The Impact of the Prehistoric Out-of-Africa Migration on Cultural Diversity*

Bravo Working Paper # 2023-002

Oded Galor, Marc Klemp, Daniel C. Wainstock

May 18, 2023

Abstract

Abstract: *Evidence suggests that the prehistoric out-of-Africa migration has impacted the degree of intra-population genetic and phenotypic diversity across the globe. This paper provides the first evidence that this migration has shaped cultural diversity. Leveraging a folklore catalogue of 958 oral traditions across the world, we find that ethnic groups further away from East Africa along the migratory routes have lower folkloric diversity. This pattern is consistent with the compression of genetic, phenotypic, and phonemic traits along the out-of-Africa migration routes, setting conditions for the emergence and proliferation of differential cultural diversity and economic development across the world.*

Keywords: Diversity, Culture, Out-of-Africa, Folklore

JEL codes: O10, Z10

*Oded Galor (oded_galor@brown.edu): Department of Economics and Population Studies and Training Center, Brown University, 64 Waterman St., Providence, RI 02912, USA. Marc Klemp (marc.klemp@econ.ku.dk): Department of Economics, University of Copenhagen, Øster Farimagsgade 5, building 26, DK-1353 Copenhagen K, Denmark, and Department of Economics and Population Studies and Training Center, Brown University, 64 Waterman St., Providence, RI 02912, USA. Daniel C. Wainstock (daniel_crisostomo_wainstock@brown.edu): Department of Economics, Brown University, 64 Waterman St., Providence, RI 02912, USA.

The Impact of the Prehistoric Out-of-Africa Migration on Cultural Diversity*

Oded Galor, Marc Klemp, Daniel C. Wainstock

May 18, 2023

Abstract

Abstract: *Evidence suggests that the prehistoric out-of-Africa migration has impacted the degree of intra-population genetic and phenotypic diversity across the globe. This paper provides the first evidence that this migration has shaped cultural diversity. Leveraging a folklore catalogue of 958 oral traditions across the world, we find that ethnic groups further away from Africa along the migratory routes have lower folkloric diversity. This pattern is consistent with the compression of genetic, phenotypic, and phonemic traits along the out-of-Africa migration routes, setting conditions for the emergence and proliferation of differential cultural diversity and economic development across the world.*

Keywords: Diversity, Culture, Out-of-Africa, Folklore

JEL codes: O10, Z10

*Oded Galor (oded_galor@brown.edu): Department of Economics and Population Studies and Training Center, Brown University, 64 Waterman St., Providence, RI 02912, USA. Marc Klemp (marc.klemp@econ.ku.dk): Department of Economics, University of Copenhagen, Øster Farimagsgade 5, building 26, DK-1353 Copenhagen K, Denmark, and Department of Economics and Population Studies and Training Center, Brown University, 64 Waterman St., Providence, RI 02912, USA. Daniel C. Wainstock (daniel_crisostomo_wainstock@brown.edu): Department of Economics, Brown University, 64 Waterman St., Providence, RI 02912, USA.

1 Introduction

The prehistoric *out-of-Africa migration* is one of the most important chapters of human history as it has shaped the initial conditions for the evolution of human settlements across the world. According to the prevailing hypothesis, this migration was largely characterized by a stepwise expansion, where in each step a subgroup of individuals left their ancestral settlement to establish a new colony farther away, carrying only a subset of the traits of their ancestral settlement. The resulting *Serial Founder Effect* has generated a negative effect of migratory distance from Africa on genetic, phenotypic, and arguably phonemic diversity.[1]

It remains uncertain, however, how deep was the imprint of this process on cultural diversity. Resolving this mystery is critical for the understanding of the extent of the impact of this migratory process on intra-population diversity and its potential persistent effect on the process of economic development over the entire course of human history. In particular, establishing an impact of this migratory process on cultural diversity would contribute to the understanding of the cultural channel for the consequences of the *out-of-Africa* on societal outcomes such as interpersonal trust, political preferences, conflict and comparative economic development (Ashraf, Galor, and Klemp 2021), providing a different viewpoint on the impact of ethnolinguistic fractionalization on economic prosperity.[2]

In view of the impact of the *Serial Founder Effect* on the compression of genetic, phenotypic, and possibly phonemic traits along the out-of-Africa migration routes, we hypothesize that this process set the stage for differential emergence and proliferation of cultural diversity across the world, resulting in a negative relationship between cultural diversity and migratory distance from Africa.

In this paper, we study one important dimension of cultural diversity – *folkloric diversity* – as reflected by the variation in the total number of motifs across oral traditions (i.e., ethnic groups or clusters of ethnic groups along linguistic lines). Leveraging worldwide data on 958 oral traditions (Figure 1) and 2,564 motifs from the Berezkin's Folklore and Mythology Catalogue,[3] we investigate whether the degree of folkloric diversity was shaped by this migratory process.

[1]See Atkinson (2011), Betti (2009), Betti et al. (2013), Betti and Manica (2018), Hanihara (2008), Manica et al. (2007), Ramachandran et al. (2005), von Cramon-Taubadel and Lycett (2008).

[2]Alesina and La Ferrara 2005; Alesina et al. 2003; Easterly and Levine 1997.

[3]Berezkin 2015; Michalopoulos and Xue 2021.

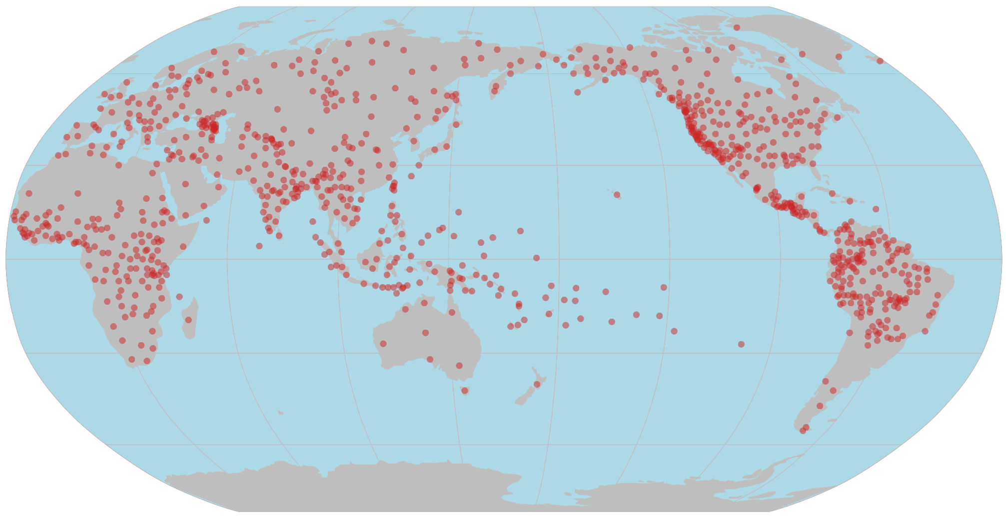

Figure 1. Locations of Berezkin's Oral Traditions.

Notes: The figure depicts the locations of the 958 oral traditions (i.e., ethnic groups or clusters of ethnic groups along linguistic lines) surveyed in Berezkin's Folklore and Mythology Catalogue.

2 Data and Empirical Strategy

The empirical analysis leverages worldwide variation in folkloric motifs across ethnic group, as well as migratory distances from Africa to each of these groups, to test the proposed hypothesis, using data on 2,564 folkloric motifs across 958 oral traditions, as assembled in Berezkin's Folklore and Mythology Catalogue. As depicted in Figure 1, the locations of these oral traditions is widely spread across the globe, permitting the exploration of the impact of the prehistorical dispersal of Homo sapiens across the planet on folkloric diversity.

The catalogue documents folkloric motifs across the globe based on a consistent classification methodology, permitting the assessment of the imprint of the *out-of-Africa migration* on the diversity in the narratives of indigenous societies across the world as a whole.[4] These motifs, often shared by several oral traditions, are the fundamental building blocks of oral traditions, capturing significant events, experiences, and images. Thus, our dependent variable, the total number of motifs in each oral tradition, reflects the degree of folkloric diversity within ethnic groups.

Yet, the distribution of the total number of motifs in each oral tradition is rather skewed as depicted in histograms in Figure 2. Half of the groups have fewer than 62 motifs while the groups in top 1% of the distribution have more than 450 motifs. Hence, in view of

[4]The catalogue is the monumental work of the anthropologist and folklorist Yuri Berezkin who has dedicated his career to compiling and classifying oral traditions for ethnic groups across world regions.

this skewness of the distribution of motifs across oral traditions, the baseline estimation of the impact of the prehistoric migration out-of-Africa on folkloric diversity will be on the logarithmic transformation of the number of motifs.

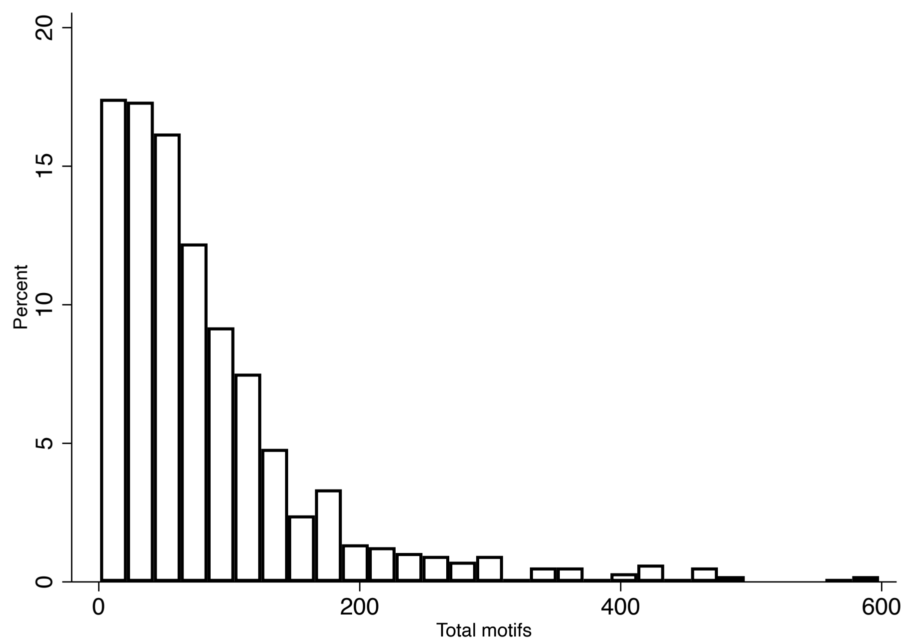

Figure 2. The Distribution of Motifs Across Oral Traditions.

Notes: The figure provides a histogram of the number of motifs per oral tradition (i.e., ethnic groups or clusters of ethnic groups along linguistic lines).

Following the traditional view in the out-of-Africa literature, the independent variable is the shortest migratory distance from East Africa to the interior centroid of each oral tradition. While there is some uncertainty about the precise origin of humans within the African continent (Ragsdale et al., 2023), this location has no impact of folkloric diversity out of Africa, as long as human dispersed to the rest of the world via East Africa. Yet, migratory distance from East Africa has been consistently shown to be a weak predictor of the decline in the level of diversity in phenotypic and genotypic traits within the African continent (e.g., Ramachandran et al., 2005). Nevertheless, no viable alternative has emerged thus far.

In view of the fact that folkloric diversity had not affected the migratory distance out-of-Africa of ethnic groups tens of thousands of years earlier, our empirical strategy is immune from concerns about reverse causality. Yet, to the extent that migratory distance out-of-

Africa could be correlated with other ancestral determinants of cultural diversity, our analysis could be plagued by omitted variable bias.

To mitigate this concern, we account for a range of potentially confounding ancestral geographical characteristics which could have arguably also shaped cultural diversity. In particular: (i) ecological diversity and its influence on trade and ethnolinguistic fragmentation (Michalopoulos 2012, Fenske 2014), (ii) absolute distance from the equator and its adverse effect on biodiversity, (iii) geographical isolation that tend to reduce biodiversity and cultural diversity, and (iv) the suitability of land for agriculture and its impact on population density and the scale of the community (Ashraf and Galor, 2011 and Galor and Ozak, 2016) and therefore on the creation and the diffusion of motifs.

Moreover, in view of the potential observable and unobservable differences in continent-specific factors (e.g., suitability of land for agriculture, disease environment, climatic conditions, as well as colonialism, exploitation, and slavery) which may have affected the observed folkloric diversity, we account for continent-specific fixed effects.

Furthermore, ethnographic characteristics may have impacted the emergence and the diffusion of motifs. Thus the empirical analysis accounts for potential possibly endogenous ethnographic confounders, such as the mean size of local communities, the degree of political centralization, the degree of cultural complexity as measured by the presence of high gods (based on the Ethnographic Atlas), as well as the extent to which ethnic groups are engaged in trade.[5]

Finally, the catalogue may represent an uneven research of oral traditions and motifs across regions and continents. In particular, the coverage is especially rich for the Americas both in terms of oral traditions and motifs.[6] Hence, the statistical model accounts for continental fixed effects, thereby isolating the intra-continental differences and exclusively exploiting the variation that exists *within* continents in estimating the impact of migratory distance from Africa on Folkloric diversity. Moreover, the varying intensity and the timing in which each oral tradition was studied by folklorists could have affected the reported number of motifs for each of these groups. Hence, the estimated effect accounts for the number of publications consulted by Berezkin to construct the universe of motifs for each oral tradition, as well as the year of first publication of these reports.

[5]The first 3 ethnographic characteristics are captured respectively variables *v31*, *v33*, and *v34* in the Ethnographic Atlas, and the engagement in trade is captured by the frequency of the concept *trade* in the set of motifs of each oral tradition.

[6]This may partly reflect the fact that Berezkin began his exploration focusing in the American continent.

3 Empirical Model

Following our hypothesis, we estimate the following linear regression model:

$$M_r = \alpha + \beta D_r + \gamma X_r + \delta Z_r + \zeta_c + \epsilon_r; \qquad r = 1, 2, 3, ...R,$$

where (i) the baseline dependent variable M_r is the log total number of motifs in oral tradition r, r = 1, 2, 3, ... R, (ii) γ and δ are row vectors,(iii) the independent variable D_r is the migratory distance from Africa to oral tradition r, (iv) X_r is a vector of controls for the log number of publications consulted by Berezkin and the log year of their first publication, (v) Z_r is a vector of confounding geographical characteristics which include absolute latitude, mean caloric suitability, standard deviation in caloric suitability and elevation, an island dummy, and log distance to coast, and (vi) ζ_c are continent fixed-effects, The coefficient of interest, β, is hypothesized to be negative.

Moreover, we also employ alternative estimation methods designed to account for count data and dispersion in the dependent variable. In particular, we show the robustness of our results to the use of either linear regression and negative binomial regressions.

4 Main Findings

The raw data indeed suggests that oral traditions further along the migratory routes out-of-Africa have fewer motifs in their set of narratives (Figure 3 (Left Panel)).

The estimated effect of the prehistoric migratory distance from Africa on folkloric diversity is reported in Figure 1, accounting for geographic and ethnographic control variables, and the intensity of folkloric sources. Three estimation methods are employed: log-linear regression (Panel A) - our baseline specification, linear regression (Panel B), and negative binomial regression (Panel C).

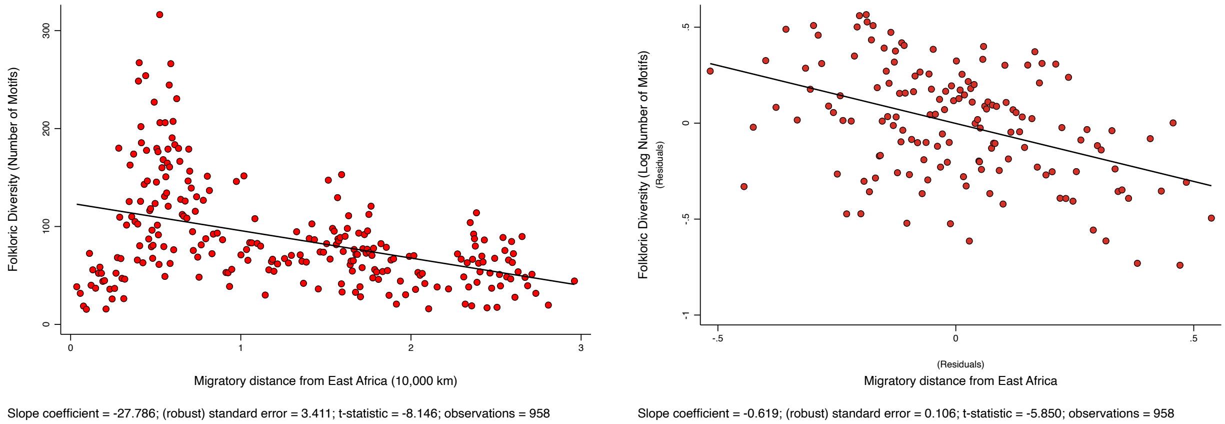

Figure 3. Number of motifs and migratory distance from Africa.

Notes: The figures presents binned scatterplots and the association between the total number of motifs and migratory distance from Africa across oral traditions (i.e., ethnic groups or clusters of ethnic groups along linguistic lines). The left Panel shows the raw association between total number of motifs and migratory distance from Africa with a fitted line estimated using a linear model. The point estimate is statistically significant at the 1% level. The right Panel shows the association between residualized log total of motifs and residualized migratory distance from Africa, accounting for continent fixed-effects, log year of first publication, and log number of publications. The point estimate is statistically significant at the 1% level.

The estimates in Column (1) suggest that indeed, unconditionally, there exists a highly significant negative association between the migration out-of-Africa and folkloric diversity on a global scale.[7] Moreover, the impact of the prehistoric migration out-of-Africa on folkloric diversity have operated within as well as across continents. In view of the potential differences in continent-specific factors that may have affected the observed folkloric diversity, some of which may be unobserved, the confounding effects of geographic and institutional continent-specific characteristics (e.g., suitability of land for agriculture, disease environment, climatic conditions, as well as colonialism, exploitation, and slavery), as well as uneven research of oral traditions and motifs across continents, are accounted for in Column (2) by the inclusion of continent-specific fixed-effects (i.e., the association of an oral tradition with either Africa, Asia, Europe, North America, Oceania, and South America). Hence, we estimate the impact of migratory distance from Africa on folkloric diversity, based on variation in oral traditions within each continent. Interestingly, the impact of the migration out-of-Africa is larger within continents in comparison to its global impact.

[7]Table S.2 rejects the possibility that the impact of the out-of-Africa migration on folkloric diversity is non-monotonic and that there is a sweet spot level of migratory distance from Africa that is conducive for folkloric diversity. The quadratic estimation suggests that if the dependent variable is the total number of motifs, the quadratic term is insignificant statistically, and if the dependent variable in either the log-linear, or negative binomial, number of motifs, the overall effect is negative and concave in the feasible domain, consistent with the main findings of a native impact of migratory distance from Africa on folkloric diversity.

Moreover, since the varying intensity and the timing in which each an oral tradition was studied by folklorists could have affected the reported number of motifs for each of these groups, the number of publications consulted by Berezkin to construct the universe of motifs for each oral tradition, as well as the year of first publication of these reports, is accounted for in Column (3). The scatter plot of this estimated negative association between folkloric diversity and migratory distance out-of-Africa is depicted in (Figure 3 (Right Panel)).[8]

Columns (4)-(8) further establish that accounting for a range of potentially confounding ancestral geographical characteristics which could have arguably also shaped cultural diversity has no qualitative impact on the point estimate and the statistical significance of migratory distance from Africa on folkloric diversity. In light of the fact that distance from the equator is correlated with declining biodiversity, Column (4) accounts the the adverse effect of absolute distance from the equator on biodiversity.

Column (5) accounts for the potential impact of ecological diversity (as captured by the standard deviation of: (i) suitability of land for agriculture and (ii) elevation), on the extent of trade (Fenske 2014), the degree of ethnolinguistic fragmentation (Michalopoulos 2012), and thus folkloric diversity. Column (6) controls for the impact of the suitability of land for agriculture (as captured by the caloric suitability index) on population density (Ashraf and Galor, 2011 and Galor and Ozak, 2016) and thus potentially on the production and the diffusion of motifs.[9]

Moreover, in view of the potential influence of the degree of isolation of an ethnic group on the diffusion of oral narratives across groups we account for determinant of the degree of contact across ethnic groups. Column (7) accounts for this dimension of isolation by including a dummy variable for whether an oral tradition is located in an island, whereas Column (8), account for it by controlling for the distance of oral traditions form to the coast.

Overall, once continental fixed effects are accounted for (in Column (2)), exploiting variations within continents, the estimated effect of migratory distance from Africa is very stable, sizable, and highly significant statistically. In particular, Column 8 in Panel B suggests that a 10,000 km increase in migratory distance from Africa to a given ethnic group (which corresponds to the migratory distance between the Inupiat at the Bering Strait and the Kuna in Colombia) is associated with 56 fewer motifs in comparison with the median number of 62 motifs per ethnic group.

[8]The noisier relationship in closer migratory distances from East Africa reflects the earlier discussion of the uncertainty about the precise origin of humans within the African continent (Ragsdale et al., 2023). In fact, migratory distance from East Africa has been consistently shown to be a weak predictor of the decline in the level of diversity within the African continent (e.g., Ramachandran et al., 2005).

[9]Since caloric suitability is not reported for the Polynesian island of Kapingamarangi , the number of observations in Columns (6)-(8) drops to 957.

Table 1. Baseline cross-oral tradition regressions on the relationship between log total number of motifs and the migratory distance from Africa.

	Total Number of Motifs							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				A. Log-Linear				
Migratory distance from East Africa	−0.19*** (0.043)	−0.73*** (0.15)	−0.62*** (0.11)	−0.50*** (0.10)	−0.53*** (0.11)	−0.57*** (0.11)	−0.54*** (0.11)	−0.54*** (0.11)
Adjusted R^2	0.021	0.19	0.58	0.59	0.59	0.59	0.59	0.59
				B. Linear				
Migratory distance from East Africa	−27.8*** (3.41)	−71.2*** (11.5)	−63.9*** (9.30)	−52.4*** (9.45)	−56.2*** (9.63)	−62.8*** (10.2)	−56.4*** (10.3)	−55.9*** (10.3)
Adjusted R^2	0.061	0.32	0.54	0.56	0.56	0.57	0.57	0.57
				C. Negative Binomial				
Migratory distance from East Africa	−0.35*** (0.037)	−0.81*** (0.13)	−0.64*** (0.098)	−0.54*** (0.097)	−0.58*** (0.099)	−0.60*** (0.10)	−0.54*** (0.10)	−0.54*** (0.10)
Effect Size	−25.6	−52.3	−42.6	−37.3	−39.3	−40.3	−37.3	−37.3
Continent FE	no	yes	yes	yes	yes	yes	yes	yes
Data source controls	no	no	yes	yes	yes	yes	yes	yes
Absolute latitude	no	no	no	yes	yes	yes	yes	yes
Ecological diversity	no	no	no	no	yes	yes	yes	yes
Caloric suitability	no	no	no	no	no	yes	yes	yes
Island	no	no	no	no	no	no	yes	yes
Distance to coast	no	no	no	no	no	no	no	yes
Observations	958	958	958	958	957	957	957	957

parentheses *Notes:* The table establishes a significant negative effect of migratory distance from Africa on folkloric diversity. Panel A presents the results from log-linear regressions, in which the outcome variable is the log total number of motifs. Panel B presents the results from linear regressions. Panel C presents the results from negative binomial regressions. Since caloric suitability is not reported for the Polynesian island of Kapingamarangi, the number of observations in Columns (5)-(8) drops to 957. Data source controls include the log year of first publication and the log number of publications. Heteroskedasticity robust standard errors are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

5 Robustness of the Main Findings

5.1 Accounting for Country Fixed Effects

As reported in Table 1, the estimated effect of the prehistoric migratory distance from Africa on the reduction in folkloric diversity is established based on variation in oral traditions in the world as a whole, but importantly even based on the compression between oral traditions *within* each continent. Importantly, as established in Table 2, the effect operates even within the modern national boundaries in which these ethnic groups and oral traditions currently reside. Accounting for observable characteristics as well as unobservable country fixed-effects, greater migratory distance from Africa has a highly significant negative impact on folkloric diversity across oral traditions within countries.

Table 2. Robustness to Country Fixed-Effects.

	Total Number of Motifs							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Log-Linear							
Migratory distance from East Africa	−0.19*** (0.043)	−0.71*** (0.24)	−0.32* (0.16)	−0.34** (0.17)	−0.34** (0.17)	−0.40** (0.17)	−0.40** (0.17)	−0.40** (0.17)
Adjusted R^2	0.021	0.26	0.60	0.60	0.60	0.61	0.61	0.61
	B. Linear							
Migratory distance from East Africa	−27.8*** (3.41)	−88.2*** (22.3)	−64.4*** (17.8)	−67.4*** (18.0)	−69.7*** (18.3)	−75.5*** (18.8)	−76.7*** (18.8)	−76.8*** (18.7)
Adjusted R^2	0.061	0.46	0.63	0.63	0.63	0.63	0.63	0.64
	C. Negative Binomial							
Migratory distance from East Africa	−0.35*** (0.037)	−0.78*** (0.18)	−0.47*** (0.14)	−0.49*** (0.14)	−0.50*** (0.13)	−0.52*** (0.13)	−0.54*** (0.13)	−0.54*** (0.13)
Effect Size	−25.6	−50.2	−32.5	−33.9	−34.4	−35.8	−36.7	−37.0
Country FE	no	yes	yes	yes	yes	yes	yes	yes
Data source controls	no	no	yes	yes	yes	yes	yes	yes
Absolute latitude	no	no	no	yes	yes	yes	yes	yes
Ecological diversity	no	no	no	no	yes	yes	yes	yes
Caloric suitability	no	no	no	no	no	yes	yes	yes
Island	no	no	no	no	no	no	yes	yes
Distance to coast	no	no	no	no	no	no	no	yes
Observations	958	958	958	958	957	957	957	957

Notes: The table establishes that the impact of migratory distance from Africa on folkloric diversity operates even within the modern national boundaries of these ethnic groups. Since caloric suitability is not reported for the Polynesian island of Kapingamarangi, the number of observations in Columns (5)-(8) drops to 957. Data source controls include the log year of first publication and the log number of publications. Heteroskedasticity robust standard errors are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

5.2 Accounting for Ethnographic Characteristics

Ethnographic characteristics may have impacted the emergence and the diffusion of motifs across the globe. In particular, a range of ethnographic confounders, such as the mean size of local communities, the degree of political centralization, the degree of cultural complexity as measured by the presence of high gods may have affected the evolution of motifs across ethnic groups. Moreover, the extent to which ethnic groups have been engaged in trade could have affected the diffusion of motifs across some groups rather than others.

Table 3. Robustness to Ethnographic Characteristics.

	Log Total Number of Motifs											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Migratory distance from East Africa	−0.53***		−0.53***	−0.49***		−0.50***	−0.51***		−0.52***	−0.54***		−0.43***
	(0.14)		(0.14)	(0.15)		(0.15)	(0.16)		(0.16)	(0.11)		(0.11)
Political centralization		0.054*	0.052									
		(0.032)	(0.032)									
High gods					0.016	0.025						
					(0.034)	(0.033)						
Mean size of local communities								0.044	0.049			
								(0.039)	(0.039)			
Trade intensity											0.20***	0.19***
											(0.033)	(0.033)
Continent FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
All controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	578	578	578	459	459	459	397	397	397	957	957	957
Adjusted R^2	0.61	0.61	0.62	0.62	0.61	0.62	0.64	0.63	0.64	0.59	0.61	0.62

Notes: The table establishes the robustness of the results in Table 1 to the inclusion of control variables for political centralization, belief in high gods, size of local communities, and the engagement with trade. In Columns (1) to (3), we restrict the analysis to the set of oral traditions for which we observe the level of political centralization. In Columns (4) to (6), we restrict the analysis to the set of oral traditions for which we observe the type of high gods. In Columns (7) to (9), we restrict the analysis to the set of oral traditions for which we observe the mean size of local communities. Controls include data source controls (i.e., log year of first publication and log number of publications), and geographical controls (i.e., absolute latitude, ecological diversity, caloric suitability, island, and log distance to coast). Heteroskedasticity robust standard errors are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table 3 established these potentially confounding cultural and institutional characteristics has no qualitative impact on the point estimate and its statistical significance of migratory distance from Africa on folkloric diversity.[10] In particular, Columns (1) to (3), shows that for the set of oral traditions for which we observe the level of political centralization, while

[10]The assignment of these ethnographic characteristics to an oral tradition relies on Michalopoulos and Xue (2021)'s the match between Berezkin's Folklore and Mythology Catalogue and the Ethnographic Atlas. In particular, if a single oral tradition consists of multiple ethnicities, the average value of the ethnographic variable is assigned to the oral tradition.

political centralization is mildly associated with folkloric diversity, it is orthogonal to the impact of migratory distance from Africa.

Columns (4)-(6) shows that despite the evidence that knowledgeable and punitive gods is a fundamental characteristic in the understanding of the evolution of human sociality (Purzycki et al. 2016), it has no association with neither the number of motifs nor with the impact of the migration out-of-Africa of this motifs. Columns (7) to (9) shows that the mean size of local communities is not associated with folkloric diversity, and is orthogonal to the impact of migratory distance from Africa on folkloric diversity. Finally, Columns (10) to (12), establish that the engagement in trade, as is captured by the frequency of the concept *trade* in the set of motifs of each oral tradition, is indeed highly significantly associated with the number of motifs. Nevertheless, the effect of migratory distance from Africa remains highly significant and sizable.[11]

5.3 Assessing the Dominating Role of some Continents

The finding are not driven by a pattern that is present in any single continent. As established in Table S.1, dropping one continent at time has no qualitative impact of the results. The estimated effect remain sizable and highly significant.

Reasuringly, in view of the uncertainty about the precise origins of humanity within the African continent, and in light of the weak explanatory power of migratory distance from Africa for folkloric diversity within the African continent, the removal of the African continent (Column (1) of Table S.1) increases the estimated effect and its statistical significance.

5.4 Spatial Dependence and Selection on Unobservables

The main findings are unaffected qualitatively when accounting for spatial dependence. As established in Table S.3, using the Spatial Error Model method, although the impact of migratory distance from Africa on the total number of motifs is reduced slightly from 0.49 to 0.41 and remains sizable and highly significant statistically. Moreover as established in Table S.4, using Conley (1999)'s method, the estimated effect remain highly significant statistically .

Furthermore, it is very improbable that omitted variables could have affected the qualitative results. In particular, as established in Table S.4 based on Oster (2019)'s methodology,

[11]Based on Michalopoulos and Xue (2021) we use the intensity at which oral traditions mention trade related terms, which were shown by them to be strongly correlated with distance to ancient trade routes.

if unobservables were as correlated with the dependent variable as the observables, the estimated effect of the coefficient on migratory distance from Africa, β^*, is qualitatively similar to the predicted baseline coefficients, under continental fixed-effects and country fixed-effects.

6 Concluding Remarks

This paper provides the first evidence about the impact of migratory distance from Africa on cultural diversity. In particular, it shows that ethnic groups further from Africa along the migratory routes have fewer motifs in their set of narratives. This result is consistent with the compression of genetic, phenotypic, and possibly phonemic traits along the out-of-Africa migration routes, setting the stage for the differential emergence and proliferation of cultural diversity across the world. This finding underscores the importance of the exploration of the long-lasting effects of the migration of humans out-of-Africa on cultural and institutional characteristics, as well as comparative development across countries and ethnic groups, highlighting a unique source of unexplored variation in cultural diversity across regions.

References

[1] Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003): "Fractionalization", Journal of Economic Growth, 8, 155-194.

[2] Alesina, A., and E. La Ferrara (2005): "Ethnic diversity and economic performance", Journal of Economic Literature, 43(3), 762-800.

[3] Ashraf, Q., and Galor, O. (2011). Dynamics and Stagnation in the Malthusian Epoch. American Economic Review, 101(5), 2003-2041.

[4] Ashraf, Q., O. Galor, M. Klemp (2021): "The ancient origins of the wealth of nations", The Handbook of Historical Economics, 675–717.

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

[6] Bates, R. (1983): "Essays on the Political Economy of Rural Africa", University of California Press.

[7] Berezkin, Y. (2015): "Folklore and Mythology Catalogue: Its Lay-out and Potential for Research", The Retrospective Methods Network Newsletter, 10, 58–70.

[8] Betti, L., F. Balloux, W. Amos, T. Hanihara, and A. 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.

[9] Betti, L., N. von Cramon-Taubadel, A. Manica, and S. 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.

[10] Betti, L., and A. Manica (2018): "Human Variation in the Shape of the Birth Canal is Significant and Geographically Structured", Proceedings of the Royal Society B: Biological Sciences, 285(1889), 20181807.

[11] Colella, F., R. Lalive, S. Sakalli, and M. Thoenig (2019): "Inference with Arbitrary Clustering", IZA Discussion Paper n. 12584.

[12] Conley, T. G. (1999). GMM estimation with cross sectional dependence, Journal of econometrics 92(1): 1–45.

[13] Easterly, W., and R. Levine (1997): "Africa's growth tragedy: policies and ethnic divisions", Quarterly Journal of Economics, 112(4), 1203-1250.

[14] Fenske, J. "Ecology, Trade, and States in Pre-Colonial Africa." Journal of the European Economic Association 12.3 (2014): 612-640.

[15] Fick, S., and R. Hijmans (2017): "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas," International Journal of Climatology, 37(12), 4302-4315.

[16] Galor, O., and O. Ozak (2016): "The Agricultural Origins of Time Preference", American Economic Review, 2016, 106(10): 3064–3103.

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

[18] Manica, A., W. Amos, F. Balloux, and T. Hanihara (2007): "The Effect of Ancient Population Bottlenecks on Human Phenotypic Variation", Nature, 448(7151), 346–348.

[19] Michalopoulos, S. (2012): "The Origins of Ethnolinguistic Diversity", American Economic Review, 102(4), 1508–1539.

[20] Michalopoulos, S., and M. M. Xue (2021): "Folklore", Quarterly Journal of Economics, 136(4), 1993–2046.

[21] Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. Journal of Business and Economic Statistics, 37(2), 187-204.

[22] Purzycki, B. et al. (2016): "Moralistic gods, supernatural punishment and the expansion of human sociality", Nature, 530, 327–330.

[23] Ramachandran, S., et al. (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.

[24] Ragsdale, A.P., Weaver, T.D., Atkinson, E.G. et al. A weakly structured stem for human origins in Africa. Nature (2023). https://doi.org/10.1038/s41586-023-06055-y

[25] von Cramon-Taubadel, N., and S. 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.

Appendix

A. Robustness of the Main Findings

A1. Assessing the Dominating Role of some Continents

Table S.1. Robustness to dropping one continent at a time.

	Log Total Number of Motifs					
	Africa	*Asia*	*Europe*	*North America*	*Oceania*	*South America*
	(1)	(2)	(3)	(4)	(5)	(6)
Migratory distance from East Africa	−0.72*** (0.11)	−0.50*** (0.16)	−0.55*** (0.11)	−0.31** (0.15)	−0.51*** (0.12)	−0.48*** (0.12)
Continent FE	yes	yes	yes	yes	yes	yes
All controls	yes	yes	yes	yes	yes	yes
Observations	814	667	903	707	902	792
Adjusted R^2	0.59	0.59	0.55	0.62	0.60	0.61

Notes: The table establishes that the impact of migratory distance from Africa on folkloric diversity is a global phenomenon that is not drive by any single continent. Controls include data source controls (i.e., log year of first publication and log number of publications), and geographical controls (i.e., absolute latitude, ecological diversity, caloric suitability, island, and log distance to coast). Heteroskedasticity robust standard errors are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

A2. Absence of a Quadratic Fit

Table S.2. Robustness to Quadratic Fit.

	Total Number of Motifs					
	Linear		*Log-Linear*		*Negative Binomial*	
	(1)	(2)	(3)	(4)	(5)	(6)
Migratory distance from East Africa	−60.3**	−58.2***	0.12	−0.013	−0.16	−0.080
	(26.0)	(21.2)	(0.33)	(0.24)	(0.27)	(0.20)
Squared migratory distance from East Africa	−4.71	1.00	−0.37***	−0.23***	−0.28***	−0.21***
	(7.26)	(6.01)	(0.11)	(0.084)	(0.091)	(0.071)
Continent FE	yes	yes	yes	yes	yes	yes
All controls	no	yes	no	yes	no	yes
Observations	958	957	958	957	958	957
Adjusted R^2	0.32	0.57	0.20	0.60		

Notes: The table establishes that the impact of migratory distance from Africa on folkloric diversity is not hump-shaped in the feasible domain. The Quadratic estimation suggests that if the dependent variable is the total number of motifs (Columns (1)-(2)), the quadratic term is insignificant statistically, and if the dependent variable is either the log-linear number of motifs (Columns (3)-(4)), or negative binomial (Columns (5)-(6)), the first order term is insignificantly different than zero and the overall effect is negative in the feasible domain. Since caloric suitability is not reported for the Polynesian island of Kapingamarangi, the number of observations in Columns (2),(4), and (6) drops to 957. Controls include data source controls (i.e., log year of first publication and log number of publications), and geographical controls (i.e., absolute latitude, ecological diversity, caloric suitability, island, and log distance to coast). Heteroskedasticity robust standard errors are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

A3. Spatial Dependence and Selection on Unobservables

Table S.3. Robustness to Spatial Dependence.

	Log Total Number of Motifs							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migratory distance	−0.54***	−0.75***	−0.55***	−0.38***	−0.39***	−0.45***	−0.44***	−0.44***
from East Africa	(0.11)	(0.20)	(0.14)	(0.13)	(0.13)	(0.13)	(0.14)	(0.14)
Continent FE	no	yes	yes	yes	yes	yes	yes	yes
Data source controls	no	no	yes	yes	yes	yes	yes	yes
Absolute latitude	no	no	no	yes	yes	yes	yes	yes
Ecological diversity	no	no	no	no	yes	yes	yes	yes
Caloric suitability	no	no	no	no	no	yes	yes	yes
Island	no	no	no	no	no	no	yes	yes
Distance to coast	no	no	no	no	no	no	no	yes
Observations	958	958	958	958	957	957	957	957

Notes: The table establishes that the impact of migratory distance from Africa on folkloric diversity is unaffected by accounting for spatial dependence across oral traditions, using the Spatial Error Model method. Since caloric suitability is not reported for the Polynesian island of Kapingamarangi, the number of observations in Columns (5)-(8) drops to 957. Data source controls include the log year of first publication and the log number of publications. Spatial correlation robust standard errors are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table S.4. Robustness to Conley's Spatial Correlation.

	Log Total Number of Motifs							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migratory distance	−0.19***	−0.73***	−0.62***	−0.50***	−0.53***	−0.57***	−0.54***	−0.54***
from East Africa	(0.069)	(0.19)	(0.13)	(0.11)	(0.11)	(0.11)	(0.12)	(0.12)
Continent FE	no	yes	yes	yes	yes	yes	yes	yes
Data source controls	no	no	yes	yes	yes	yes	yes	yes
Absolute latitude	no	no	no	yes	yes	yes	yes	yes
Ecological diversity	no	no	no	no	yes	yes	yes	yes
Caloric suitability	no	no	no	no	no	yes	yes	yes
Island	no	no	no	no	no	no	yes	yes
Distance to coast	no	no	no	no	no	no	no	yes
Observations	958	958	958	958	957	957	957	957

Notes: The table establishes that the impact of migratory distance from Africa on folkloric diversity remains significant if spatial autocorrelations across oral traditions are accounted for using the Conley's method. Since caloric suitability is not reported for the Polynesian island of Kapingamarangi, the number of observations in Columns (5)-(8) drops to 957. Data source controls include the log year of first publication and the log number of publications. Conley standard errors (500 km cutoff) are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table S.5. Robustness to Selection on Unobservables

	Log Total Number of Motifs			
	(1)	(2)	(3)	(4)
Migratory distance from East Africa	−0.73*** (0.15)	−0.54*** (0.11)	−0.71*** (0.24)	−0.40** (0.17)
Continent FE	yes	yes	no	no
Country FE	no	no	yes	yes
All controls	no	yes	no	yes
Observations	958	957	958	957
Adjusted R^2	0.19	0.59	0.26	0.61
β^*		−0.43		−0.17

Notes: The table establishes that accounting for selection on unobservables, using Oster's method, if selection on unobservables is of equal proportion to selection on observables and the maximum R^2 is equal to 1.3 times the observed R^2, the impact of migratory distance from Africa on folkloric diversity, β^*, is qualitatively similar to the predicted baseline effects under continental fixed-effects (Columns (1)-(2)) and country fixed-effects (Columns (3)-(4)). Since caloric suitability is not reported for the Polynesian island of Kapingamarangi, the number of observations in Columns (2) and (4) drops to 957. Controls include data source controls (i.e., log year of first publication and log number of publications), and geographical controls (i.e., absolute latitude, ecological diversity, caloric suitability, island, and log distance to coast). Robust standard errors are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

B. Variable Definitions, Sources and Summary Statistics

B1. Variable Definition and Sources

B1.1. Dependent Variable - Folkloric Diversity

- **Total number of motifs**: The total number of motifs that are prevalent in a given oral tradition (i.e., ethnic groups or clusters of ethnic groups along linguistic lines). Data Source: Berezkin's Folklore and Mythology Catalogue.

B1.2. Independent Variable - Migratory Distance from Africa

- **Migratory distance from Africa:** The shortest migratory distance from Addis Ababa to the interior centroid of each oral tradition. Data Source: Authors' computation. Data Source: Authors' computation.

Computation method: In estimating the migratory distance from Addis Ababa (East Africa) for each of the oral traditions in the data, the shortest migratory path from Addis Ababa to the centroid of each oral tradition is computed. Considering the limited ability of humans to cross large bodies of water in prehistory, the traversable area included bodies of water within 100km of land mass. Yet, in view of existing evidence, migration from Africa to Europe via Italy or Spain is excluded. For oral traditions that reside in a distance that exceed 100km from the shore, the sailing distance to the nearest coast that minimizes the traversable distance from Addis Ababa was added.

B1.3. Data source Controls

- **Year of first publication**: Year of earliest publication (i.e., books and journal articles) consulted by Berezkin to encode motifs. Data Source: Berezkin (2015), Xue and Michalopoulos (2021).
- **Number of publications**: Number of publications (i.e., books and journal articles) consulted by Berezkin to encode motifs. Data Source: Berezkin (2015), Xue and Michalopoulos (2021).

B1.4. Geographical Controls

- **Absolute latitude**: The absolute value of the latitude of the geodesic centroid of each oral tradition. Data Source: Authors' computation.
- **Caloric Suitability**: The maximum potential caloric yield attainable in each oral tradition, given the set of crops that are suitable for cultivation in the pre-1500 period. Data Source: Galor and Ozak (2016).
- **Distance to Coast**: Distance from the centroid of the oral tradition's to nearest point in the coastline. Data Source: Authors' computation.
- **Ecological Diversity**: Standard deviation of caloric suitability and elevation within the territory of each oral tradition. Data Source: Authors' computation based on Galor and Ozak (2016) and Fick and Hijmans (2017) respectively.
- **Island**: A dummy variable that captures whether the oral tradition is located on an island. Data Source: Authors' assignment.
- **Continental Fixed-Effects**: Dummy variables capturing the location of an oral tradition in either: Africa, Asia, Europe, North America, South America, or Oceania. Data Source: Xue and Michalopoulos (2021).

B1.5. Ethnographic Controls

- **Political centralization**: The number of jurisdictional levels beyond the local community, ranging from 1 for stateless societies, through 2 or 3 for petty and larger paramount chiefdoms or their equivalent, to 4 or 5 for large states. Polities imposed recently by colonial regimes are excluded. Data Source: the Ethnographic Atlas. Matched to Berezkin's Folklore and Mythology Catalogue using Xue and Michalopoulos (2021).

- **High Gods**: The range of beliefs in high gods. A high god is defined as a spiritual being who is believed to have created all reality and to be its ultimate governor, even if his sole act was to create other spirits who, in turn, created or control the natural world. Data Source: the Ethnographic Atlas. Matched to Berezkin's Folklore and Mythology Catalogue using Xue and Michalopoulos (2021).

- **Mean size of local communities**: The average population of local communities. Data Source: the Ethnographic Atlas. Matched to Berezkin's Folklore and Mythology Catalogue using Xue and Michalopoulos (2021).

- **Trade Intensity**: The intensity at which trade related terms are mentioned in oral traditions. Data Source: Xue and Michalopoulos (2021).

B2. Summary Statistics

Table S.6. Summary Statistics

	Mean	SD	Median	Min	Max	N
Total number of motifs	87.23	88.09	62	1	598.0	958
Migratory distance from East Africa	1.26	0.79	1	0	3.0	958
Year of first publication	1908.65	35.91	1,904	1,638	2015.0	958
Number of publications	14.57	13.24	10	1	104.0	958
Absolute latitude	27.53	19.34	24	0	77.5	958
Caloric suitability (mean)	5931.09	3879.64	6,262	0	18952.8	957
Caloric suitability (std)	1183.50	927.44	1,020	0	5011.9	957
Elevation (std)	355.58	345.15	254	0	2148.6	958
Island	0.14	0.35	0	0	1.0	958
Distance to coast (km)	406.52	458.22	205	0	2289.6	958
Political centralization	1.90	1.05	2	1	5.0	579
High gods	2.12	1.17	2	1	4.0	460
Mean size of local communities	3.62	2.25	3	1	8.0	397
Trade intensity	0.01	0.02	0	0	0.5	958

Notes: The table provides for all variables used in the data analysis the mean, the standard deviation (SD), the median, the minimum value (MIN), the maximum value (MAX), and the number of observations (N) .