.133	0.456	-0.458	0.581	-0.598**	0.297
		[0.133]	[0.273]	[0.293]	[0.293]	[0.284]	[0.271]	[0.290]	[0.351]	[0.386]	[0.246]	[0.263]
Land area		-0.933	0.026	0.342	1.139	0.352	1.152	1.480	2.065	3.697*	-0.248	1.051
		[1.515]	[1.809]	[1.876]	[1.928]	[1.923]	[1.929]	[1.976]	[2.128]	[1.918]	[1.793]	[1.967]
Ruggedness		0.039	-0.001	-0.003	0.004	-0.001	0.004	0.028	0.024	0.067	0.006	0.034
		[0.043]	[0.048]	[0.046]	[0.046]	[0.047]	[0.047]	[0.045]	[0.055]	[0.053]	[0.046]	[0.040]
Mean elevation		-0.015	-0.009	-0.008	-0.008	-0.009	-0.008	-0.013	-0.016	-0.022*	-0.011	-0.017*
		[0.009]	[0.010]	[0.011]	[0.011]	[0.012]	[0.011]	[0.010]	[0.012]	[0.013]	[0.010]	[0.010]
Range of elevation		0.008*	0.006	0.006	0.005	0.005	0.005	0.002	0.007	0.001	0.007	0.003
		[0.005]	[0.004]	[0.004]	[0.004]	[0.005]	[0.004]	[0.004]	[0.006]	[0.005]	[0.004]	[0.004]
Mean land suitability		0.013	0.016	0.013	0.013	0.015	0.013	-0.000	0.015	-0.005	0.017	-0.002
		[0.015]	[0.014]	[0.017]	[0.016]	[0.017]	[0.016]	[0.016]	[0.017]	[0.020]	[0.014]	[0.015]
Range of land suitability		0.023**	0.031**	0.031**	0.025*	0.034**	0.025	0.028	0.034**	0.020	0.034**	0.030*
		[0.010]	[0.013]	[0.014]	[0.013]	[0.016]	[0.016]	[0.019]	[0.016]	[0.020]	[0.014]	[0.017]
Distance to nearest waterway		0.009	0.004	0.003	-0.006	0.002	-0.006	-0.012	0.000	-0.015	0.003	-0.014
		[0.009]	[0.010]	[0.012]	[0.013]	[0.012]	[0.013]	[0.012]	[0.011]	[0.012]	[0.010]	[0.010]
Executive constraints, 1960–2005 average				-0.000	-0.002	0.000	-0.002	0.002		0.002		0.003
				[0.004]	[0.004]	[0.004]	[0.004]	[0.004]		[0.005]		[0.004]
Fraction of years under democracy, 1960–2005				-0.009	-0.005	-0.009	-0.005	-0.006		0.003		-0.009
				[0.025]	[0.024]	[0.025]	[0.025]	[0.024]		[0.024]		[0.021]
Fraction of years under autocracy, 1960–2005				-0.012	-0.014	-0.012	-0.014	-0.016		-0.014		-0.013
				[0.020]	[0.019]	[0.020]	[0.019]	[0.018]		[0.020]		[0.016]
Log oil production per capita, 1960–2005 average								0.003**		0.002*		0.003**
								[0.001]		[0.001]		[0.001]
Log population, 1960–2005 average								-0.000		0.002		0.000
								[0.003]		[0.004]		[0.003]
Log GDP per capita, 1960–2005 average								-0.022***		-0.022***		-0.022***
								[0.005]		[0.005]		[0.005]
Effect of increasing genetic diversity	0.013**	0.029***	0.027**	0.027**	0.028**	0.028**	0.028**	0.030**	0.036***	0.045***	0.045***	0.056***
from the 10^{th} to the 90^{th} percentile	[0.006]	[0.008]	[0.011]	[0.013]	[0.012]	[0.013]	[0.013]	[0.013]	[0.013]	[0.016]	[0.017]	[0.018]
Continent dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Colonial history dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Sample	Global	Global	Global	Global	Global	Global	Global	Global	Old World	Old World	Global	Global
Observations	137	137	137	137	137	137	137	137	116	116	137	137
Partial R^2 of genetic diversity	–	0.098	0.054	0.052	0.060	0.057	0.058	0.071	0.104	0.146	–	–
Partial R^2 sum of other diversity measures	–	–	–	–	0.063	0.008	0.057	0.057	–	0.080	–	–
First-stage adjusted R^2	–	–	–	–	–	–	–	–	–	–	0.786	0.779
First-stage partial R^2 of migratory distance	–	–	–	–	–	–	–	–	–	–	0.515	0.474
First-stage F statistic	–	–	–	–	–	–	–	–	–	–	215.900	109.242
Adjusted R^2	0.020	0.129	0.166	0.133	0.181	0.132	0.174	0.246	0.196	0.283	–	–

Notes: This table exploits cross-country variations to establish that the significant positive reduced-form impact of contemporary genetic diversity on the annual frequency of new ethnic (WCM09) civil conflict onsets during the 1960–2005 time period, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict, is robust to accounting for the influence of linguistic rather than ethnic fractionalization. For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of colonial history dummies includes indicators for whether a country was ever a colony of the U.K., France, and any other major colonizing power. The 2SLS regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated effect associated with increasing genetic diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the actual number of new conflict onsets per year. Robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A.5: Genetic Diversity and the Frequency of Overall Civil Conflict Onset across Countries – Robustness to Accounting for Spatial Dependence

	(1) Spatial OLS	(2) Spatial OLS	(3) Spatial OLS	(4) Spatial OLS	(5) Spatial OLS	(6) Spatial OLS	(7) Spatial OLS	(8) Spatial OLS	(9) Spatial OLS	(10) Spatial OLS	(11) Spatial GMM	(12) Spatial GMM
	Log number of new PRIO25 civil conflict onsets per year during the 1960–2008 time period											
Genetic diversity (ancestry adjusted)	0.182***	0.422***	0.322**	0.366**	0.350**	0.390**	0.377**	0.398**	0.639***	0.855***	0.599***	0.805***
	[0.029]	[0.086]	[0.150]	[0.166]	[0.161]	[0.157]	[0.160]	[0.162]	[0.137]	[0.163]	[0.138]	[0.160]
Ethnic fractionalization					0.011**		0.006	0.007*		0.012***		-0.002
					[0.005]		[0.004]	[0.004]		[0.004]		[0.008]
Ethnolinguistic polarization						0.013*	0.010	0.010		0.007		0.019**
						[0.007]	[0.007]	[0.006]		[0.007]		[0.008]
Absolute latitude		-0.404***	-0.440***	-0.331***	-0.225**	-0.356***	-0.292***	0.149*	-0.333*	0.255***	-0.529**	-0.116
		[0.055]	[0.161]	[0.082]	[0.115]	[0.074]	[0.099]	[0.077]	[0.193]	[0.081]	[0.208]	[0.177]
Land area		0.765	1.825*	1.709	1.972*	1.719	1.862*	1.586	4.177***	4.114***	1.626	1.311
		[1.047]	[0.978]	[1.104]	[1.132]	[1.109]	[1.101]	[1.112]	[1.013]	[0.801]	[1.187]	[1.338]
Ruggedness		0.038***	0.028**	0.030**	0.036***	0.032**	0.035**	0.056***	0.041**	0.080***	0.034**	0.054***
		[0.014]	[0.013]	[0.013]	[0.013]	[0.014]	[0.014]	[0.013]	[0.017]	[0.020]	[0.017]	[0.020]
Mean elevation		-0.016**	-0.015***	-0.017***	-0.018***	-0.018***	-0.018***	-0.020***	-0.019***	-0.025***	-0.016***	-0.023***
		[0.006]	[0.005]	[0.003]	[0.003]	[0.003]	[0.003]	[0.004]	[0.006]	[0.006]	[0.006]	[0.006]
Range of elevation		0.009***	0.009***	0.009***	0.008***	0.008***	0.008***	0.004**	0.009***	0.003*	0.010***	0.005***
		[0.003]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Mean land suitability		0.013***	0.018***	0.016***	0.019***	0.019***	0.020***	0.006	0.018***	0.006	0.018***	0.003
		[0.004]	[0.003]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.005]
Range of land suitability		0.013***	0.014***	0.012***	0.011***	0.014***	0.013***	0.010**	0.019***	0.007	0.017***	0.015**
		[0.002]	[0.005]	[0.004]	[0.003]	[0.005]	[0.005]	[0.004]	[0.004]	[0.004]	[0.003]	[0.006]
Distance to nearest waterway		0.008**	0.005	0.007	0.006	0.007	0.007	0.002	0.000	-0.001	0.004	0.001
		[0.004]	[0.004]	[0.006]	[0.006]	[0.005]	[0.005]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Executive constraints, 1960–2008 average				0.004***	0.004**	0.004**	0.004**	0.006***		0.005***		0.008***
				[0.001]	[0.001]	[0.002]	[0.002]	[0.001]		[0.001]		[0.002]
Fraction of years under democracy, 1960–2008				-0.015*	-0.014	-0.015*	-0.014	-0.012*		-0.002		-0.017
				[0.009]	[0.009]	[0.009]	[0.009]	[0.007]		[0.008]		[0.010]
Fraction of years under autocracy, 1960–2008				-0.006	-0.005	-0.005	-0.005	-0.008		-0.009*		-0.007
				[0.005]	[0.005]	[0.005]	[0.005]	[0.006]		[0.005]		[0.005]
Log oil production per capita, 1960–2008 average								0.002***		0.002***		0.002***
								[0.000]		[0.000]		[0.000]
Log population, 1960–2008 average								0.003**		0.004***		0.003**
								[0.001]		[0.001]		[0.001]
Log GDP per capita, 1960–2008 average								-0.015***		-0.016***		-0.016***
								[0.001]		[0.001]		[0.001]
Continent dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Colonial history dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Sample	Global	Global	Global	Global	Global	Global	Global	Global	Old World	Old World	Global	Global
Observations	143	143	143	143	143	143	143	143	119	119	143	143
Adjusted R^2	0.325	0.443	0.447	0.444	0.443	0.445	0.441	0.473	0.496	0.531	–	–

Notes: This table exploits cross-country variations to establish that the significant positive reduced-form impact of contemporary genetic diversity on the annual frequency of new overall (PRIO25) civil conflict onsets during the 1960–2008 time period, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict, is robust to accounting for spatial dependence across observations, following the method of Conley (1999). To perform this robustness check, the spatial distribution of observations is specified on the Euclidean plane using the full set of pairwise geodesic distances between country centroids, and the spatial autoregressive process across residuals is modeled as varying inversely with distance from each observation up to a maximum threshold of 25,000 kilometers, thus admitting the possibility of spatial dependence at a global scale. For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of colonial history dummies includes indicators for whether a country was ever a colony of the U.K., France, and any other major colonizing power. The GMM regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. Standard errors, corrected for spatial autocorrelation, are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A.6: Genetic Diversity and the Frequency of Ethnic Civil Conflict Onset across Countries – Robustness to Accounting for Spatial Dependence

	(1) Spatial OLS	(2) Spatial OLS	(3) Spatial OLS	(4) Spatial OLS	(5) Spatial OLS	(6) Spatial OLS	(7) Spatial OLS	(8) Spatial OLS	(9) Spatial OLS	(10) Spatial OLS	(11) Spatial GMM	(12) Spatial GMM
	Log number of new WCM09 ethnic civil conflict onsets per year during the 1960–2005 time period											
Genetic diversity (ancestry adjusted)	0.217***	0.418***	0.375**	0.385**	0.352**	0.408***	0.366**	0.391**	0.780***	0.904***	0.707***	0.795***
	[0.051]	[0.087]	[0.172]	[0.170]	[0.160]	[0.158]	[0.158]	[0.161]	[0.103]	[0.109]	[0.125]	[0.122]
Ethnic fractionalization					0.021***		0.019***	0.016***		0.021***		0.005
					[0.004]		[0.005]	[0.006]		[0.006]		[0.008]
Ethnolinguistic polarization						0.012**	0.005	0.002		-0.008		0.011
						[0.006]	[0.007]	[0.006]		[0.007]		[0.010]
Absolute latitude		-0.366***	-0.564***	-0.419***	-0.215	-0.445***	-0.250**	0.292***	-0.537***	0.298***	-0.673***	0.017
		[0.071]	[0.113]	[0.132]	[0.140]	[0.118]	[0.120]	[0.107]	[0.144]	[0.096]	[0.143]	[0.134]
Land area		-0.917	0.193	0.528	1.018	0.551	0.969	1.365	2.072**	2.920***	-0.034	1.109
		[0.795]	[0.880]	[0.961]	[0.975]	[0.969]	[0.971]	[0.887]	[1.015]	[0.745]	[1.121]	[1.070]
Ruggedness		0.033***	-0.002	-0.003	0.007	-0.001	0.007	0.025*	0.023	0.062***	0.004	0.022
		[0.013]	[0.014]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.015]	[0.016]	[0.019]	[0.023]
Mean elevation		-0.012	-0.006	-0.007	-0.007*	-0.007	-0.007*	-0.011*	-0.014*	-0.020***	-0.008	-0.014**
		[0.008]	[0.006]	[0.005]	[0.004]	[0.005]	[0.004]	[0.006]	[0.008]	[0.007]	[0.008]	[0.007]
Range of elevation		0.007*	0.006*	0.005*	0.004	0.005*	0.004	0.001	0.006*	0.000	0.006**	0.002
		[0.004]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.002]	[0.004]	[0.003]	[0.003]	[0.002]
Mean land suitability		0.013***	0.017***	0.013**	0.019***	0.016***	0.019***	0.005	0.017***	0.001	0.017***	0.001
		[0.005]	[0.003]	[0.005]	[0.006]	[0.006]	[0.006]	[0.006]	[0.005]	[0.007]	[0.005]	[0.006]
Range of land suitability		0.026***	0.030***	0.031***	0.029***	0.033***	0.030***	0.034***	0.033***	0.026***	0.034***	0.038***
		[0.003]	[0.006]	[0.005]	[0.005]	[0.005]	[0.006]	[0.006]	[0.007]	[0.006]	[0.005]	[0.008]
Distance to nearest waterway		0.009	0.004	0.003	0.001	0.003	0.002	-0.005	0.001	-0.007*	0.003	-0.007
		[0.005]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.005]	[0.006]	[0.004]	[0.005]	[0.004]
Executive constraints, 1960–2005 average				-0.000	-0.001	-0.000	-0.001	0.003**		0.003**		0.005***
				[0.001]	[0.001]	[0.001]	[0.001]	[0.001]		[0.002]		[0.002]
Fraction of years under democracy, 1960–2005				-0.011	-0.007	-0.011	-0.007	-0.008		0.003		-0.013
				[0.011]	[0.011]	[0.011]	[0.011]	[0.010]		[0.010]		[0.012]
Fraction of years under autocracy, 1960–2005				-0.015*	-0.013	-0.014	-0.013	-0.015**		-0.011		-0.015*
				[0.009]	[0.008]	[0.009]	[0.008]	[0.007]		[0.008]		[0.008]
Log oil production per capita, 1960–2005 average								0.003***		0.003***		0.003***
								[0.001]		[0.001]		[0.001]
Log population, 1960–2005 average								-0.000		0.002**		-0.000
								[0.001]		[0.001]		[0.001]
Log GDP per capita, 1960–2005 average								-0.021***		-0.023***		-0.022***
								[0.002]		[0.002]		[0.002]
Continent dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Colonial history dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Sample	Global	Global	Global	Global	Global	Global	Global	Global	Old World	Old World	Global	Global
Observations	141	141	141	141	141	141	141	141	117	117	141	141
Adjusted R^2	0.239	0.319	0.349	0.328	0.334	0.327	0.329	0.381	0.398	0.427	–	–

Notes: This table exploits cross-country variations to establish that the significant positive reduced-form impact of contemporary genetic diversity on the annual frequency of new ethnic (WCM09) civil conflict onsets during the 1960–2005 time period, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict, is robust to accounting for spatial dependence across observations, following the method of Conley (1999). To perform this robustness check, the spatial distribution of observations is specified on the Euclidean plane using the full set of pairwise geodesic distances between country centroids, and the spatial autoregressive process across residuals is modeled as varying inversely with distance from each observation up to a maximum threshold of 25,000 kilometers, thus admitting the possibility of spatial dependence at a global scale. For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of colonial history dummies includes indicators for whether a country was ever a colony of the U.K., France, and any other major colonizing power. The GMM regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. Standard errors, corrected for spatial autocorrelation, are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A.7: Genetic Diversity and the Frequency of Overall Civil Conflict Onset across Countries in the Ethnic Civil Conflict Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	2SLS	2SLS
	Log number of new WCM09 civil conflict onsets per year during the 1960–2005 time period											
Genetic diversity (ancestry adjusted)	0.219**	0.529***	0.598***	0.615***	0.593**	0.642***	0.623**	0.665***	1.017***	1.225***	0.908***	1.100***
	[0.110]	[0.151]	[0.191]	[0.231]	[0.227]	[0.241]	[0.247]	[0.240]	[0.345]	[0.417]	[0.316]	[0.352]
Ethnic fractionalization					0.014		0.008	0.007		0.016		-0.004
					[0.015]		[0.017]	[0.018]		[0.022]		[0.019]
Ethnolinguistic polarization						0.014	0.011	0.008		-0.002		0.018
						[0.018]	[0.020]	[0.018]		[0.020]		[0.017]
Absolute latitude		-0.741***	-0.922**	-0.688*	-0.554	-0.718*	-0.632	0.077	-0.889*	0.024	-1.024***	-0.219
		[0.179]	[0.373]	[0.391]	[0.453]	[0.376]	[0.421]	[0.386]	[0.473]	[0.515]	[0.331]	[0.375]
Land area		-2.269	-1.114	-0.465	-0.141	-0.438	-0.252	-0.214	1.732	2.600	-1.326	-0.490
		[2.405]	[2.784]	[2.858]	[2.913]	[2.895]	[2.970]	[3.161]	[3.293]	[3.371]	[2.748]	[3.058]
Ruggedness		0.068	0.031	0.037	0.044	0.040	0.043	0.070	0.055	0.108	0.037	0.068
		[0.055]	[0.060]	[0.059]	[0.063]	[0.060]	[0.063]	[0.063]	[0.071]	[0.075]	[0.058]	[0.056]
Mean elevation		-0.026**	-0.019	-0.023	-0.023	-0.024	-0.024	-0.028*	-0.026	-0.037**	-0.021	-0.031**
		[0.012]	[0.013]	[0.016]	[0.016]	[0.016]	[0.016]	[0.014]	[0.017]	[0.018]	[0.013]	[0.014]
Range of elevation		0.013***	0.010**	0.010*	0.009	0.009*	0.009	0.005	0.010	0.003	0.011**	0.006
		[0.005]	[0.005]	[0.005]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.007]	[0.005]	[0.006]
Mean land suitability		0.018	0.021	0.011	0.015	0.014	0.015	-0.007	0.018	-0.014	0.021	-0.011
		[0.017]	[0.019]	[0.021]	[0.020]	[0.021]	[0.021]	[0.023]	[0.023]	[0.028]	[0.018]	[0.021]
Range of land suitability		0.022*	0.030*	0.030*	0.029*	0.033*	0.032*	0.032	0.038**	0.027	0.033**	0.037*
		[0.012]	[0.016]	[0.016]	[0.016]	[0.018]	[0.018]	[0.021]	[0.019]	[0.024]	[0.017]	[0.020]
Distance to nearest waterway		0.022	0.017	0.012	0.011	0.012	0.011	0.004	0.012	-0.002	0.016	0.002
		[0.014]	[0.016]	[0.018]	[0.018]	[0.018]	[0.018]	[0.018]	[0.017]	[0.019]	[0.015]	[0.016]
Executive constraints, 1960–2005 average				-0.004	-0.004	-0.004	-0.004	-0.000		-0.000		0.002
				[0.005]	[0.005]	[0.005]	[0.005]	[0.005]		[0.006]		[0.005]
Fraction of years under democracy, 1960–2005				-0.012	-0.009	-0.012	-0.010	-0.010		0.003		-0.016
				[0.030]	[0.029]	[0.030]	[0.030]	[0.028]		[0.028]		[0.026]
Fraction of years under autocracy, 1960–2005				-0.028	-0.026	-0.027	-0.026	-0.030		-0.030		-0.029
				[0.026]	[0.025]	[0.026]	[0.025]	[0.025]		[0.027]		[0.022]
Log oil production per capita, 1960–2005 average								0.003*		0.003		0.003*
								[0.002]		[0.002]		[0.002]
Log population, 1960–2005 average								0.002		0.004		0.002
								[0.004]		[0.005]		[0.004]
Log GDP per capita, 1960–2005 average								-0.026***		-0.028***		-0.027***
								[0.008]		[0.008]		[0.007]
Effect of increasing genetic diversity	0.014**	0.035***	0.039***	0.041***	0.039***	0.042***	0.041**	0.044***	0.051***	0.061***	0.060***	0.072***
from the 10th to the 90th percentile	[0.007]	[0.010]	[0.013]	[0.015]	[0.015]	[0.016]	[0.016]	[0.016]	[0.017]	[0.021]	[0.021]	[0.023]
Continent dummies	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Colonial history dummies	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Sample	Global	Global	Global	Global	Global	Global	Global	Global	Old World	Old World	Global	Global
Observations	141	141	141	141	141	141	141	141	117	117	141	141
Partial R^2 of genetic diversity	–	0.095	0.079	0.079	0.073	0.084	0.075	0.092	0.119	0.145	–	–
Partial R^2 sum of other diversity measures	–	–	–	–	0.005	0.006	0.004	0.003	–	0.005	–	–
First-stage adjusted R^2	–	–	–	–	–	–	–	–	–	–	0.753	0.760
First-stage partial R^2 of migratory distance	–	–	–	–	–	–	–	–	–	–	0.475	0.438
First-stage F statistic	–	–	–	–	–	–	–	–	–	–	206.014	97.246
Adjusted R^2	0.012	0.210	0.237	0.220	0.217	0.218	0.213	0.277	0.285	0.326	–	–

Notes: This table exploits cross-country variations to establish a significant positive reduced-form impact of contemporary genetic diversity on the annual frequency of new overall civil conflict onsets during the 1960–2005 time period, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict, using the same conflict sample and source data (i.e., WCM09) as the ones employed for the analyses of ethnic civil conflict throughout the paper. The analysis thus provides an appropriate benchmark for assessing the influence of genetic diversity on overall versus ethnic civil conflict. For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of colonial history dummies includes indicators for whether a country was ever a colony of the U.K., France, and any other major colonizing power. The 2SLS regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated effect associated with increasing genetic diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the actual number of new conflict onsets per year. Robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A.8: Genetic Diversity and the Frequency of Overall Civil Conflict Onset across Countries – Robustness to the Elimination of Regions from the Estimation Sample

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) 2SLS	(7) 2SLS	(8) 2SLS	(9) 2SLS	(10) 2SLS
	Log number of new PRIO25 civil conflict onsets per year during the 1960–2008 time period									
Genetic diversity (ancestry adjusted)	0.411***	0.505***	0.437***	0.340**	0.578**	0.613***	1.210***	0.637***	0.616***	0.736***
	[0.142]	[0.179]	[0.154]	[0.151]	[0.226]	[0.201]	[0.379]	[0.204]	[0.223]	[0.248]
Effect of increasing genetic diversity	0.027***	0.031***	0.030***	0.021**	0.027**	0.040***	0.074***	0.044***	0.037***	0.034***
from the 10^{th} to the 90^{th} percentile	[0.009]	[0.011]	[0.011]	[0.009]	[0.011]	[0.013]	[0.023]	[0.014]	[0.014]	[0.012]
Continent dummies	No	No	No	No	No	No	No	No	No	No
All other baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Global	Global	Global	Global	Global	Global	Global	Global	Global	Global
Omitted region	None	SSA	MENA	EAP	LAC	None	SSA	MENA	EAP	LAC
Observations	143	102	127	128	123	143	102	127	128	123
Partial R^2 of genetic diversity	0.100	0.107	0.111	0.060	0.089	–	–	–	–	–
Partial R^2 sum of other diversity measures	0.015	0.007	0.016	0.029	0.006	–	–	–	–	–
First-stage adjusted R^2	–	–	–	–	–	0.688	0.635	0.686	0.697	0.804
First-stage partial R^2 of migratory distance	–	–	–	–	–	0.510	0.282	0.519	0.508	0.703
First-stage F statistic	–	–	–	–	–	71.295	25.274	72.047	56.085	69.174
Adjusted R^2	0.233	0.169	0.236	0.239	0.255	–	–	–	–	–

Notes: This table exploits cross-country variations to establish that the significant positive reduced-form impact of contemporary genetic diversity on the annual frequency of new overall (PRIO25) civil conflict onsets during the 1960–2008 time period is robust to the one-at-a-time elimination of world regions from the global sample, including Sub-Saharan Africa (SSA), Middle East and North Africa (MENA), East Asia and Pacific (EAP), and Latin America and the Caribbean (LAC). With the exception of continent dummies, all regressions include controls for other well-known diversity measures (namely, ethnic fractionalization and ethnolinguistic polarization) as well as the proximate geographical, institutional, and development-related correlates of conflict, as considered by the baseline analysis of overall civil conflict frequency in Table 3. In light of the lower degrees of freedom afforded by the regression samples with eliminated regions, the omission of continent dummies from the specification reflects the need to preserve as much of the cross-country variation in conflict as possible in order to permit the independent variables to possess some explanatory power, and the regressions presented in Columns 1 and 6 should therefore be used as the relevant baselines for assessing the current robustness exercise. The 2SLS regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated effect associated with increasing genetic diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the actual number of new conflict onsets per year. Robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A.9: Genetic Diversity and the Frequency of Ethnic Civil Conflict Onset across Countries – Robustness to the Elimination of Regions from the Estimation Sample

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) 2SLS	(7) 2SLS	(8) 2SLS	(9) 2SLS	(10) 2SLS
	Log number of new WCM09 ethnic civil conflict onsets per year during the 1960–2005 time period									
Genetic diversity (ancestry adjusted)	0.350**	0.376*	0.375**	0.354**	0.451*	0.575***	0.875**	0.634***	0.586***	0.516**
	[0.141]	[0.195]	[0.151]	[0.144]	[0.244]	[0.185]	[0.348]	[0.183]	[0.211]	[0.238]
Effect of increasing genetic diversity	0.023**	0.023*	0.026**	0.021**	0.021*	0.037***	0.053**	0.043***	0.036***	0.024**
from the 10th to the 90th percentile	[0.009]	[0.012]	[0.010]	[0.009]	[0.011]	[0.012]	[0.021]	[0.013]	[0.013]	[0.011]
Continent dummies	No	No	No	No	No	No	No	No	No	No
All other baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Global	Global	Global	Global	Global	Global	Global	Global	Global	Global
Omitted region	None	SSA	MENA	EAP	LAC	None	SSA	MENA	EAP	LAC
Observations	141	101	126	126	121	141	101	126	126	121
Partial R^2 of genetic diversity	0.056	0.053	0.066	0.049	0.042	–	–	–	–	–
Partial R^2 sum of other diversity measures	0.007	0.028	0.005	0.007	0.002	–	–	–	–	–
First-stage adjusted R^2	–	–	–	–	–	0.690	0.639	0.688	0.700	0.807
First-stage partial R^2 of migratory distance	–	–	–	–	–	0.495	0.288	0.496	0.491	0.704
First-stage F statistic	–	–	–	–	–	68.639	25.768	67.661	53.825	69.922
Adjusted R^2	0.163	0.185	0.168	0.135	0.153	–	–	–	–	–

Notes: This table exploits cross-country variations to establish that the significant positive reduced-form impact of contemporary genetic diversity on the annual frequency of new ethnic (WCM09) civil conflict onsets during the 1960–2005 time period is robust to the one-at-a-time elimination of world regions from the global sample, including Sub-Saharan Africa (SSA), Middle East and North Africa (MENA), East Asia and Pacific (EAP), and Latin America and the Caribbean (LAC). With the exception of continent dummies, all regressions include controls for other well-known diversity measures (namely, ethnic fractionalization and ethnolinguistic polarization) as well as the proximate geographical, institutional, and development-related correlates of conflict, as considered by the baseline analysis of ethnic civil conflict frequency in Table 4. In light of the lower degrees of freedom afforded by the regression samples with eliminated regions, the omission of continent dummies from the specification reflects the need to preserve as much of the cross-country variation in conflict as possible in order to permit the independent variables to possess some explanatory power, and the regressions presented in Columns 1 and 6 should therefore be used as the relevant baselines for assessing the current robustness exercise. The 2SLS regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated effect associated with increasing genetic diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the actual number of new conflict onsets per year. Robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A.10: Genetic Diversity and the Incidence of Civil Conflict in Annually Repeated Cross-Country Data

	(1) Probit	(2) Probit	(3) IV Probit	(4) IV Probit	(5) Probit	(6) Probit	(7) IV Probit	(8) IV Probit
	PRIO25 civil conflict incidence				WCM09 ethnic civil conflict incidence			
Genetic diversity (ancestry adjusted)	9.439***	10.033***	10.631***	12.154***	12.918***	12.265***	9.379**	8.197*
	[3.627]	[3.609]	[3.521]	[3.975]	[3.958]	[4.137]	[3.662]	[4.361]
Ethnic fractionalization		-0.322		-0.496*		0.027		-0.400
		[0.302]		[0.299]		[0.374]		[0.391]
Ethnolinguistic polarization		0.192		0.470		-0.084		0.462
		[0.275]		[0.288]		[0.346]		[0.430]
Absolute latitude	-11.528**	-7.624	-17.757***	-18.740**	-16.524***	-9.058	-21.623***	-22.114**
	[5.141]	[7.017]	[5.141]	[7.657]	[6.354]	[9.143]	[5.907]	[9.470]
Land area	27.165	17.807	-2.623	-16.975	16.586	17.083	2.588	2.836
	[28.637]	[28.302]	[33.877]	[37.375]	[30.667]	[35.924]	[41.574]	[48.197]
Ruggedness	0.939	1.873**	0.462	0.889	0.720	1.613	-0.430	-0.434
	[0.934]	[0.928]	[0.802]	[0.846]	[0.968]	[1.054]	[1.107]	[1.346]
Mean elevation	-0.417**	-0.550***	-0.255	-0.336*	-0.441*	-0.585**	0.071	0.053
	[0.212]	[0.206]	[0.172]	[0.176]	[0.239]	[0.231]	[0.274]	[0.296]
Range of elevation	0.096*	0.032	0.124**	0.079	0.025	-0.037	-0.059	-0.087
	[0.056]	[0.077]	[0.057]	[0.073]	[0.060]	[0.080]	[0.079]	[0.098]
Mean land suitability	0.231	0.015	0.153	-0.025	0.420	0.449	0.167	0.361
	[0.245]	[0.302]	[0.254]	[0.317]	[0.274]	[0.355]	[0.333]	[0.411]
Range of land suitability	0.621***	0.703**	0.589***	0.719**	0.953***	1.129***	1.056***	1.453***
	[0.238]	[0.290]	[0.226]	[0.290]	[0.304]	[0.381]	[0.358]	[0.453]
Distance to nearest waterway	0.113	0.245	0.113	0.224	0.281	0.435**	0.141	0.291
	[0.150]	[0.158]	[0.136]	[0.158]	[0.194]	[0.219]	[0.197]	[0.214]
Executive constraints, lagged		0.057		0.043		0.099*		0.067
		[0.038]		[0.037]		[0.055]		[0.053]
Democracy dummy, lagged		-0.293*		-0.306**		-0.315*		-0.414***
		[0.158]		[0.140]		[0.178]		[0.159]
Autocracy dummy, lagged		-0.169		-0.281**		-0.107		-0.322*
		[0.115]		[0.114]		[0.162]		[0.174]
Log oil production per capita, lagged		0.009		0.004		0.021		0.013
		[0.022]		[0.021]		[0.023]		[0.025]
Log population, lagged		0.054		0.058		0.009		-0.031
		[0.060]		[0.057]		[0.084]		[0.084]
Log GDP per capita, lagged		-0.171**		-0.115		-0.184**		-0.076
		[0.075]		[0.071]		[0.090]		[0.126]
Conflict incidence, lagged	2.725***	2.645***	2.690***	2.622***	3.685***	3.610***	3.470***	3.366***
	[0.122]	[0.119]	[0.116]	[0.116]	[0.149]	[0.147]	[0.186]	[0.199]
Marginal effect	0.885**	0.921***	1.045***	1.177***	0.592***	0.549***	0.541**	0.484*
	[0.346]	[0.343]	[0.381]	[0.434]	[0.184]	[0.188]	[0.219]	[0.262]
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	Yes	No	Yes	No	Yes	No	Yes
Colonial history dummies	No	Yes	No	Yes	No	Yes	No	Yes
Sample	Old World	Old World	Global	Global	Old World	Old World	Global	Global
Observations	4,756	4,756	5,797	5,797	4,331	4,316	4,855	4,840
Countries	119	119	141	141	117	117	129	129
Time horizon	1960–2008	1960–2008	1960–2008	1960–2008	1960–2005	1960–2005	1960–2005	1960–2005
Time frequency	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual
Pseudo R^2	0.617	0.627	–	–	0.785	0.792	–	–

Notes: This table exploits variations in an annually repeated cross-section of countries to establish a significant positive reduced-form impact of contemporary genetic diversity on the likelihood of observing the incidence of (i) an overall (PRIO25) civil conflict in any given year during the 1960–2008 time horizon; and (ii) an ethnic (WCM09) civil conflict in any given year during the 1960–2005 time horizon, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict. To account for temporal dependence in conflict outcomes, all regressions control for the incidence of conflict in the previous year, following Esteban, Mayoral and Ray (2012). For regressions based on the global sample, the set of continent dummies includes indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes indicators for Africa and Asia, implying that Europe is treated as the omitted category in all cases. The set of legal origin dummies includes indicators for British and French legal origins, and the set of time-varying (lagged) colonial history dummies includes indicators for whether a country was a colony of the U.K., France, and any other major colonizing power in the previous year. The IV probit regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated marginal effect of a 1 percentage point increase in genetic diversity is the average marginal effect across the entire cross-section of observed diversity values, and it reflects the increase in the likelihood of a conflict incidence in any given year, expressed in percentage points. Robust standard errors, clustered at the country level, are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A.11: Genetic Diversity and the Incidence of Civil Conflict in Quinquennially Repeated Cross-Country Data – Robustness to Employing the Logit and Rare-Events Logit Estimators

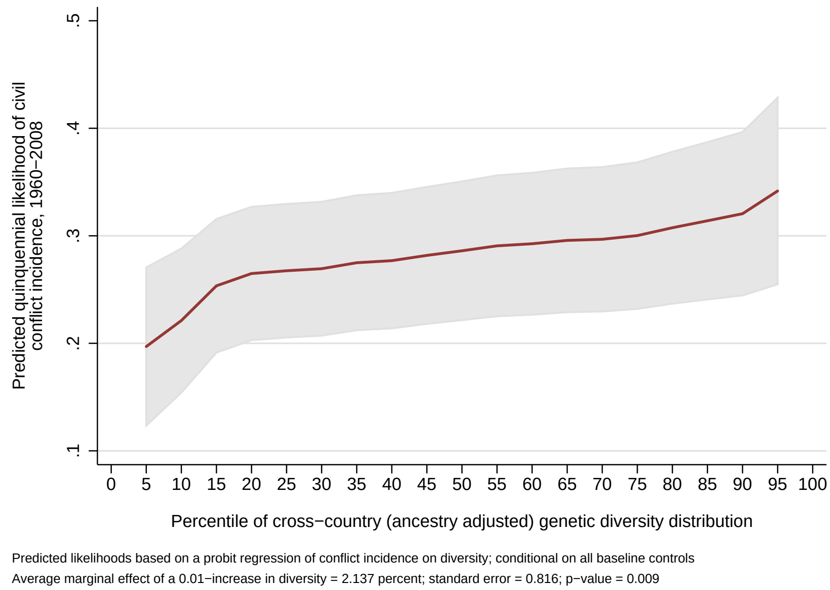
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Logit	Logit	RE Logit	RE Logit	Logit	Logit	RE Logit	RE Logit
	PRIO25 civil conflict incidence				WCM09 ethnic civil conflict incidence			
Genetic diversity (ancestry adjusted)	22.265***	21.459**	21.485***	20.122**	38.212***	40.717***	36.234***	36.533***
	[8.137]	[8.332]	[7.952]	[8.034]	[11.150]	[11.820]	[10.891]	[11.390]
Ethnic fractionalization		-0.528		-0.504		0.294		0.276
		[0.678]		[0.654]		[0.920]		[0.886]
Ethnolinguistic polarization		0.351		0.355		0.148		0.158
		[0.636]		[0.614]		[0.821]		[0.791]
Absolute latitude	-34.466***	-18.189	-32.977***	-16.722	-59.087***	-42.438*	-55.697***	-37.695*
	[12.601]	[16.097]	[12.314]	[15.521]	[16.229]	[22.677]	[15.853]	[21.852]
Land area	14.171	-0.289	17.305	2.685	56.718	60.255	55.353	53.914
	[72.189]	[63.334]	[70.547]	[61.046]	[69.143]	[89.279]	[67.540]	[86.033]
Ruggedness	2.435	5.066**	2.377	4.777**	2.926	5.122**	2.829	4.740**
	[2.142]	[2.056]	[2.093]	[1.983]	[2.507]	[2.379]	[2.449]	[2.293]
Mean elevation	-0.952*	-1.298***	-0.910*	-1.218***	-1.053*	-1.283**	-0.987*	-1.166**
	[0.513]	[0.463]	[0.501]	[0.446]	[0.573]	[0.517]	[0.560]	[0.499]
Range of elevation	0.190	-0.060	0.182	-0.054	0.003	-0.312	0.002	-0.272
	[0.133]	[0.172]	[0.130]	[0.166]	[0.145]	[0.206]	[0.141]	[0.198]
Mean land suitability	0.204	-0.580	0.198	-0.520	0.682	0.241	0.641	0.281
	[0.610]	[0.723]	[0.596]	[0.697]	[0.703]	[0.922]	[0.687]	[0.888]
Range of land suitability	1.568***	1.806***	1.508***	1.689***	2.390***	2.662***	2.245***	2.390**
	[0.562]	[0.647]	[0.549]	[0.624]	[0.808]	[0.968]	[0.789]	[0.932]
Distance to nearest waterway	0.573	0.901**	0.557	0.851**	0.886**	1.096**	0.861**	1.029**
	[0.378]	[0.382]	[0.369]	[0.369]	[0.399]	[0.457]	[0.389]	[0.440]
Average executive constraints, lagged		0.168		0.157		0.330***		0.295**
		[0.116]		[0.112]		[0.127]		[0.123]
Fraction of years under democracy, lagged		-0.545		-0.523		-0.823		-0.751
		[0.512]		[0.494]		[0.568]		[0.548]
Fraction of years under autocracy, lagged		-0.384		-0.360		-0.300		-0.279
		[0.350]		[0.338]		[0.569]		[0.549]
Log average oil production per capita, lagged		0.075		0.071		0.073		0.071
		[0.056]		[0.054]		[0.064]		[0.062]
Log average population, lagged		0.149		0.139		0.224		0.202
		[0.133]		[0.128]		[0.216]		[0.208]
Log average GDP per capita, lagged		-0.674***		-0.629***		-0.644***		-0.570***
		[0.216]		[0.208]		[0.227]		[0.219]
Conflict incidence, lagged	3.003***	2.849***	2.873***	2.653***	3.796***	3.635***	3.562***	3.281***
	[0.242]	[0.227]	[0.236]	[0.219]	[0.351]	[0.379]	[0.343]	[0.365]
Marginal effect	3.319***	2.908**	3.543***	3.138**	2.668***	2.343***	3.206***	3.087***
	[1.150]	[1.150]	[1.328]	[1.402]	[0.699]	[0.669]	[1.018]	[1.067]
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5-year period dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	Yes	No	Yes	No	Yes	No	Yes
Colonial history controls	No	Yes	No	Yes	No	Yes	No	Yes
Sample	Old World	Old World	Old World	Old World	Old World	Old World	Old World	Old World
Observations	944	944	944	944	927	927	927	927
Countries	119	119	119	119	117	117	117	117
Time horizon	1960–2008	1960–2008	1960–2008	1960–2008	1960–2005	1960–2005	1960–2005	1960–2005
Time frequency	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly	5-yearly
Pseudo R^2	0.423	0.457	–	–	0.514	0.546	–	–

Notes: This table exploits variations in a quinquennially repeated cross-section of countries from the Old World to establish that the significant positive reduced-form impact of contemporary genetic diversity on the likelihood of observing the incidence of (i) an overall (PRIO25) civil conflict in any given 5-year interval during the 1960–2008 time horizon; and (ii) an ethnic (WCM09) civil conflict in any given 5-year interval during the 1960–2005 time horizon, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict, is robust to estimation under the logit and rare-events logit estimators, rather than the probit estimator. To account for temporal dependence in conflict outcomes, all regressions control for the incidence of conflict in the previous 5-year interval, following Esteban, Mayoral and Ray (2012). Since all regressions are based on the Old-World sample, the set of continent dummies includes indicators for Africa and Asia, with Europe being treated as the omitted category. The set of legal origin dummies includes indicators for British and French legal origins, and the set of time-varying (lagged) colonial history controls includes variables that reflect the fraction of years from the previous 5-year interval that a country served as a colony of the U.K., France, and any other major colonizing power. Given the absence of a rare-events logit estimator that permits instrumentation, the current analysis is unable to employ the strategy of exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated marginal effect of a 1 percentage point increase in genetic diversity is the marginal effect at the mean value of diversity in the cross-section, and it reflects the increase in the likelihood of a conflict incidence in any given 5-year interval, expressed in percentage points. Robust standard errors, clustered at the country level, are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

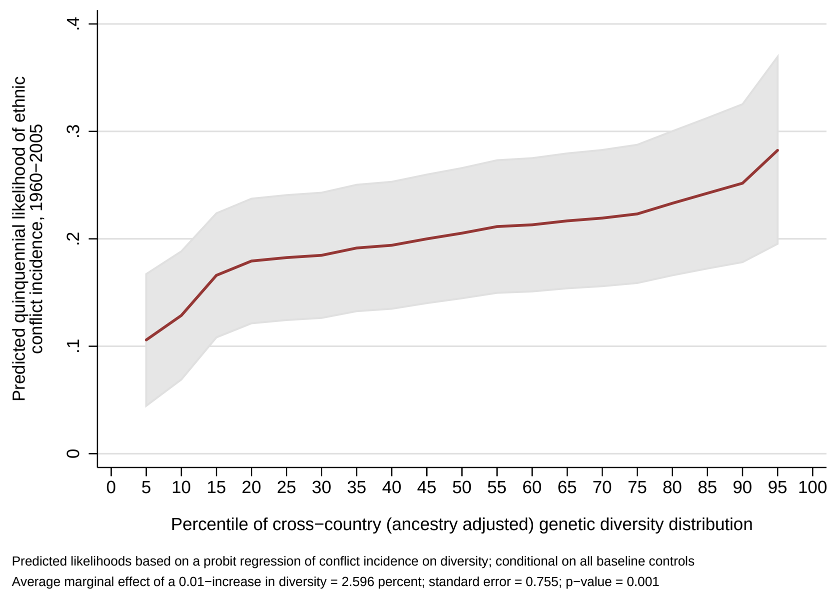
Table A.12: Genetic Diversity and the Onset of Civil Conflict in Annually Repeated Cross-Country Data – Robustness to Employing the Logit and Rare-Events Logit Estimators

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Logit	Logit	RE Logit	RE Logit	Logit	Logit	RE Logit	RE Logit	Logit	Logit	RE Logit	RE Logit
	PRIO25 civil conflict onset [PRIO2]				New PRIO25 civil conflict onset [PRIO-NC]				New WCM09 ethnic civil conflict onset			
Genetic diversity (ancestry adjusted)	20.269***	24.552***	19.592***	23.342***	21.515***	23.446***	20.470***	21.425***	29.656***	33.500***	28.034***	30.012***
	[6.229]	[6.020]	[6.138]	[5.950]	[7.293]	[8.085]	[7.203]	[7.995]	[8.505]	[9.137]	[8.258]	[9.040]
Ethnic fractionalization		-0.191		-0.175		0.488		0.442		1.267		1.214
		[0.570]		[0.565]		[0.656]		[0.649]		[1.160]		[1.118]
Ethnolinguistic polarization		0.722		0.741		0.538		0.585		-0.137		-0.007
		[0.571]		[0.565]		[0.604]		[0.592]		[0.812]		[0.814]
Absolute latitude	-31.974***	-28.490**	-30.808***	-27.394**	-20.601*	-10.446	-19.410*	-9.471	-38.801***	-11.229	-36.383***	-10.095
	[8.600]	[13.332]	[8.487]	[13.141]	[10.858]	[14.420]	[10.766]	[14.289]	[14.048]	[23.630]	[13.913]	[23.211]
Land area	58.467	49.461	59.336	42.477	119.139**	117.641*	114.309**	102.873	36.916	69.194	40.857	53.457
	[51.370]	[60.139]	[50.729]	[59.240]	[58.029]	[63.892]	[57.555]	[63.448]	[65.477]	[89.771]	[64.833]	[89.992]
Ruggedness	2.289	2.751*	2.287	2.815*	1.534	1.655	1.559	1.748	2.858	2.984	2.790	2.988
	[1.747]	[1.594]	[1.731]	[1.583]	[1.814]	[1.877]	[1.799]	[1.843]	[2.352]	[2.498]	[2.332]	[2.454]
Mean elevation	-1.177***	-1.417***	-1.129***	-1.330***	-1.390***	-1.380***	-1.309***	-1.249***	-1.055**	-1.297**	-0.975*	-1.105**
	[0.425]	[0.428]	[0.420]	[0.418]	[0.432]	[0.443]	[0.427]	[0.431]	[0.523]	[0.553]	[0.513]	[0.534]
Range of elevation	0.325***	0.276*	0.309***	0.245*	0.415***	0.268**	0.396***	0.233*	0.196*	0.022	0.180	-0.018
	[0.091]	[0.143]	[0.089]	[0.139]	[0.094]	[0.127]	[0.093]	[0.121]	[0.118]	[0.159]	[0.116]	[0.148]
Mean land suitability	-0.349	-0.401	-0.335	-0.379	1.039**	0.782	1.004**	0.727	0.869	1.032	0.831	0.946
	[0.499]	[0.609]	[0.493]	[0.604]	[0.499]	[0.636]	[0.494]	[0.627]	[0.763]	[0.946]	[0.751]	[0.929]
Range of land suitability	1.647***	1.873***	1.586***	1.738***	0.527	0.284	0.468	0.249	2.261***	2.189	2.094**	1.841
	[0.538]	[0.571]	[0.533]	[0.565]	[0.484]	[0.624]	[0.480]	[0.617]	[0.878]	[1.395]	[0.868]	[1.328]
Distance to nearest waterway	0.200	0.252	0.217	0.285	0.262	0.251	0.285	0.286	0.695*	0.544	0.719*	0.579
	[0.299]	[0.361]	[0.295]	[0.353]	[0.334]	[0.396]	[0.330]	[0.388]	[0.417]	[0.548]	[0.415]	[0.541]
Executive constraints, lagged		0.225***		0.214***		0.202*		0.187		0.287**		0.249*
		[0.081]		[0.081]		[0.117]		[0.116]		[0.145]		[0.144]
Democracy dummy, lagged		-0.755**		-0.732**		-0.802*		-0.765*		-0.920*		-0.825
		[0.350]		[0.354]		[0.434]		[0.439]		[0.521]		[0.539]
Autocracy dummy, lagged		-0.228		-0.264		-0.613*		-0.629*		-0.335		-0.468
		[0.292]		[0.284]		[0.340]		[0.332]		[0.347]		[0.345]
Log oil production per capita, lagged		0.081**		0.084**		0.047		0.053		0.092		0.106*
		[0.040]		[0.040]		[0.051]		[0.050]		[0.062]		[0.063]
Log population, lagged		0.018		0.043		0.093		0.104		0.008		0.059
		[0.117]		[0.117]		[0.096]		[0.097]		[0.252]		[0.244]
Log GDP per capita, lagged		-0.534***		-0.502***		-0.458**		-0.427**		-0.960***		-0.877***
		[0.161]		[0.165]		[0.203]		[0.204]		[0.253]		[0.252]
Conflict incidence, lagged					-0.610	-0.669*	-0.596	-0.513	-0.663*	-0.717*	-0.611	-0.609
					[0.397]	[0.386]	[0.395]	[0.386]	[0.385]	[0.389]	[0.377]	[0.379]
Marginal effect	0.422***	0.453***	0.451***	0.520***	0.290***	0.283***	0.253**	0.255**	0.315***	0.279***	0.369***	0.354***
	[0.125]	[0.105]	[0.155]	[0.143]	[0.095]	[0.097]	[0.110]	[0.125]	[0.094]	[0.078]	[0.139]	[0.129]
Continent dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peace duration cubic splines	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal origin dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Colonial history dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Sample	Old World	Old World	Old World	Old World	Old World	Old World	Old World	Old World	Old World	Old World	Old World	Old World
Observations	4,376	4,354	4,756	4,756	3,849	3,828	4,756	4,756	3,607	3,585	4,331	4,331
Countries	119	119	119	119	119	119	119	119	117	117	117	117
Time horizon	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2008	1960–2005	1960–2005	1960–2005	1960–2005
Time frequency	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual	Annual
Pseudo R^2	0.153	0.173	–	–	0.120	0.142	–	–	0.146	0.182	–	–

Notes: This table exploits variations in an annually repeated cross-section of countries from the Old World to establish that the significant positive reduced-form impact of contemporary genetic diversity on the likelihood of observing the onset of (i) a new or recurring episode of an overall (PRIO2) civil conflict, following two or more years of uninterrupted peace, in any given year during the 1960–2008 time horizon; (ii) a new overall (PRIO-NC) civil conflict in any given year during the 1960–2008 time horizon; and (iii) a new ethnic (WCM09) civil conflict in any given year during the 1960–2005 time horizon, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict, is robust to estimation under the logit and rare-events logit estimators, rather than the probit estimator. To account for duration and temporal dependence in conflict outcomes, all regressions control for a set of cubic splines of the number of peace years, following Beck, Katz and Tucker (1998). In addition, with the exception of regressions explaining PRIO2 onset, for which a mechanical correlation with conflict incidence in the previous year would follow by definition, all regressions control for the lagged incidence of conflict, following Esteban, Mayoral and Ray (2012). Since all regressions are based on the Old-World sample, the set of continent dummies includes indicators for Africa and Asia, with Europe being treated as the omitted category. The set of legal origin dummies includes indicators for British and French legal origins, and the set of time-varying (lagged) colonial history dummies includes indicators for whether a country was a colony of the U.K., France, and any other major colonizing power in the previous year. Given the absence of a rare-events logit estimator that permits instrumentation, the current analysis is unable to employ the strategy of exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary genetic diversity. The estimated marginal effect of a 1 percentage point increase in genetic diversity is the marginal effect at the mean value of diversity in the cross-section, and it reflects the increase in the likelihood of a conflict onset in any given year, expressed in percentage points. Robust standard errors, clustered at the country level, are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.



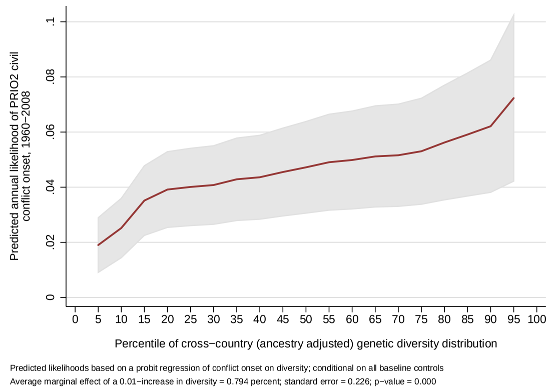
(A) Effect on overall civil conflict incidence



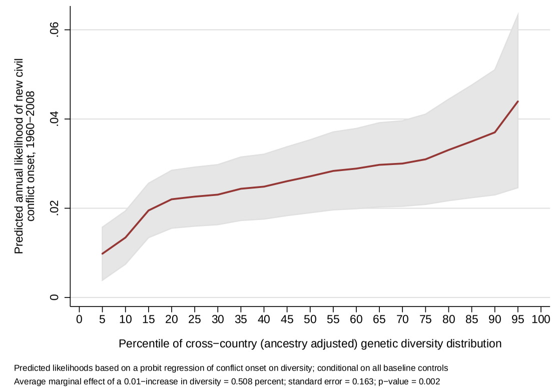
(B) Effect on ethnic civil conflict incidence

FIGURE A.1: The Effect of Genetic Diversity on the Quinquennial Likelihood of Civil Conflict Incidence in the Old-World Sample

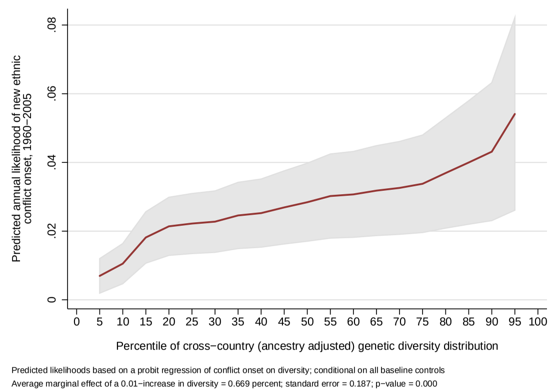
Notes: This figure depicts the influence of contemporary genetic diversity at the country level on the *predicted* likelihood of observing the incidence of (i) an overall (PRIO25) civil conflict in any given 5-year interval during the 1960–2008 time horizon [Panel A]; and (ii) an ethnic (WCM09) civil conflict in any given 5-year interval during the 1960–2005 time horizon [Panel B], conditional on other well-known diversity measures, the proximate geographical, institutional, and development-related correlates of conflict, controls for temporal dependence in conflict outcomes, and continent and 5-year time-interval dummies. In each panel, the predicted likelihood of conflict incidence is illustrated as a function of the percentile of the cross-country genetic diversity distribution, and the prediction is based on the relevant probit regression from Table 5, conducted using the Old-World sample of countries and the full set of covariates considered by the analysis of the conflict outcome in question. The shaded area in each plot reflects the 95-percent confidence-interval region of the depicted relationship.



(A) Effect on PRIO2 overall civil conflict onset



(B) Effect on new overall civil conflict onset



(C) Effect on new ethnic civil conflict onset

FIGURE A.2: The Effect of Genetic Diversity on the Annual Likelihood of Civil Conflict Onset in the Old-World Sample

Notes: This figure depicts the influence of contemporary genetic diversity at the country level on the *predicted* likelihood of observing the onset of (i) a new or recurring episode of an overall (PRIO2) civil conflict, following two or more years of uninterrupted peace, in any given year during the 1960–2008 time horizon [Panel A]; (ii) a new overall (PRIO-NC) civil conflict in any given year during the 1960–2008 time horizon [Panel B]; and (iii) a new ethnic (WCM09) civil conflict in any given year during the 1960–2005 time horizon [Panel C], conditional on other well-known diversity measures, the proximate geographical, institutional, and development-related correlates of conflict, controls for temporal dependence in conflict outcomes, and continent and year dummies. In each panel, the predicted likelihood of conflict onset is illustrated as a function of the percentile of the cross-country genetic diversity distribution, and the prediction is based on the relevant probit regression from Table 8, conducted using the Old-World sample of countries and the full set of covariates considered by the analysis of the conflict outcome in question. The shaded area in each plot reflects the 95-percent confidence-interval region of the depicted relationship.

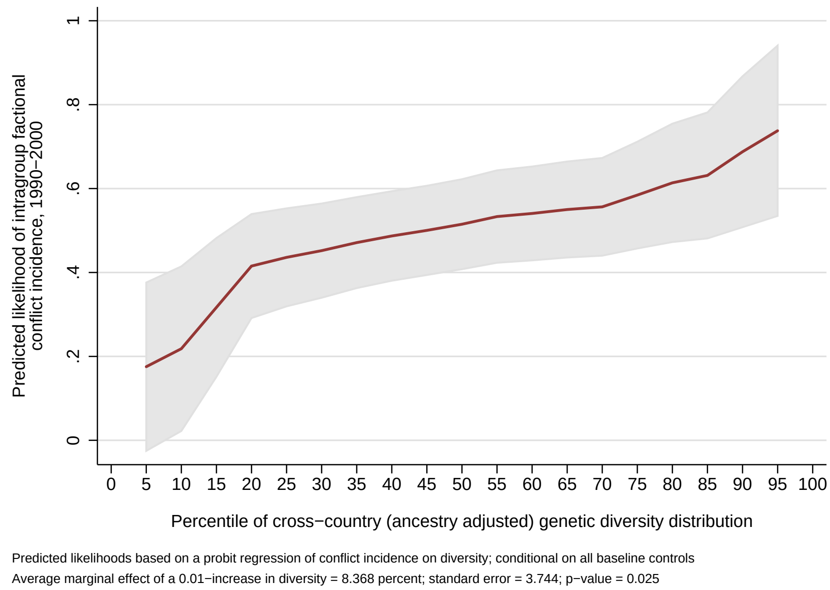
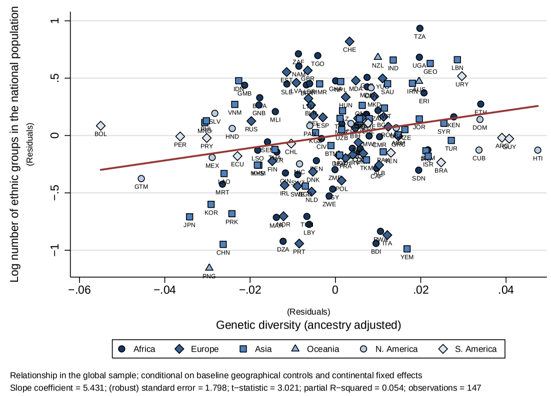
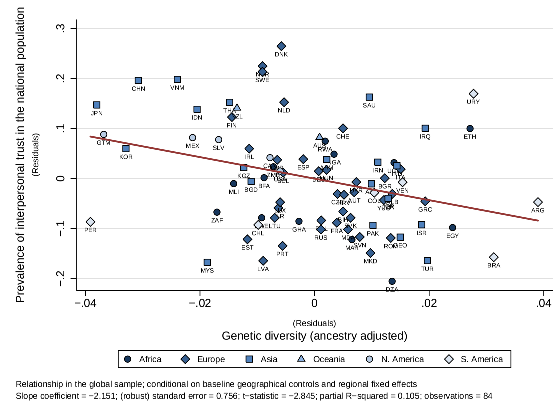


Figure A.3: The Effect of Genetic Diversity on the Likelihood of Intragroup Factional Conflict Incidence in the Old-World Sample

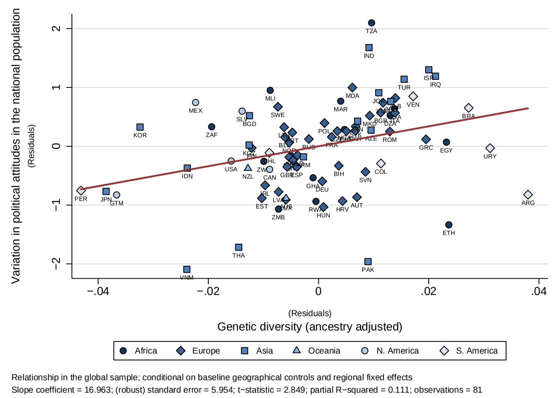
Notes: This figure depicts the influence of contemporary genetic diversity at the country level on the *predicted* likelihood of observing one or more factional conflicts *within* the "minorities at risk" (MAR) groups of a country's population in the 1990–1999 time period, conditional on other well-known diversity measures, the proximate geographical, institutional, and development-related correlates of conflict, and continent dummies. The predicted likelihood of observing one or more intragroup factional conflicts is illustrated as a function of the percentile of the cross-country genetic diversity distribution, and the prediction is based on the relevant probit regression from Table 11, conducted using the Old-World sample of countries and the baseline set of geographical, institutional, and development-related covariates. The shaded area reflects the 95-percent confidence-interval region of the depicted relationship.



(A) Relationship with the number of ethnic groups



(B) Relationship with interpersonal trust



(C) Relationship with variation in political attitudes

FIGURE A.4: Genetic Diversity and the Proximate Determinants of the Frequency of Civil Conflict Onset across Countries

Notes: This figure depicts the global cross-country relationship between contemporary genetic diversity and each of three potential proximate determinants of conflict, namely (i) the number of ethnic groups in the national population [Panel A]; (ii) the prevalence of generalized interpersonal trust at the country level [Panel B]; and (iii) the intracountry dispersion in individual political attitudes on a politically "left"–"right" categorical scale [Panel C], conditional on the baseline geographical correlates of conflict, as considered by the analysis in Table 12. Each of panels A, B, and C presents an added-variable plot with a partial regression line, corresponding to the estimated coefficient associated with genetic diversity in Columns 1, 4, and 7, respectively, of Table 12.

B Descriptive Statistics

Table B.1: Descriptive Statistics of the Regression Sample for Explaining the Frequency of Overall Civil Conflict across Countries

(a) Summary statistics

						Percentile	
		Mean	SD	Min	Max	10th	90th
(1)	Log PRIO25 civil conflict onsets per year, 1960–2008	0.020	0.030	0.000	0.186	0.000	0.054
(2)	Genetic diversity (ancestry adjusted)	0.728	0.027	0.628	0.774	0.688	0.752
(3)	Ethnic fractionalization	0.461	0.259	0.002	0.930	0.107	0.792
(4)	Ethnolinguistic polarization	0.450	0.241	0.000	0.957	0.097	0.734
(5)	Absolute latitude	0.028	0.017	0.001	0.064	0.007	0.051
(6)	Land area	0.001	0.002	0.000	0.016	0.000	0.002
(7)	Ruggedness	0.122	0.117	0.004	0.585	0.018	0.278
(8)	Mean elevation	0.585	0.537	0.001	2.837	0.104	1.232
(9)	Range of elevation	1.685	1.370	0.040	6.176	0.322	3.442
(10)	Mean land suitability	0.388	0.248	0.003	0.951	0.046	0.718
(11)	Range of land suitability	0.717	0.266	0.000	0.999	0.317	0.994
(12)	Distance to nearest waterway	0.359	0.463	0.014	2.386	0.038	1.010
(13)	Ever a U.K. colony dummy	0.245	0.431	0.000	1.000	0.000	1.000
(14)	Ever a French colony dummy	0.196	0.398	0.000	1.000	0.000	1.000
(15)	Ever a non-U.K./non-French colony dummy	0.308	0.463	0.000	1.000	0.000	1.000
(16)	British legal origin dummy	0.238	0.427	0.000	1.000	0.000	1.000
(17)	French legal origin dummy	0.462	0.500	0.000	1.000	0.000	1.000
(18)	Executive constraints, 1960–2008 average	3.939	1.884	1.000	7.000	1.537	7.000
(19)	Fraction of years under democracy, 1960–2008	0.377	0.384	0.000	1.000	0.000	1.000
(20)	Fraction of years under autocracy, 1960–2008	0.393	0.337	0.000	1.000	0.000	0.918
(21)	Log oil production per capita, 1960–2008 average	2.507	2.758	0.000	9.748	0.000	6.501
(22)	Log population, 1960–2008 average	9.163	1.405	5.863	13.841	7.513	10.951
(23)	Log GDP per capita, 1960–2008 average	8.051	1.057	6.210	9.943	6.664	9.596

(b) Pairwise correlations

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
(1)	Log PRIO25 civil conflict onsets per year, 1960–2008	1.000																					
(2)	Genetic diversity (ancestry adjusted)	0.162	1.000																				
(3)	Ethnic fractionalization	0.203	0.198	1.000																			
(4)	Ethnolinguistic polarization	0.048	0.096	0.243	1.000																		
(5)	Absolute latitude	-0.138	0.083	-0.549	0.202	1.000																	
(6)	Land area	0.184	-0.056	-0.003	0.020	0.110	1.000																
(7)	Ruggedness	0.061	-0.201	-0.264	0.021	0.190	-0.115	1.000															
(8)	Mean elevation	0.138	-0.009	0.118	0.133	-0.050	0.054	0.645	1.000														
(9)	Range of elevation	0.293	-0.324	0.102	0.100	-0.028	0.360	0.408	0.604	1.000													
(10)	Mean land suitability	-0.023	-0.234	-0.348	-0.340	0.063	-0.251	0.281	-0.122	-0.142	1.000												
(11)	Range of land suitability	0.195	-0.198	0.001	-0.158	0.161	0.276	0.153	0.196	0.429	0.112	1.000											
(12)	Distance to nearest waterway	0.209	0.191	0.264	0.140	-0.000	0.443	-0.023	0.405	0.335	-0.403	0.129	1.000										
(13)	Ever a U.K. colony dummy	0.050	0.240	0.246	0.182	-0.223	0.099	-0.201	-0.019	-0.018	-0.236	-0.150	0.019	1.000									
(14)	Ever a French colony dummy	0.060	0.144	0.247	-0.055	-0.243	0.086	-0.180	-0.150	-0.097	-0.194	-0.031	0.118	-0.076	1.000								
(15)	Ever a non-U.K./non-French colony dummy	0.123	-0.345	0.173	-0.304	-0.511	0.053	-0.057	0.011	0.210	0.136	-0.026	-0.152	-0.062	-0.062	1.000							
(16)	British legal origin dummy	0.062	0.164	0.186	0.093	-0.195	0.103	-0.125	0.062	0.015	-0.144	0.003	0.003	0.714	-0.151	-0.052	1.000						
(17)	French legal origin dummy	-0.012	-0.124	0.178	-0.131	-0.413	-0.080	-0.148	-0.082	0.022	-0.039	-0.164	-0.084	-0.299	0.321	0.416	-0.517	1.000					
(18)	Executive constraints, 1960–2008 average	-0.101	-0.207	-0.349	-0.009	0.503	0.050	0.170	-0.092	-0.025	0.244	0.277	-0.309	-0.035	-0.281	-0.074	0.123	-0.289	1.000				
(19)	Fraction of years under democracy, 1960–2008	-0.157	-0.212	-0.371	-0.036	0.468	0.042	0.075	-0.192	-0.066	0.248	0.220	-0.328	-0.070	-0.272	-0.045	0.093	-0.203	0.919	1.000			
(20)	Fraction of years under autocracy, 1960–2008	0.048	0.238	0.223	0.076	-0.276	0.015	-0.204	-0.010	-0.021	-0.309	-0.287	0.215	0.079	0.239	-0.048	-0.120	0.151	-0.849	-0.804	1.000		
(21)	Log oil production per capita, 1960–2008 average	-0.004	0.005	0.006	0.182	0.103	0.282	-0.232	-0.175	0.111	-0.386	-0.050	0.077	0.076	0.004	-0.026	-0.085	0.084	-0.132	-0.094	0.238	1.000	
(22)	Log population, 1960–2008 average	0.311	-0.186	-0.120	-0.111	0.059	0.466	0.036	0.081	0.522	0.115	0.586	0.092	0.038	-0.024	0.045	0.080	-0.073	0.105	0.063	-0.063	0.063	1.000
(23)	Log GDP per capita, 1960–2008 average	-0.289	-0.152	-0.485	0.092	0.661	0.117	0.043	-0.271	-0.092	0.004	0.078	-0.298	-0.131	-0.302	-0.171	-0.124	-0.150	0.597	0.599	-0.394	0.437	0.006

Table B.2: Descriptive Statistics of the Regression Sample for Explaining the Frequency of Ethnic Civil Conflict across Countries

(A) Summary statistics

		Mean	SD	Min	Max	Percentile 10th	Percentile 90th
(1)	Log WCM09 ethnic conflict onsets per year, 1960–2005	0.018	0.033	0.000	0.182	0.000	0.063
(2)	Genetic diversity (ancestry adjusted)	0.727	0.027	0.628	0.774	0.688	0.752
(3)	Ethnic fractionalization	0.460	0.259	0.002	0.930	0.107	0.792
(4)	Ethnolinguistic polarization	0.446	0.240	0.000	0.957	0.097	0.733
(5)	Absolute latitude	0.028	0.017	0.001	0.064	0.007	0.051
(6)	Land area	0.001	0.002	0.000	0.016	0.000	0.002
(7)	Ruggedness	0.124	0.118	0.004	0.585	0.019	0.278
(8)	Mean elevation	0.591	0.538	0.001	2.837	0.106	1.232
(9)	Range of elevation	1.706	1.367	0.043	6.176	0.393	3.442
(10)	Mean land suitability	0.393	0.246	0.003	0.951	0.054	0.718
(11)	Range of land suitability	0.726	0.256	0.002	0.999	0.347	0.994
(12)	Distance to nearest waterway	0.363	0.465	0.020	2.386	0.038	1.010
(13)	Ever a U.K. colony dummy	0.241	0.429	0.000	1.000	0.000	1.000
(14)	Ever a French colony dummy	0.199	0.400	0.000	1.000	0.000	1.000
(15)	Ever a non-U.K./non-French colony dummy	0.305	0.462	0.000	1.000	0.000	1.000
(16)	British legal origin dummy	0.241	0.429	0.000	1.000	0.000	1.000
(17)	French legal origin dummy	0.454	0.500	0.000	1.000	0.000	1.000
(18)	Executive constraints, 1960–2005 average	3.903	1.891	1.000	7.000	1.537	7.000
(19)	Fraction of years under democracy, 1960–2005	0.368	0.388	0.000	1.000	0.000	1.000
(20)	Fraction of years under autocracy, 1960–2005	0.403	0.343	0.000	1.000	0.000	0.935
(21)	Log oil production per capita, 1960–2005 average	2.335	2.652	0.000	9.743	0.000	6.063
(22)	Log population, 1960–2005 average	9.179	1.363	6.579	13.822	7.519	10.932
(23)	Log GDP per capita, 1960–2005 average	8.003	1.042	6.180	9.910	6.642	9.556

(B) Pairwise correlations

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
(1)	Log WCM09 ethnic conflict onsets per year, 1960–2005	1.000																					
(2)	Genetic diversity (ancestry adjusted)	0.177	1.000																				
(3)	Ethnic fractionalization	0.215	0.198	1.000																			
(4)	Ethnolinguistic polarization	0.058	0.086	0.236	1.000																		
(5)	Absolute latitude	-0.110	0.093	-0.560	0.217	1.000																	
(6)	Land area	0.068	-0.052	-0.001	0.028	0.106	1.000																
(7)	Ruggedness	0.064	-0.195	-0.261	0.037	0.185	-0.120	1.000															
(8)	Mean elevation	0.135	-0.001	0.125	0.153	-0.059	0.049	0.641	1.000														
(9)	Range of elevation	0.208	-0.317	0.109	0.123	-0.041	0.357	0.401	0.599	1.000													
(10)	Mean land suitability	-0.002	-0.224	-0.347	-0.323	0.049	-0.263	0.269	-0.142	-0.168	1.000												
(11)	Range of land suitability	0.217	-0.184	0.016	-0.119	0.144	0.274	0.131	0.175	0.412	0.067	1.000											
(12)	Distance to nearest waterway	0.163	0.199	0.270	0.155	-0.007	0.441	-0.031	0.400	0.328	-0.424	0.111	1.000										
(13)	Ever a U.K. colony dummy	0.040	0.240	0.235	0.168	-0.232	0.104	-0.195	-0.009	-0.007	-0.228	-0.132	0.026	1.000									
(14)	Ever a French colony dummy	0.038	0.150	0.251	-0.047	-0.251	0.083	-0.187	-0.157	-0.106	-0.207	-0.050	0.114	-0.073	1.000								
(15)	Ever a non-U.K./non-French colony dummy	0.029	-0.356	0.186	-0.311	-0.505	0.056	-0.056	0.014	0.218	0.146	-0.017	-0.152	-0.049	-0.059	1.000							
(16)	British legal origin dummy	-0.001	0.171	0.190	0.104	-0.203	0.100	-0.133	0.056	0.006	-0.158	-0.017	-0.002	0.729	-0.156	-0.049	1.000						
(17)	French legal origin dummy	-0.003	-0.136	0.175	-0.154	-0.407	-0.074	-0.138	-0.070	0.040	-0.018	-0.134	-0.074	-0.314	0.332	0.417	-0.514	1.000					
(18)	Executive constraints, 1960–2005 average	-0.152	-0.193	-0.344	0.031	0.501	0.045	0.148	-0.114	-0.044	0.202	0.238	-0.321	-0.008	-0.295	-0.075	0.121	-0.280	1.000				
(19)	Fraction of years under democracy, 1960–2005	-0.196	-0.196	-0.370	-0.013	0.462	0.041	0.052	-0.210	-0.081	0.213	0.192	-0.329	-0.051	-0.274	-0.046	0.092	-0.194	0.912	1.000			
(20)	Fraction of years under autocracy, 1960–2005	0.076	0.242	0.230	0.037	-0.294	0.020	-0.193	0.014	-0.007	-0.294	-0.258	0.228	0.061	0.259	-0.039	-0.112	0.153	-0.845	-0.801	1.000		
(21)	Log oil production per capita, 1960–2005 average	-0.002	-0.017	-0.020	0.150	0.135	0.301	-0.203	-0.152	0.153	-0.346	0.032	0.083	0.059	0.006	-0.026	-0.067	0.051	-0.067	-0.036	0.173	1.000	
(22)	Log population, 1960–2005 average	0.231	-0.171	-0.122	-0.074	0.044	0.472	0.013	0.053	0.508	0.077	0.555	0.070	0.056	-0.046	0.058	0.060	-0.044	0.073	0.043	-0.036	0.149	1.000
(23)	Log GDP per capita, 1960–2005 average	-0.291	-0.160	-0.497	0.067	0.672	0.124	0.053	-0.264	-0.082	0.021	0.117	-0.293	-0.152	-0.295	-0.163	-0.116	-0.157	0.639	0.625	-0.449	0.440	0.044

Table B.3: List of Countries with at least one New PRIO25 Civil Conflict Onset in the 1960–2008 Time Period

	Total count of new conflict onsets, 1960–2008	Fraction of active-conflict years, 1960–2008	Decile of global genetic diversity distribution
Africa			
Ethiopia	8	0.918	10
Nigeria	4	0.122	9
Congo, Dem. Rep.	4	0.347	10
Niger	3	0.143	7
Morocco	2	0.327	5
Sudan	2	0.755	10
Angola	2	0.882	9
South Africa	2	0.469	6
Guinea	1	0.041	5
Rwanda	1	0.255	10
Zimbabwe	1	0.205	9
Ghana	1	0.061	7
Madagascar	1	0.020	9
Gabon	1	0.020	9
Somalia	1	0.388	10
Mauritania	1	0.082	4
Guinea-Bissau	1	0.057	4
Lesotho	1	0.023	7
Mozambique	1	0.471	8
Togo	1	0.020	8
Cameroon	1	0.061	9
Burundi	1	0.362	10
Sierra Leone	1	0.208	5
Côte d'Ivoire	1	0.061	7
Gambia, The	1	0.023	4
Tunisia	1	0.020	7
Liberia	1	0.143	5
Egypt, Arab Rep.	1	0.122	10
Chad	1	0.694	10
Senegal	1	0.184	4
Mali	1	0.082	6
Burkina Faso	1	0.020	7
Algeria	1	0.383	6
Kenya	1	0.022	10
Uganda	1	0.681	10
Central African Republic	1	0.061	10
Congo, Rep.	1	0.102	9
Asia			
India	10	0.837	3
Georgia	3	0.278	8
Indonesia	3	0.673	1
Pakistan	3	0.265	3
Myanmar	2	0.980	2
Sri Lanka	2	0.490	3
Asia (contd.)			
Iran, Islamic Rep.	2	0.531	7
Turkey	2	0.510	9
Azerbaijan	2	0.333	7
Syrian Arab Republic	1	0.102	9
Nepal	1	0.286	3
Saudi Arabia	1	0.020	8
Oman	1	0.143	5
Uzbekistan	1	0.167	4
Iraq	1	0.796	9
Afghanistan	1	0.612	4
Philippines	1	0.816	1
Israel	1	1.000	8
Bangladesh	1	0.474	2
Malaysia	1	0.163	2
Cambodia	1	0.612	2
Tajikistan	1	0.333	4
Thailand	1	0.306	2
Europe			
Russian Federation	6	0.327	5
Serbia and Montenegro	3	0.061	7
Croatia	1	0.167	6
United Kingdom	1	0.449	4
Moldova	1	0.056	7
France	1	0.041	4
Macedonia, FYR	1	0.056	7
Romania	1	0.020	8
Spain	1	0.204	3
North America			
Panama	1	0.020	1
United States	1	0.143	3
Dominican Republic	1	0.020	5
Nicaragua	1	0.245	1
Haiti	1	0.061	8
El Salvador	1	0.286	1
Mexico	1	0.041	1
South America			
Colombia	1	0.918	1
Chile	1	0.020	1
Venezuela, RB	1	0.061	2
Uruguay	1	0.020	3
Peru	1	0.429	1

Table B.4: List of Countries with at least one New WCM09 Ethnic Civil Conflict Onset in the 1960–2005 Time Period

	Total count of new ethnic conflict onsets, 1960–2005	Fraction of active-conflict years, 1960–2005	Decile of global genetic diversity distribution		Total count of new ethnic conflict onsets, 1960–2005	Fraction of active-conflict years, 1960–2005	Decile of global genetic diversity distribution
Africa				**Asia (contd.)**			
Ethiopia	6	0.674	10	Iraq	3	0.826	9
Chad	5	0.674	10	Georgia	3	0.200	8
Sudan	3	0.717	10	Indonesia	3	0.935	1
Nigeria	3	0.130	9	Pakistan	3	0.261	3
Liberia	3	0.261	5	Syrian Arab Republic	2	0.109	9
Burundi	2	0.364	10	Philippines	2	0.783	1
Rwanda	2	0.250	10	Azerbaijan	2	0.267	7
Niger	2	0.130	7	Sri Lanka	1	0.457	3
Uganda	2	0.477	10	Lao PDR	1	0.043	2
Congo, Dem. Rep.	2	0.109	10	Thailand	1	0.065	2
Togo	2	0.043	8	Bangladesh	1	0.559	2
Côte d'Ivoire	1	0.065	7	Lebanon	1	0.348	9
South Africa	1	0.174	6	Turkey	1	0.478	9
Congo, Rep.	1	0.217	9				
Mozambique	1	0.516	8	**Europe**			
Senegal	1	0.304	4	Russian Federation	3	0.304	5
Zimbabwe	1	0.195	9	Serbia and Montenegro	3	0.065	7
Angola	1	0.452	9	United Kingdom	1	0.609	4
Cameroon	1	0.022	9	Croatia	1	0.286	6
Mali	1	0.109	6	Moldova	1	0.067	7
Central African Republic	1	0.043	10	Spain	1	0.283	3
				Macedonia, FYR	1	0.077	7
Asia							
India	8	0.804	3	**North America**			
Myanmar	5	1.000	2	Nicaragua	1	0.196	1
Iran, Islamic Rep.	4	0.478	7	Guatemala	1	0.674	1
Afghanistan	3	0.217	4	Mexico	1	0.022	1