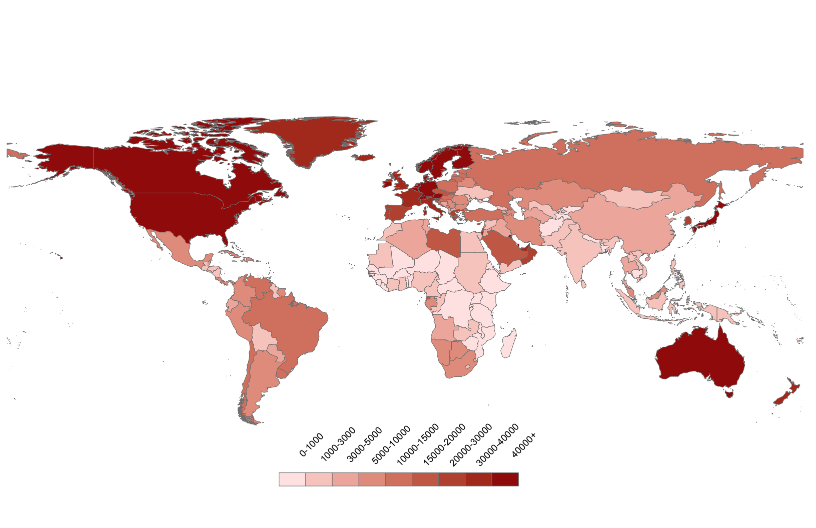
# Genetic Diversity and Comparative Development

Oded Galor

# Income per Capita across the Globe in 2010



# Fundamental Research Questions

- What is the origin of the vast inequality in income per capita across countries and regions?
- What is the impact of deep-rooted factors on the observed patterns of comparative development?
- What fraction of the variation in income per capita across countries could be attributed to the long shadow of history?
- How can policy mitigate the persistent effect of historical factors on comparative development?

# Main Hypothesis

- The migration of humans out of Africa 70,000-90,000 BP affected
  - The distribution of genetic diversity across the globe
  - Comparative economic development
    - Accounts for 16% of the variation in the income per capita across countries

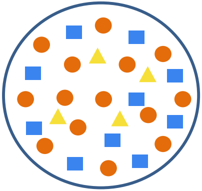
# Main Building Blocks of the Proposed Hypothesis

- The Serial Founder Effect:
  - Lower genetic diversity among indigenous populations at greater migratory distances from East Africa
- Existence of an optimal level of genetic diversity (for each stage of development)
  - Balances between:
    - The negative effect on the cohesiveness of society
    - The positive effect on innovations

# The Serial Founder Effect

- Exodus of modern humans from Africa (70-90K BP)
- Departing populations carry only a subset of the genetic diversity of their parental colonies
    - $\Longrightarrow$ Lower genetic diversity among indigenous populations at greater migratory distances from East Africa

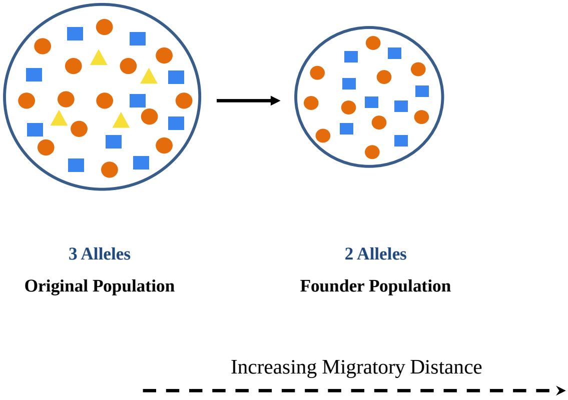
# An Illustration of the Serial Founder Effect



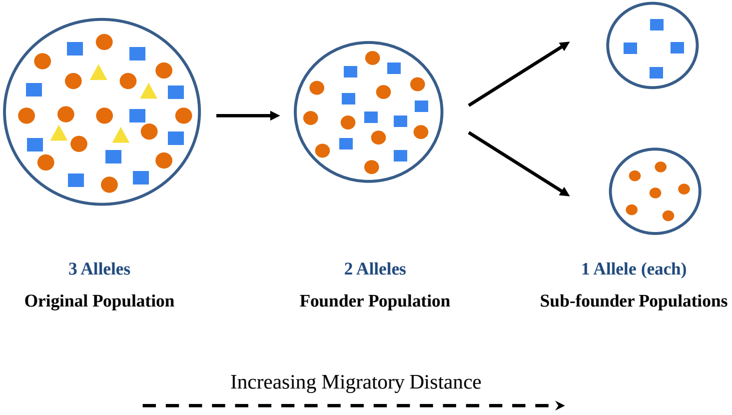
**3 Alleles**

**Original Population**

# An Illustration of the Serial Founder Effect



# An Illustration of the Serial Founder Effect



# Genetic Diversity – Measurement

- Expected Heterozygosity – Index of Genetic Diversity:
    - The probability that two individuals, selected at random from a given population, are genetically different from one another (in a certain spectrum of genes)
    - Measuring Expected Heterozygosity:
        - Gene-specific index capturing its allelic ("gene variant") frequencies in the population
        - Average over a gene-specific index for large spectrum of genes

# The Expected Heterozygosity Index

- $H_\lambda \equiv$ Locus-specific heterozygosity:
  - For a gene $\lambda$ with $k_\lambda$ alleles, where $p_i^\lambda$ is the observed frequency of the $i$-th allele in gene $\lambda$ :

$$H_\lambda = 1 - \sum_{i=1}^{k_\lambda} (p_i^\lambda)^2$$

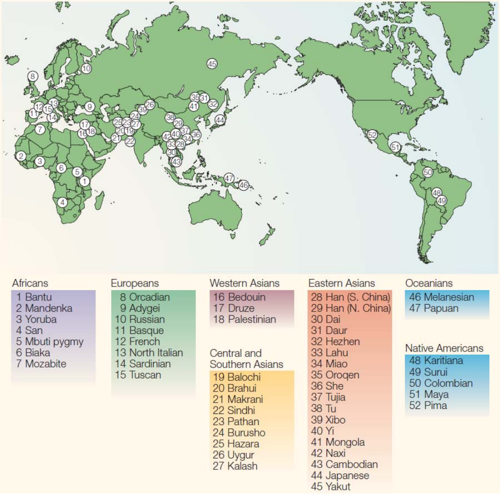
- $H \equiv$ Expected heterozygosity
  - Averaging over $m$ genes:

$$H = \frac{1}{m} \sum_{\lambda=1}^{m} H_\lambda = 1 - \frac{1}{m} \sum_{\lambda=1}^{m} \sum_{i=1}^{k_\lambda} (p_i^\lambda)^2$$

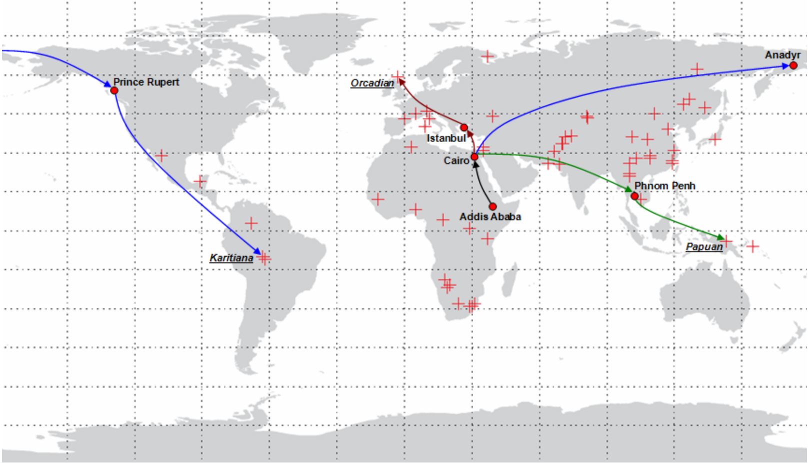
# Data: Human Genome Diversity Project (HGDP)

- Designed to study GD in isolated populations in order to shed light on:
  - The scope of human diversity
  - The journey of humankind from Africa
- Consists of 53 ethnic groups (52 originally)
  - Isolated geographically
  - Resided in the same location for a prolonged period of time
  - Display insignificant genetic admixture

# HGDP Ethnic Groups



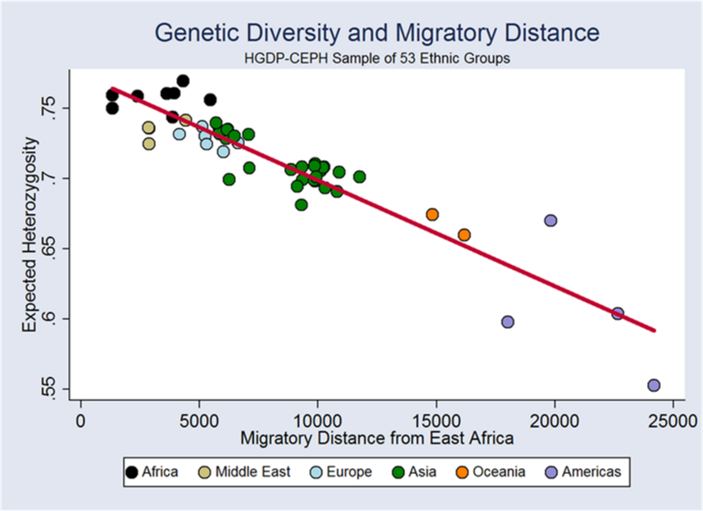
# The Spatial Distribution of the HGDP Ethnic Groups




+ Marks the location of an HGDP ethnic group.

o Marks an approximate critical juncture in the journey of humankind from Africa.

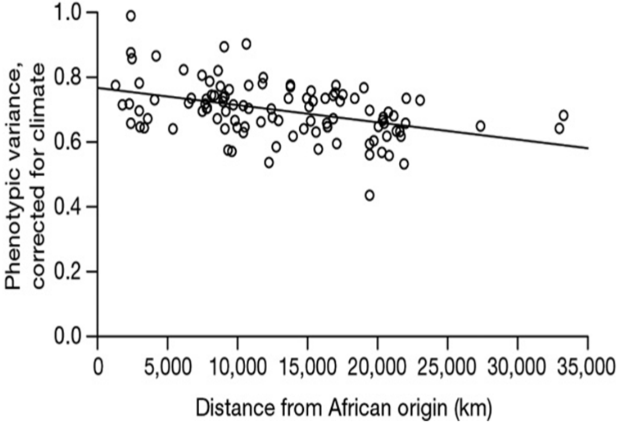
# Migratory Distance from Africa and Genetic Diversity



# Expected Heterozygosity in Microsatellites

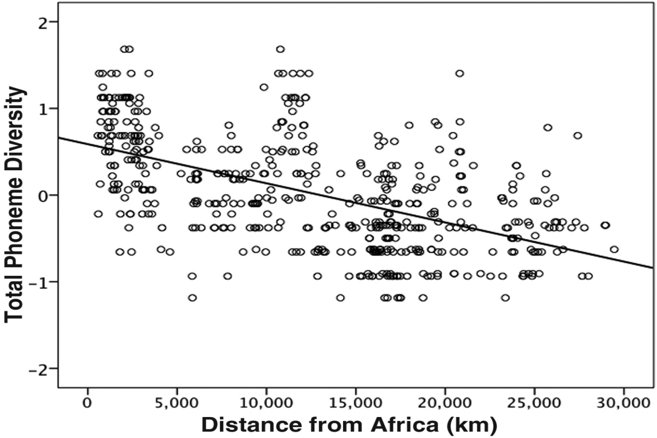
- Expected heterozygosity calculated for the 53 ethnic groups in the HGDP using allelic frequencies for 783 microsatellite loci
- Advantage of using microsatellites – a class of non-protein-coding regions of the human genome:
  - Selectively neutral
    - Ensures that the observed cross-sectional variation in diversity is not due to differential forces of natural selection
  - Mutationally active
    - Facilitates the construction of "population trees" and thus the genealogical and migratory histories of populations
  - Expected heterozygosity in microsatellites is positively correlated with phenotypic and cultural diversity
    - Permitting an exploration of the effect of GD on economic outcomes

# Distance from Africa and Craniometric Diversity



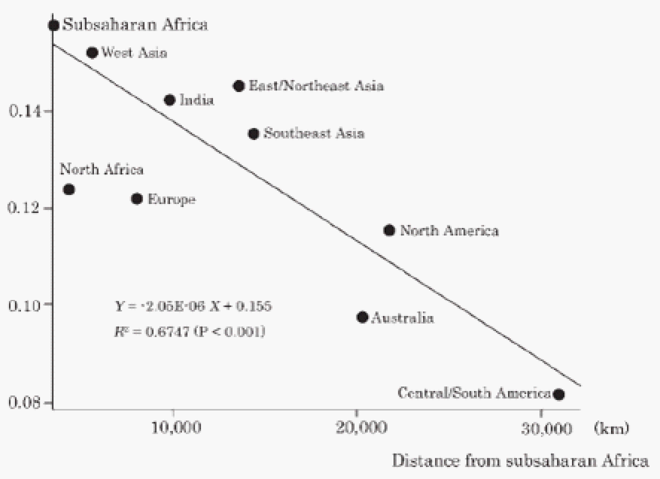

Source: Manica et al. (Nature 2007)

# Distance from Africa and Linguistic Diversity



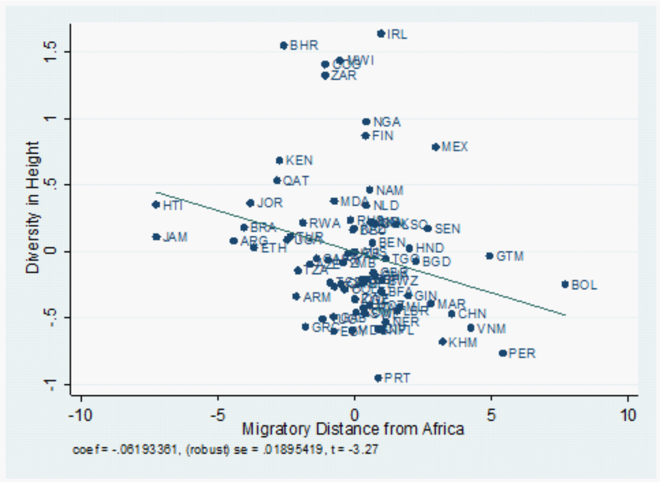

Source: Atkinson (Science 2011)

# Distance from Africa and Dental Diversity



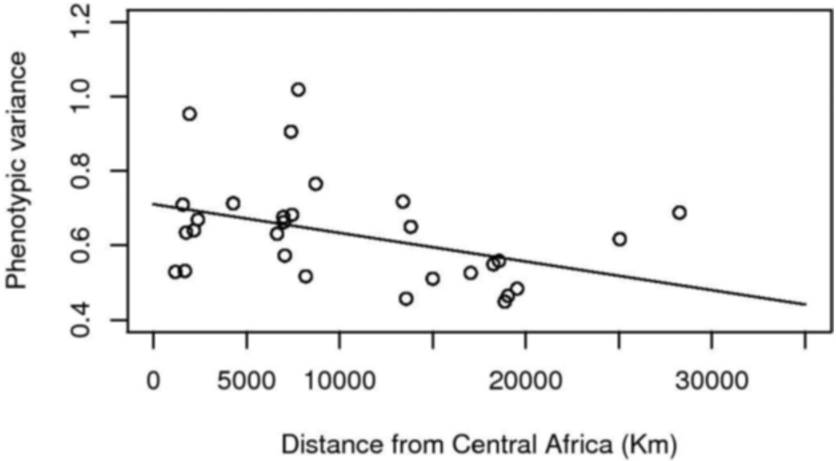

Source: Hanihara (American Journal of Physical Anthropology, 2008)

# Distance from Africa and Height Diversity




Source: Galor and Klemp (2014)

# Distance from Africa and Pelvic Bone Diversity




Source: Betti et al. (Human Biology, 2012)

# Trade-offs: Human Diversity and Productivity

- Existence of an optimal level of genetic diversity (for each stage of development)
  - Balances between:
    - The negative effect on the cohesiveness of society
    - The positive effect on innovations

# The Costs of Diversity

- Genetic diversity increases the incidence of:
  - Mistrust
  - Civil conflicts
  - $\Longrightarrow$ Inefficiency in the operation of the economy relative to its PPF

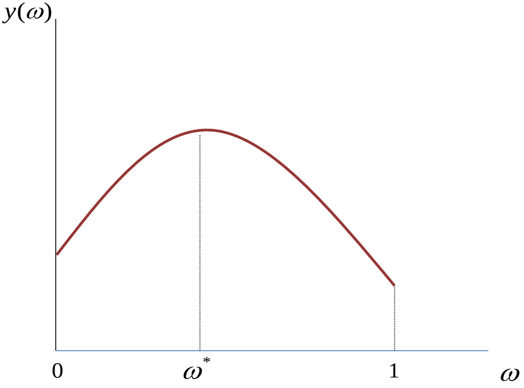
## The Benefits of Diversity

- Genetic diversity
  - Increases the likelihood for the existence of traits that are complementary to the adoption of newly technologies
  - Generates complementary in the production process
  - Increases the upper tail of the ability distribution that matters for innovations
  - $\Longrightarrow$ Diversity fosters innovations & expands the production possibilities

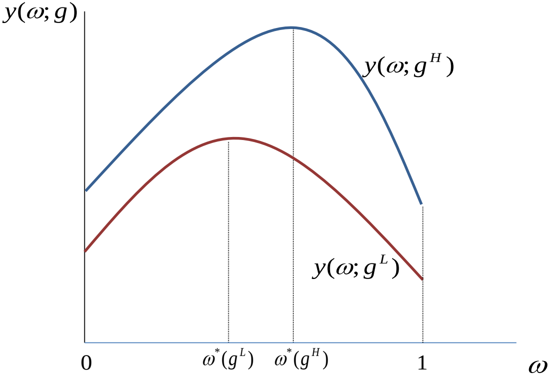
# Optimal Diversity

- Positive but diminishing effects of:
  - Genetic diversity on innovations
  - Homogeneity on cohesiveness
    - $\Longrightarrow$ A hump-shaped relationship between diversity and development
    - $\Longrightarrow$ Optimal level of genetic diversity (for each stage of development)

# The Optimal Level of Genetic Diversity



# The Rise in the Optimal Diversity – Faster Technological Progress



# Empirical Strategy

- Cross-country Analysis
  - Pre-colonial era:
    - Observed genetic diversity (21 countries)
    - Projected diversity (145 countries)
  - Contemporary analysis:
    - Projected diversity (145 countries)
- Across ethnic groups
  - Observed genetic diversity (232 ethnic groups)
  - Projected diversity (1331 ethnic groups)

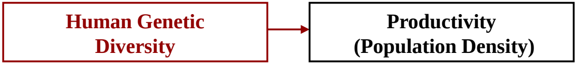
## Comparative Development in the Pre-Colonial Era

- The effect of GD on productivity in the years 1-1500 CE:
    - Productivity is captured by population density (Malthusian Epoch)
    - Disentangle effects of:
        - Genetic Diversity
        - Geographic Factors: Land productivity, Absolute latitude
        - Time elapsed since the Neolithic Revolution

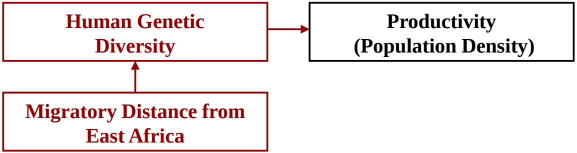
# Determinants of Productivity: Channels

**Productivity**
**(Population Density)**

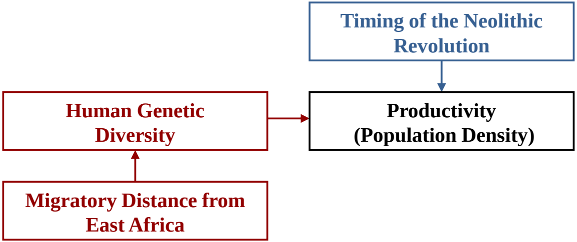
# Determinants of Productivity: Channels



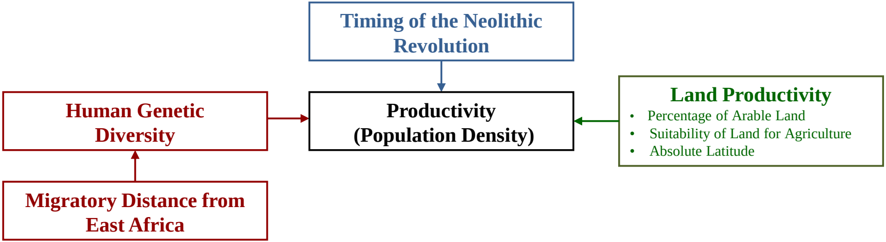
# Determinants of Productivity: Channels



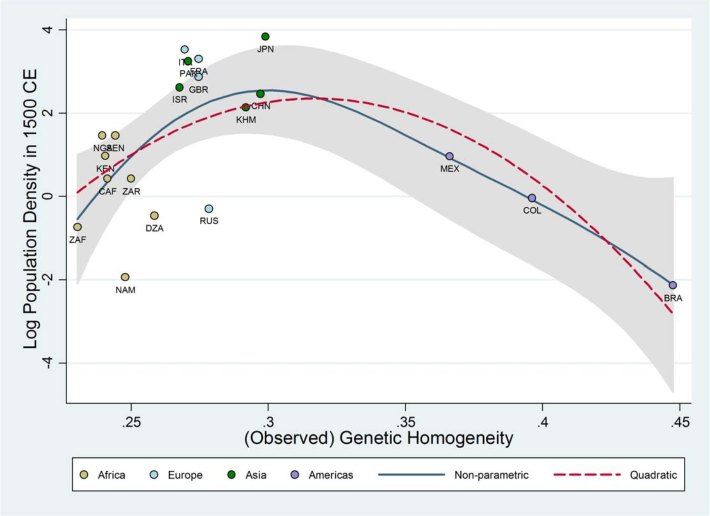
# Determinants of Productivity: Channels



# Determinants of Productivity: Channels



# Observed Diversity and Development in 1500: Unconditional Relationship



# Empirical Model I

- Testing the hypothesis using observed genetic diversity from the HGDP
  - 21-country sample
  - Empirical specification

$$\ln P_{it} = \beta_{0t} + \beta_{1t} G_i + \beta_{2t} G_i^2 + \beta_{3t} \ln T_i + \beta_{4t}' \ln X_i + \beta_{5t}' \ln \Delta_i + \varepsilon_{it}$$

  - $P_{it} \equiv$ population density in country $i$ in year t
  - $G_i \equiv$ *actual* genetic diversity of country $i$
  - $T_i \equiv$ years elapsed since the Neolithic Revolution (NR) for country $i$
  - $X_i \equiv$ vector of land productivity controls for country $i$
  - $\Delta_i \equiv$ vector of continental dummies for country $i$
  - $\varepsilon_{it} \equiv$ a country-year specific error term for country $i$

# Actual Diversity and Comparative Development in 1500

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Dependent Variable: Log Population Density in 1500 | | | | |
| Genetic Diversity | 413.51***<br>(97.32) | | | 225.44***<br>(73.78) | 203.82*<br>(97.64) |
| Genetic Diversity Sqr. | -302.65***<br>(73.34) | | | -161.16**<br>(56.16) | -145.72*<br>(80.41) |
| Log Years since NR | | 2.40***<br>(0.27) | | 1.21***<br>(0.37) | 1.14<br>(0.66) |
| Log % of Arable Land | | | 0.73**<br>(0.28) | 0.52***<br>(0.17) | 0.55*<br>(0.26) |
| Log Absolute Latitude | | | 0.15<br>(0.18) | -0.16<br>(0.13) | -0.13<br>(0.17) |
| Log Agri. Suitability | | | 0.73*<br>(0.38) | 0.57*<br>(0.29) | 0.59<br>(0.33) |
| Optimal Diversity | 0.683<br>(0.008) | | | 0.699<br>(0.015) | 0.699<br>(0.055) |
| Continent Dummies | No | No | No | No | Yes |
| Observations | 21 | 21 | 21 | 21 | 21 |
| R-squared | 0.42 | 0.54 | 0.57 | 0.89 | 0.90 |

Robust standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

# Migratory Distance from East Africa vs. Genetic Diversity

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Dependent Variable: Log Population Density in 1500 | | | | |
| Genetic Diversity | 417.003*** | | | 300.978*** | 361.421** |
| | (90.909) | | | (76.371) | (121.429) |
| Genetic Diversity Sqr. | -306.218*** | | | -241.755*** | -268.515*** |
| | (68.308) | | | (61.099) | (87.342) |
| Migratory Distance | | 0.463*** | | -0.003 | |
| | | (0.142) | | (0.178) | |
| Migratory Distance Sqr. | | -0.021*** | | -0.010 | |
| | | (0.006) | | (0.009) | |
| Mobility Index | | | 0.353** | | 0.051 |
| | | | (0.127) | | (0.154) |
| Mobility Index Sqr. | | | -0.012*** | | -0.003 |
| | | | (0.004) | | (0.006) |
| Observations | 18 | 18 | 18 | 18 | 18 |
| R-squared | 0.43 | 0.30 | 0.30 | 0.47 | 0.43 |
| P-value for: | | | | | |
| Joint Sig. of Diversity and its Sqr. | | | | 0.006 | 0.027 |
| Joint Sig. of Distance and its Sqr. | | | | 0.320 | |
| Joint Sig. of Mobility and its Sqr. | | | | | 0.905 |

Robust standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

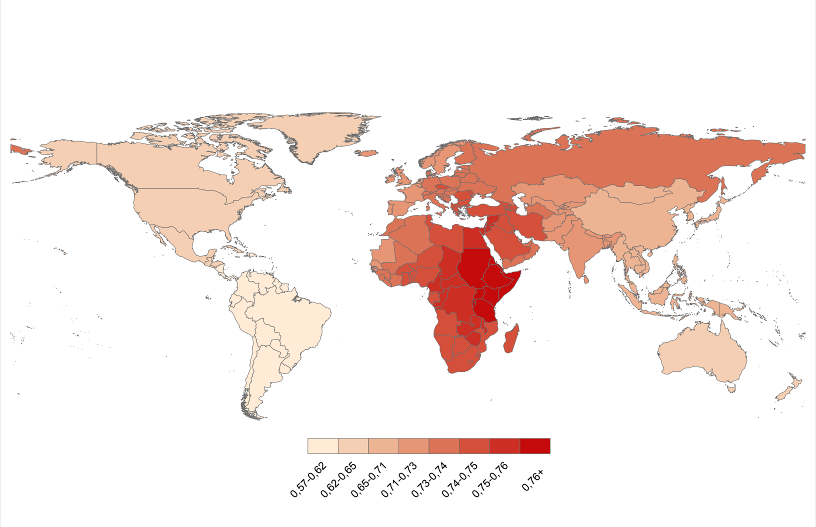
# Empirical Model II

- Testing the hypothesis using projected genetic diversity
  - 145-country sample
- Empirical specification

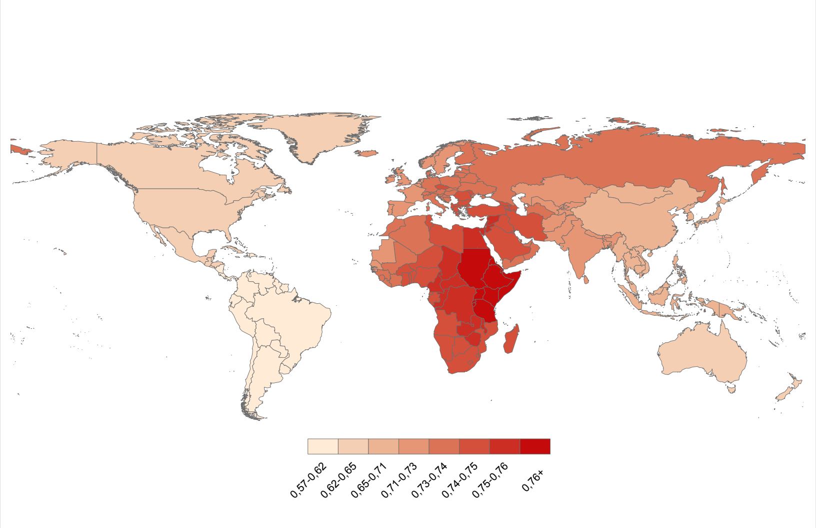
$$\ln P_{it} = \beta_{0t} + \beta_{1t}\hat{G}_i + \beta_{2t}\hat{G}_i^2 + \beta_{3t}\ln T_i + \beta_{4t}'\ln X_i + \beta_{5t}'\ln \Delta_i + \varepsilon_{it}$$

  - $P_{it} \equiv$ population density of country $i$ in year t
  - $\hat{G}_i \equiv$ genetic diversity of country $i$ *projected by migratory distance*
  - $T_i \equiv$ years elapsed since the Neolithic Revolution (NR) for country $i$
  - $X_i \equiv$ vector of land productivity controls for country $i$
  - $\Delta_i \equiv$ vector of continental dummies for country $i$
  - $\varepsilon_{it} \equiv$ a country-year specific error term for country $i$

# Projected Genetic Diversity across Countries in the Pre-Colonial Era



# Genetic Diversity across Countries in the Pre-Colonial Era

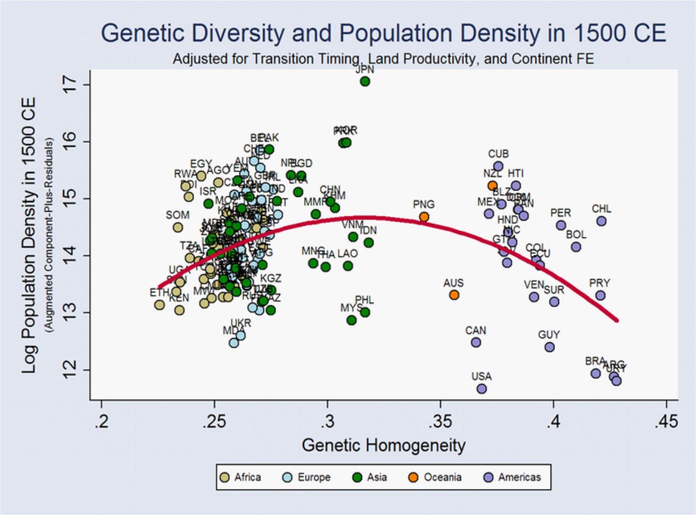


# Predicted Diversity and Comparative Development in 1500

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Dependent Variable is Log Population Density in 1500 | | | | | |
| Predicted Diversity | 250.99*** | | 213.54*** | 203.02*** | 195.42*** | 199.73** |
| | (68.26) | | (63.50) | (61.05) | (56.09) | (80.51) |
| Predicted Diversity Sqr. | -177.40*** | | -152.11*** | -141.98*** | -137.98*** | -146.17*** |
| | (50.22) | | (46.65) | (44.83) | (40.84) | (56.26) |
| Log Years since NR | | 1.29*** | 1.05*** | | 1.16*** | 1.24*** |
| | | (0.18) | (0.19) | | (0.15) | (0.24) |
| Log % of Arable Land | | | | 0.52*** | 0.40*** | 0.39*** |
| | | | | (0.12) | (0.09) | (0.10) |
| Log Absolute Latitude | | | | -0.17* | -0.34*** | -0.42*** |
| | | | | (0.09) | (0.09) | (0.12) |
| Log Agri. Suitability | | | | 0.19 | 0.31*** | 0.26*** |
| | | | | (0.12) | (0.10) | (0.10) |
| Optimal Diversity | 0.707 | | 0.702 | 0.715 | 0.708 | 0.683 |
| | (0.021) | | (0.025) | (0.110) | (0.051) | (0.110) |
| Continent Dummies | No | No | No | No | No | Yes |
| Observations | 145 | 145 | 145 | 145 | 145 | 145 |
| R-squared | 0.22 | 0.26 | 0.38 | 0.50 | 0.67 | 0.69 |

Bootstrap standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

# Predicted Diversity and Comparative Development in 1500



## Interpretations – Diversity and Comparative Development in 1500

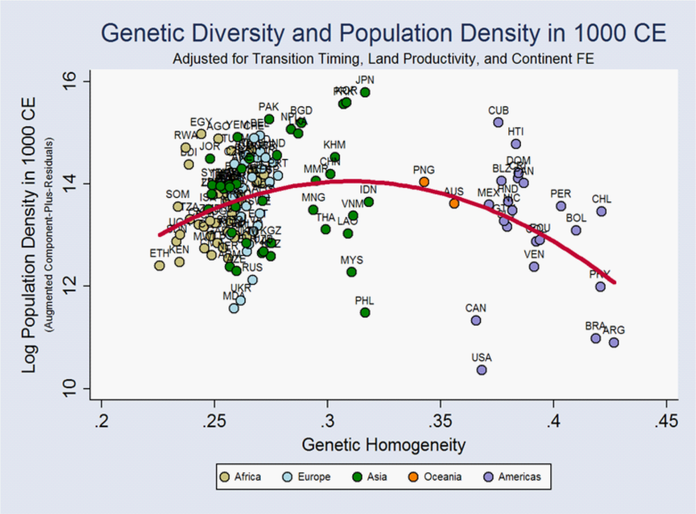
- Optimal GD in 1500 $= 0.6832 \approx$ GD in Japan $= 0.6835$
- Increasing GD of the most homogeneous population in South America by:
  - $0.11 \implies$ 6-fold increase in population density in 1500
  - $0.01 \implies$ 44% increase in population density in 1500
- Decreasing GD of the most heterogeneous population in East Africa by:
  - $0.09 \implies$ 3-fold increase in population density in 1500
  - $0.01 \implies$ 18% increase in population density in 1500
- 0.01 change from the optimal level of GD
  - $\implies$ 1.4% decrease in population density in 1500

# Predicted Diversity and Comparative Development in Earlier Periods

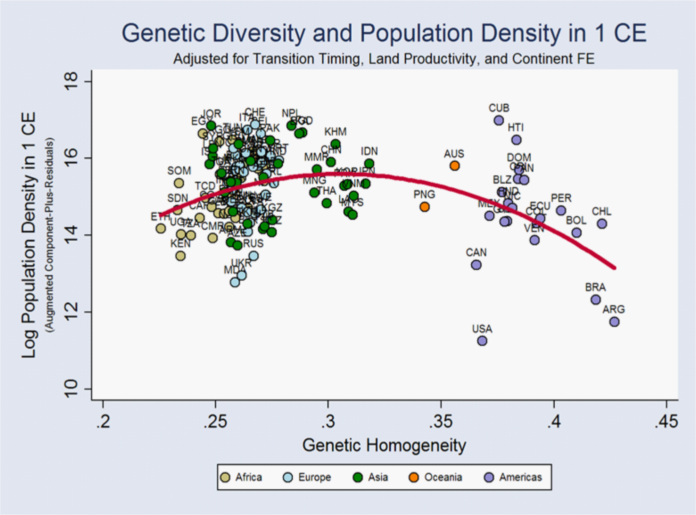
| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Dependent Variable: Log Population Density | | | |
| | 1000 CE | 1000 CE | 1 CE | 1 CE |
| Predicted Diversity | 154.91** | 201.24** | 134.77** | 231.69** |
| | (62.39) | (95.58) | (63.45) | (115.83) |
| Predicted Diversity Sqr. | -109.81** | -145.89** | -96.25** | -166.86** |
| | (45.70) | (66.79) | (46.49) | (81.13) |
| Log Years since NR | 1.37*** | 1.60*** | 1.66*** | 2.13*** |
| | (0.15) | (0.27) | (0.21) | (0.44) |
| Log % of Arable Land | 0.37*** | 0.37*** | 0.31*** | 0.35*** |
| | (0.10) | (0.12) | (0.12) | (0.13) |
| Log Absolute Latitude | -0.38*** | -0.37*** | -0.12 | -0.12 |
| | (0.10) | (0.14) | (0.12) | (0.13) |
| Log Agri. Suitability | 0.21** | 0.19* | 0.24* | 0.21* |
| | (0.10) | (0.11) | (0.12) | (0.12) |
| Optimal Diversity | 0.705 | 0.690 | 0.705 | 0.694 |
| | (0.108) | (0.293) | (0.188) | (0.194) |
| Continent Dummies | No | Yes | No | Yes |
| Observations | 140 | 140 | 126 | 126 |
| R-squared | 0.61 | 0.62 | 0.59 | 0.61 |

Bootstrap standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

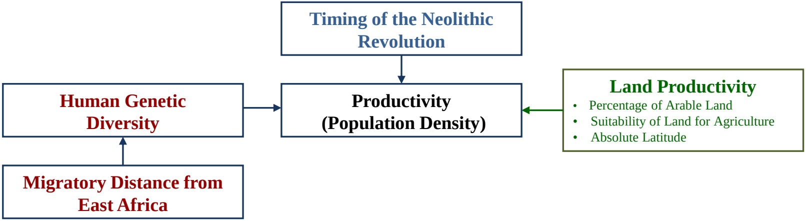
# Predicted Diversity and Comparative Development in 1000 CE



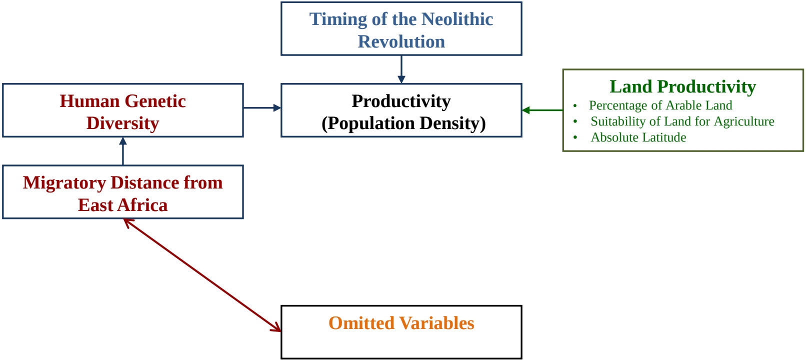
# Predicted Diversity and Comparative Development in 1 CE



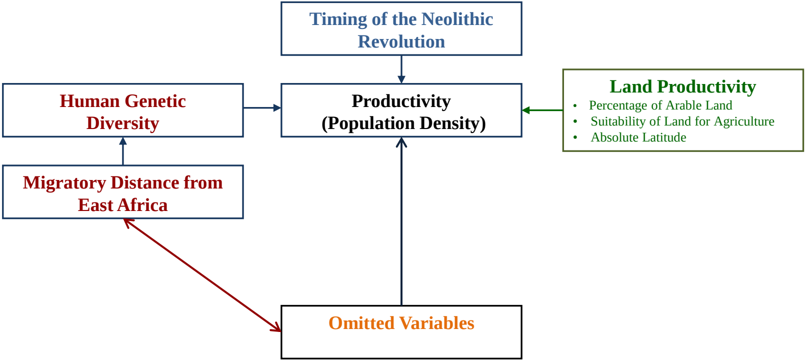
# The Role of Omitted Variables



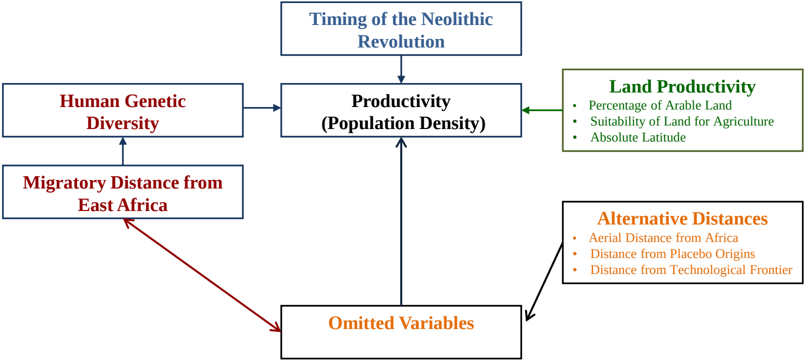
# The Role of Omitted Variables



# The Role of Omitted Variables



# The Role of Omitted Variables – Alternative Distances



# Robustness: Distances from Placebo Origins

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Dependent Variable: Log Population Density in 1500 | | | | |
| Distance calculated from: | Addis Ababa | Addis Ababa | London | Tokyo | Mexico City |
| Migratory Distance | 0.138** | | -0.040 | 0.052 | -0.063 |
| | (0.061) | | (0.063) | (0.145) | (0.099) |
| Migratory Distance Sqr. | -0.008*** | | -0.002 | -0.006 | 0.005 |
| | (0.002) | | (0.002) | (0.007) | (0.004) |
| Aerial Distance | | -0.008 | | | |
| | | (0.106) | | | |
| Aerial Distance Sqr. | | -0.005 | | | |
| | | (0.006) | | | |
| Log Years since NR | 1.160*** | 1.158*** | 1.003*** | 1.047*** | 1.619*** |
| | (0.144) | (0.138) | (0.164) | (0.225) | (0.277) |
| Log % of Arable Land | 0.401*** | 0.488*** | 0.357*** | 0.532*** | 0.493*** |
| | (0.091) | (0.102) | (0.092) | (0.089) | (0.094) |
| Log Absolute Latitude | -0.342*** | -0.263*** | -0.358*** | -0.334*** | -0.239*** |
| | (0.091) | (0.097) | (0.112) | (0.099) | (0.083) |
| Log Agri. Suitability | 0.305*** | 0.254** | 0.344*** | 0.178** | 0.261*** |
| | (0.091) | (0.102) | (0.092) | (0.080) | (0.092) |
| Observations | 145 | 145 | 145 | 145 | 145 |
| R-squared | 0.67 | 0.59 | 0.67 | 0.59 | 0.63 |

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

# Regional Technological Frontiers

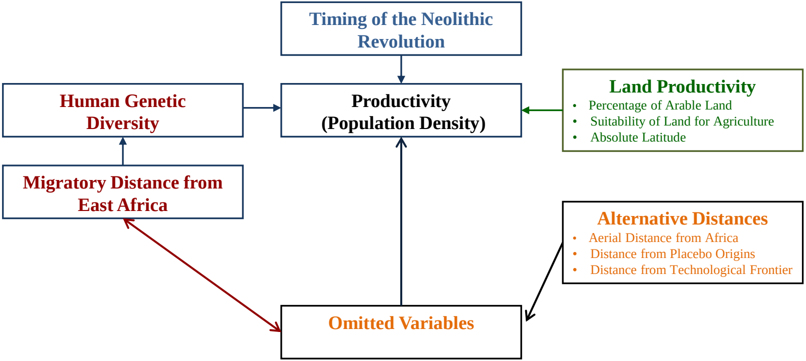
| City & Modern Location | Continent | Sociopolitical Entity | Period |
|---|---|---|---|
| Cairo, Egypt | Africa | Mamluk Sultanate | 1500 CE |
| Fez, Morocco | Africa | Marinid Kingdom of Fez | 1500 CE |
| London, UK | Europe | Tudor Dynasty | 1500 CE |
| Paris, France | Europe | Valois-Orléans Dynasty | 1500 CE |
| Constantinople, Turkey | Asia | Ottoman Empire | 1500 CE |
| Peking, China | Asia | Ming Dynasty | 1500 CE |
| Tenochtitlan, Mexico | Americas | Aztec Civilization | 1500 CE |
| Cuzco, Peru | Americas | Inca Civilization | 1500 CE |
| Cairo, Egypt | Africa | Fatimid Caliphate | 1000 CE |
| Kairwan, Tunisia | Africa | Berber Zirite Dynasty | 1000 CE |
| Constantinople, Turkey | Europe | Byzantine Empire | 1000 CE |
| Cordoba, Spain | Europe | Caliphate of Cordoba | 1000 CE |
| Baghdad, Iraq | Asia | Abbasid Caliphate | 1000 CE |
| Kaifeng, China | Asia | Song Dynasty | 1000 CE |
| Tollan, Mexico | Americas | Classic Maya Civilization | 1000 CE |
| Huari, Peru | Americas | Huari Culture | 1000 CE |
| Alexandria, Egypt | Africa | Roman Empire | 1 CE |
| Carthage, Tunisia | Africa | Roman Empire | 1 CE |
| Athens, Greece | Europe | Roman Empire | 1 CE |
| Rome, Italy | Europe | Roman Empire | 1 CE |
| Luoyang, China | Asia | Han Dynasty | 1 CE |
| Seleucia, Iraq | Asia | Seleucid Dynasty | 1 CE |
| Teotihuacán, Mexico | Americas | Pre-classic Maya Civilization | 1 CE |
| Cahuachi, Peru | Americas | Nazca Culture | 1 CE |

# Robustness to Distance from Regional Technological Frontiers

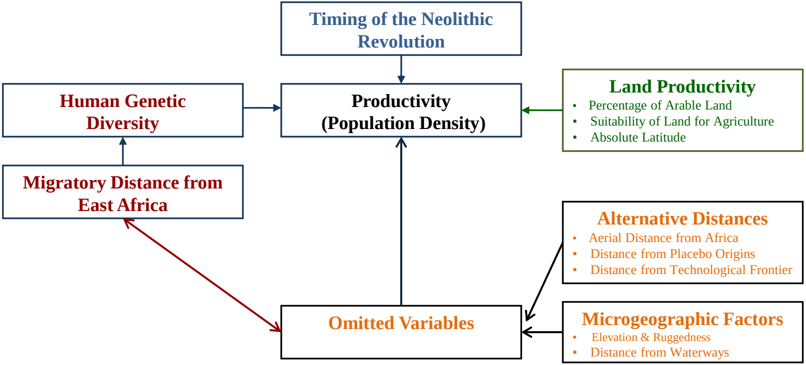
| | (1) | (2) | (3) |
|---|---|---|---|
| | Log Population Density 1500 CE | Log Population Density 1000 CE | Log Population Density 1 CE |
| Predicted Diversity | 156.74** (77.98) | 183.77** (91.20) | 215.86** (106.50) |
| Predicted Diversity Sqr. | -114.63** (54.67) | -134.61** (63.65) | -157.72** (74.82) |
| Log Years since NR | Yes | Yes | Yes |
| Land Prod. Controls | Yes | Yes | Yes |
| Log Distance to Frontier in 1500 CE | -0.19*** (0.07) | | |
| Log Distance to Frontier in 1000 CE | | -0.23** (0.11) | |
| Log Distance to Frontier in 1 CE | | | -0.30*** (0.10) |
| Optimal Diversity | 0.684 (0.169) | 0.683 (0.218) | 0.684 (0.266) |
| Continent Dummies | Yes | Yes | Yes |
| Observations | 145 | 140 | 126 |
| R-squared | 0.72 | 0.64 | 0.66 |

Bootstrap standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

# The Role of Omitted Variables



# The Role of Omitted Variables – Microgeographic Factors

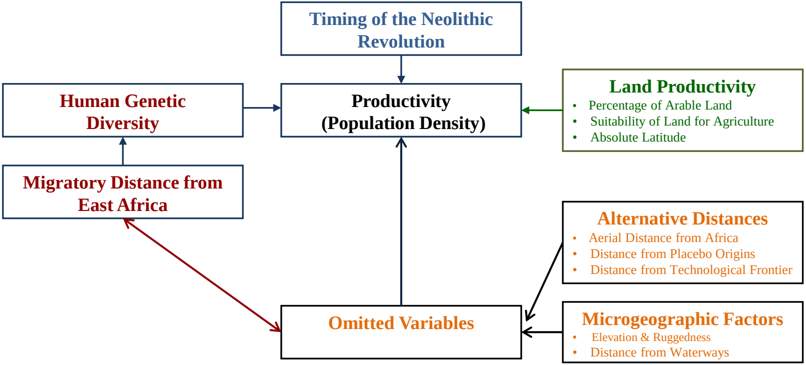


# Robustness to Microgeographic Factors

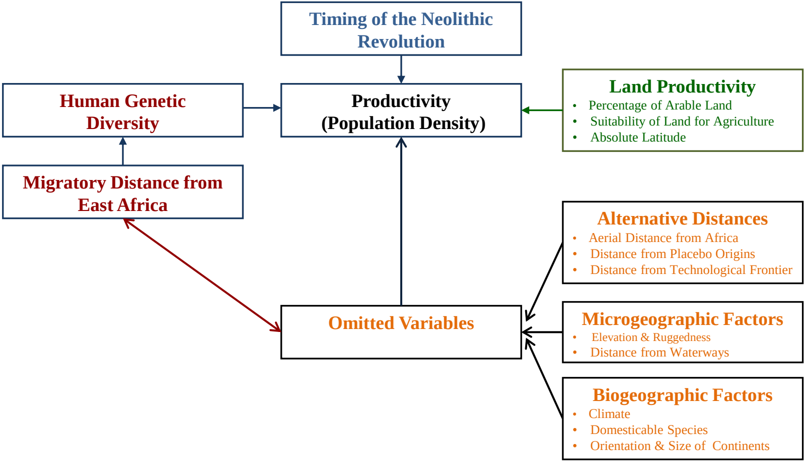
| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Dependent Variable: Log Population Density in 1500 | | | | |
| Predicted Diversity | 159.92*** | 153.20*** | 157.07** | 150.02*** | 157.06** |
| | (56.00) | (53.39) | (78.82) | (49.36) | (68.61) |
| Predicted Diversity Sqr. | -110.39*** | -105.33*** | -112.78** | -102.76*** | -114.99** |
| | (41.08) | (39.11) | (55.48) | (36.23) | (48.26) |
| Log Years since NR | Yes | Yes | Yes | Yes | Yes |
| Land Prod. Controls | Yes | Yes | Yes | Yes | Yes |
| Mean Elevation | -0.48** | | | 0.51* | 0.50* |
| | (0.23) | | | (0.27) | (0.27) |
| Roughness | 5.15*** | | | 3.09* | 4.08** |
| | (1.77) | | | (1.74) | (1.84) |
| Roughness Sqr. | -7.05** | | | -7.05** | -7.63*** |
| | (3.11) | | | (2.96) | (2.91) |
| Distance to Nearest Waterway | | -0.49*** | -0.44** | -0.47** | -0.39** |
| | | (0.18) | (0.18) | (0.18) | (0.18) |
| % Land within 100 km of Waterway | | 0.70** | 0.73** | 1.11*** | 1.18*** |
| | | (0.28) | (0.31) | (0.29) | (0.29) |
| Optimal Diversity | 0.724 | 0.727 | 0.696 | 0.730 | 0.683 |
| | (0.201) | (0.190) | (0.187) | (0.229) | (0.095) |
| Continent Dummies | No | No | Yes | No | Yes |
| Observations | 145 | 145 | 145 | 145 | 145 |
| R-squared | 0.69 | 0.74 | 0.75 | 0.76 | 0.78 |

Bootstrap standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

# The Role of Omitted Variables



# The Role of Omitted Variables – Biogeography



# Robustness to Biogeography

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Dependent Variable is Log Population Density in 1500 | | | | |
| Predicted Diversity | 216.85*** (62.06) | 252.08*** (70.81) | 174.41*** (62.75) | 212.12*** (72.13) | 274.92*** (72.12) |
| Predicted Diversity Sqr. | -154.75*** (45.19) | -180.65*** (51.89) | -125.14*** (45.72) | -151.58*** (52.79) | -197.12*** (52.40) |
| Log Years since NR | 1.30*** (0.16) | | | | 1.16*** (0.31) |
| Land Prod. Controls | Yes | Yes | Yes | Yes | Yes |
| Climate | | 0.62*** (0.14) | | 0.42 (0.27) | 0.37* (0.22) |
| Orientation of Axis | | 0.28 (0.33) | | 0.04 (0.30) | -0.17 (0.27) |
| Size of Continent | | -0.01 (0.02) | | -0.01 (0.01) | -0.01 (0.01) |
| Domesticable Plants | | | 0.02 (0.02) | -0.01 (0.02) | 0.00 (0.02) |
| Domesticable Animals | | | 0.15** (0.06) | 0.12 (0.07) | -0.01 (0.07) |
| Optimal Diversity | 0.701 (0.123) | 0.698 (0.016) | 0.697 (0.159) | 0.700 (0.045) | 0.697 (0.041) |
| Observations | 96 | 96 | 96 | 96 | 96 |
| R-squared | 0.74 | 0.70 | 0.70 | 0.72 | 0.78 |

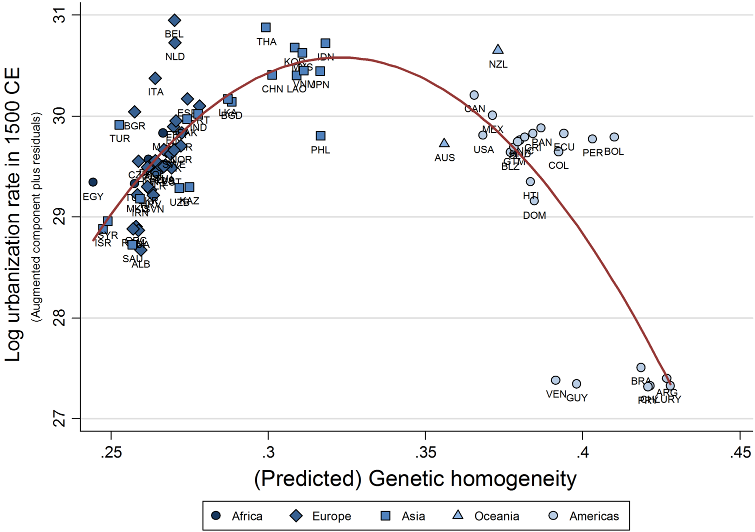
Bootstrap standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

# Robustness to the Use of Urbanization Rates in 1500

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Dependent Variable: Log Urbanization Rate in 1500 | | | | |
| Predicted Diversity | 120.583** (51.618) | 165.167*** (50.088) | 93.467* (48.769) | 148.757*** (48.373) | 234.410*** (67.321) |
| Predicted Diversity Square | -84.760** (38.423) | -120.124*** (37.208) | -62.408* (36.650) | -106.165*** (36.506) | -166.786*** (48.780) |
| Log Years since NR | | 0.457** (0.224) | | 0.402** (0.202) | 0.752*** (0.257) |
| Log % of Arable Land | | | -0.097** (0.043) | -0.116*** (0.044) | -0.119** (0.052) |
| Log Absolute Latitude | | | -0.334** (0.151) | -0.236 (0.155) | -0.151 (0.170) |
| Log Agri. Suitability | | | 0.002 (0.057) | -0.036 (0.058) | 0.031 (0.059) |
| Continent Dummies | No | No | No | No | Yes |
| Observations | 80 | 80 | 80 | 80 | 80 |
| R-squared | 0.30 | 0.35 | 0.40 | 0.44 | 0.51 |

Bootstrap standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

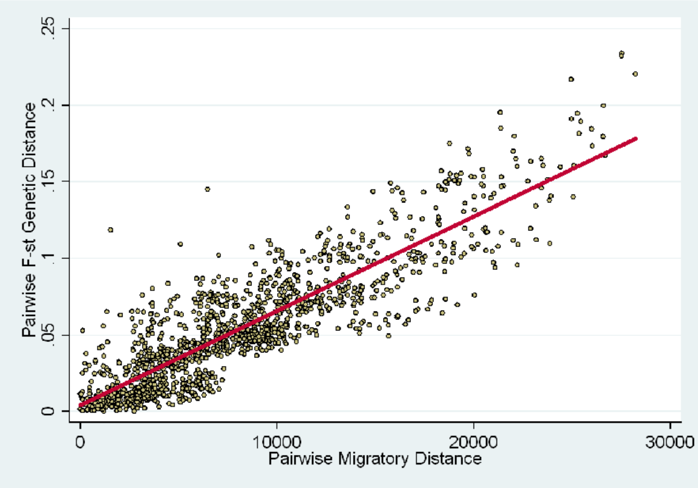
# Genetic Diversity and Urbanization Rates in 1500



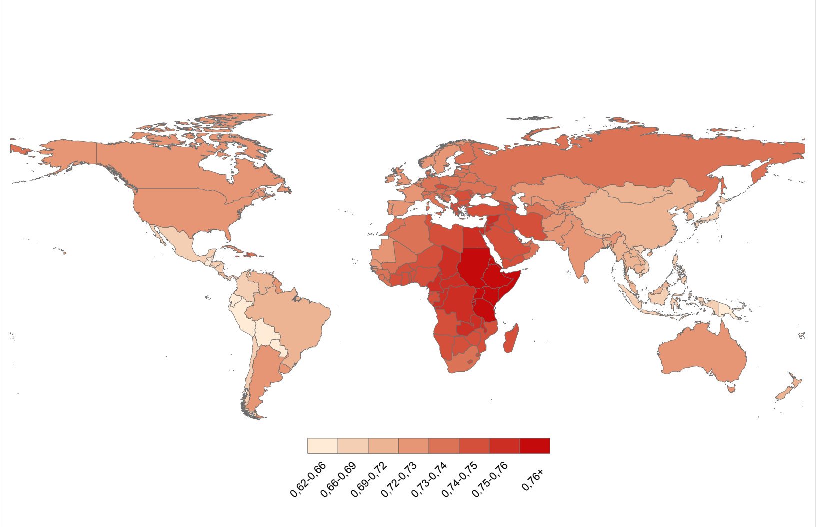
# Measure of Contemporary Genetic Diversity

- The index of contemporary genetic diversity captures:
  - Proportional representation of each ancestral population within a country
  - Genetic diversity among the ancestral populations of each country
    - Projected based on migratory distance of this ancestral population from East Africa
  - Genetic distance between all pairs of ancestral populations of each country
    - Projected based on migratory distance between these ancestral populations

# Projection of Pairwise Genetic Distances



# Genetic Diversity across Countries in 2000



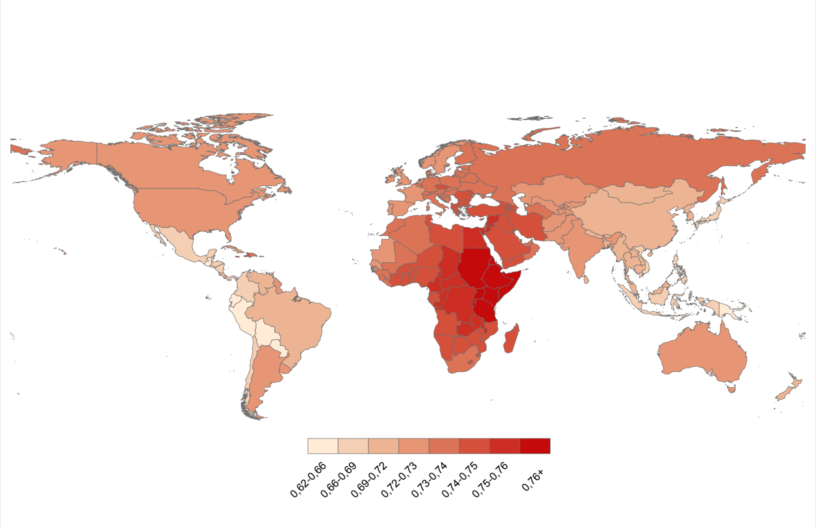
# Empirical Model III

- Testing the hypothesis using contemporary genetic diversity
  - 145-country sample
- Empirical specification

$$\ln y_i = \gamma_0 + \gamma_1 \hat{G}_i + \gamma_2 \hat{G}_i^2 + \gamma_3 \ln T_i + \gamma_4' \ln X_i + \gamma_5' \ln \Lambda_i + \gamma_6 \ln \Gamma_i + \eta_i$$

  - $y_i \equiv$ income per capita of country $i$ in the year 2000
  - $\hat{G}_i \equiv$ index of contemporary genetic diversity of country $i$
  - $T_i \equiv$ years elapsed since the Neolithic Revolution (NR) for country $i$
  - $X_i \equiv$ vector of land productivity controls for country $i$
  - $\Lambda_i \equiv$ vector of institutional and cultural controls for country $i$
  - $\Gamma_i \equiv$ vector of additional geographical controls for country $i$
  - $\eta_i \equiv$ error term for country $i$

# Genetic Diversity across Countries in 2000



# Genetic Diversity and Economic Development in 2000 and 1500

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Log Income per Capita in 2000 | | | Log Population Density in 1500 |
| Adjusted Diversity | 204.610** | 237.238*** | 244.960*** | |
| | (88.466) | (86.278) | (85.454) | |
| Adjusted Diversity Sqr. | -143.437** | -166.507*** | -171.364*** | |
| | (62.545) | (61.363) | (60.843) | |
| Unadjusted. Diversity | | | | 198.587** |
| | | | | (79.110) |
| Unadjusted. Diversity Sqr. | | | | -145.320*** |
| | | | | (55.472) |
| Log Adj. Years since NR | | 0.061 | 0.002 | |
| | | (0.262) | (0.305) | |
| Log Years since NR | -0.151 | | | 1.238*** |
| | (0.186) | | | (0.230) |
| Log % of Arable Land | -0.110 | -0.119 | -0.137 | 0.378*** |
| | (0.100) | (0.107) | (0.111) | (0.100) |
| Log Absolute Latitude | 0.164 | 0.172 | 0.192 | -0.423*** |
| | (0.125) | (0.119) | (0.143) | (0.124) |
| Log Agri. Suitability | -0.193** | -0.177* | -0.189* | 0.264*** |
| | (0.095) | (0.102) | (0.102) | (0.096) |
| Log Population Density in 1500 | | | 0.047 | |
| | | | (0.097) | |
| Optimal Diversity | 0.713 | 0.712 | 0.715 | 0.683 |
| | (0.100) | (0.036) | (0.118) | (0.095) |
| Continent Dummies | Yes | Yes | Yes | Yes |
| Observations | 143 | 143 | 143 | 143 |
| R-squared | 0.57 | 0.57 | 0.57 | 0.68 |

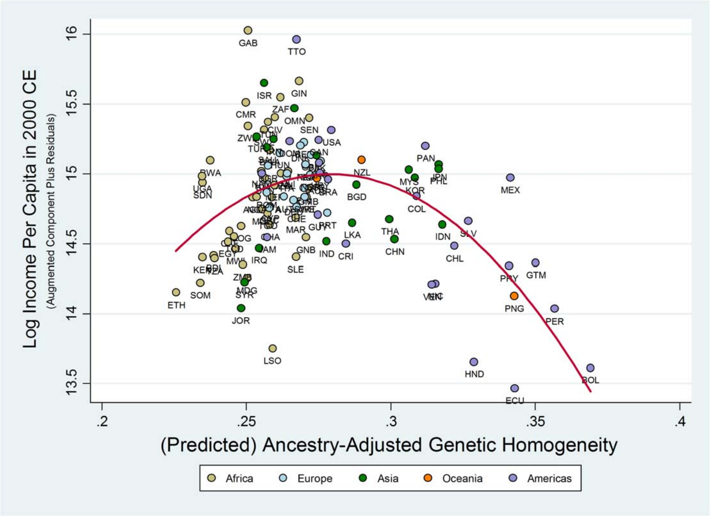
Bootstrap standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

# Genetic Diversity and Comparative Development in 2000

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Dependent Variable: Log Income per Capita in 2000 | | | | |
| Adjusted Diversity | 315.282*** | 225.858*** | 204.102*** | 277.342*** | 215.675*** |
| | (84.215) | (67.669) | (66.984) | (70.232) | (63.954) |
| Adjusted Diversity Sqr. | -220.980*** | -155.826*** | -140.850*** | -192.386*** | -150.871*** |
| | (59.562) | (47.962) | (47.393) | (49.675) | (45.554) |
| Log Adj. Time from NR | -0.273 | -0.092 | -0.062 | 0.396* | -0.046 |
| | (0.269) | (0.200) | (0.203) | (0.233) | (0.208) |
| Log % of Arable Land | -0.218*** | -0.159*** | -0.163*** | -0.183*** | -0.084 |
| | (0.061) | (0.049) | (0.050) | (0.051) | (0.056) |
| Log Absolute Latitude | 0.123 | 0.083 | 0.080 | 0.009 | -0.006 |
| | (0.122) | (0.100) | (0.101) | (0.108) | (0.087) |
| Social Infrastructure | | 2.359*** | 2.069*** | 1.826*** | 0.880** |
| | | (0.269) | (0.377) | (0.417) | (0.418) |
| Democracy | | | 0.036 | | |
| | | | (0.029) | | |
| Ethnic Fractionalization | | | | -0.333 | -0.122 |
| | | | | (0.280) | (0.265) |
| % Population at Risk of Contracting Malaria | | | | -0.502 | -0.723** |
| | | | | (0.351) | (0.353) |
| Avg. Schooling | | | | | 0.134*** |
| | | | | | (0.042) |
| Optimal Diversity | 0.713 | 0.725 | 0.725 | 0.721 | 0.715 |
| | (0.014) | (0.032) | (0.045) | (0.008) | (0.073) |
| Continent Dummies | Yes | Yes | Yes | Yes | Yes |
| Legal Origin Dummies | No | No | No | Yes | Yes |
| Major Religion Shares | No | No | No | Yes | Yes |
| Observations | 109 | 109 | 109 | 109 | 94 |
| R-squared | 0.74 | 0.84 | 0.85 | 0.90 | 0.93 |

Bootstrap standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

# Genetic Diversity and Comparative Development in 2000



## Interpretations – Diversity and Comparative Development in 2000

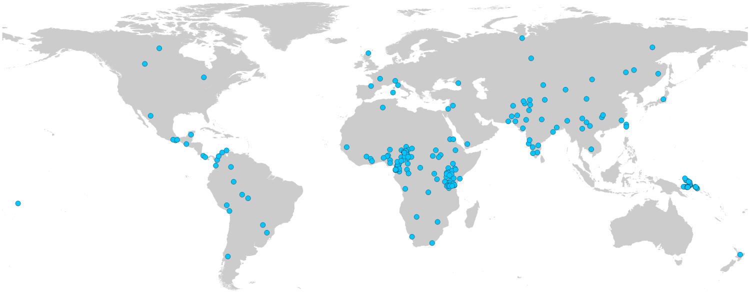
- Optimal GD in 2000 $= 0.7208 \approx$ GD in US $= 0.7206$
- Increasing GD of Bolivia (0.63), the most homogeneous country, by:
  - 0.09 $\Longrightarrow$ 5.4-fold increase income per capita in 2000
    - From 9% to 40% of that of the US
  - 0.01 $\Longrightarrow$ 39% increase income per capita in 2000
- Decreasing GD of Ethiopia (0.77), the most heterogeneous country, by:
  - 0.05 $\Longrightarrow$ 1.7-fold increase in income per capita in 2000
    - From 2% to 4% of that of the US
  - 0.01 $\Longrightarrow$ 21% increase in income per capita in 2000
- 0.01 change from the optimal level of GD
  - $\Longrightarrow$ 1.9% decrease in income per capita in 2000

# Addressing Endogenous Post-1500 Migrations

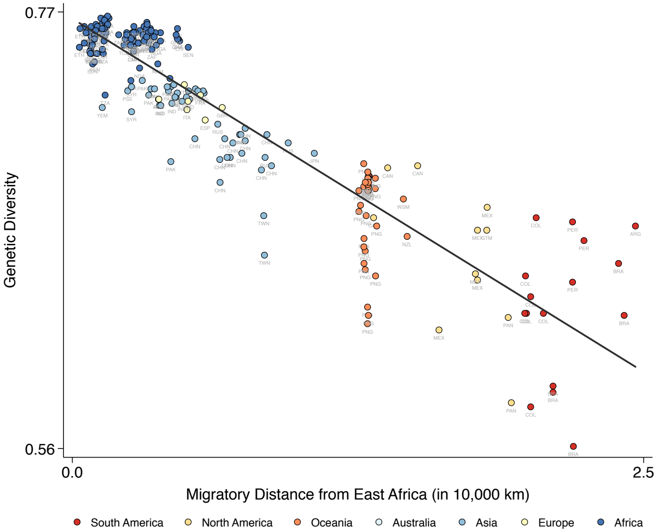
| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Full Sample | Non OECD | w/o Neo Europes | w/o Latin America | w/o Sub Sahara | >0.97 Indigenous |
| | Dependent Variable is Log Income per Capita in 2000 | | | | | |
| Adjusted Diversity | 277.342*** | 271.979*** | 261.367*** | 412.222*** | 264.805** | 304.735** |
| | (70.232) | (88.479) | (70.533) | (148.584) | (111.365) | (111.588) |
| Adjusted Diversity Sqr. | -192.386*** | -188.974*** | -181.811*** | -287.067*** | -183.863** | -213.389** |
| | (49.675) | (62.096) | (49.671) | (101.906) | (80.398) | (77.255) |
| Log Adj. Time of NR | 0.396* | 0.390 | 0.355 | 0.518* | 0.068 | 0.448* |
| | (0.233) | (0.281) | (0.231) | (0.298) | (0.442) | (0.254) |
| Log % of Arable Land | -0.183*** | -0.236*** | -0.201*** | -0.189*** | -0.211** | -0.104 |
| | (0.051) | (0.060) | (0.055) | (0.050) | (0.097) | (0.061) |
| Log Absolute Latitude | 0.009 | -0.021 | -0.025 | -0.139 | 0.218 | -0.074 |
| | (0.108) | (0.119) | (0.111) | (0.126) | (0.242) | (0.130) |
| Social Infrastructure | 1.826*** | 1.313** | 1.416*** | 2.044*** | 1.585*** | 1.311* |
| | (0.417) | (0.579) | (0.507) | (0.545) | (0.486) | (0.716) |
| Ethnic Frac. | -0.333 | -0.437 | -0.390 | -0.752** | 0.104 | -0.044 |
| | (0.280) | (0.375) | (0.300) | (0.348) | (0.408) | (0.412) |
| % Population at Risk of Malaria | -0.502 | -0.605 | -0.591 | -0.308 | -0.425 | -0.153 |
| | (0.351) | (0.381) | (0.370) | (0.486) | (0.581) | (0.434) |
| % Population Living in Tropical Zones | -0.319 | -0.196 | -0.302 | -0.520** | -0.528 | -0.339 |
| | (0.204) | (0.239) | (0.219) | (0.252) | (0.341) | (0.312) |
| Optimal Diversity | 0.721 | 0.720 | 0.719 | 0.718 | 0.720 | 0.714 |
| | (0.083) | (0.085) | (0.015) | (0.023) | (0.180) | (0.012) |
| Observations | 109 | 83 | 105 | 87 | 71 | 37 |
| R-squared | 0.90 | 0.82 | 0.89 | 0.93 | 0.86 | 0.98 |

Bootstrap standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

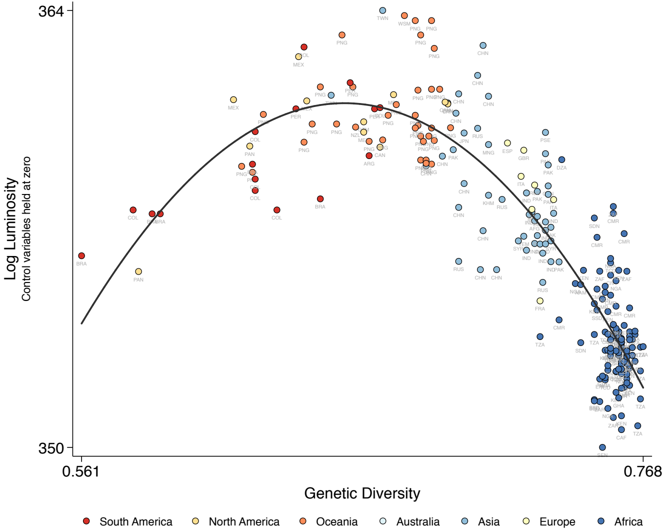
# Observed Genetic Diversity - 232 Ethnic Groups



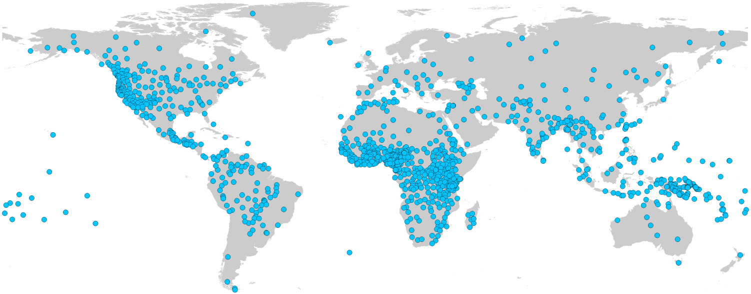
# Migratory Distance from Africa and Genetic Diversity



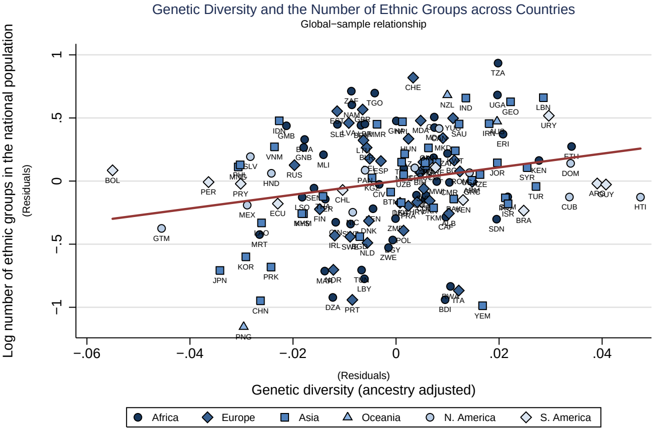
# Genetic Diversity and Productivity of Ethnic Group



# Predicted Genetic Diversity - 1331 Ethnic Groups

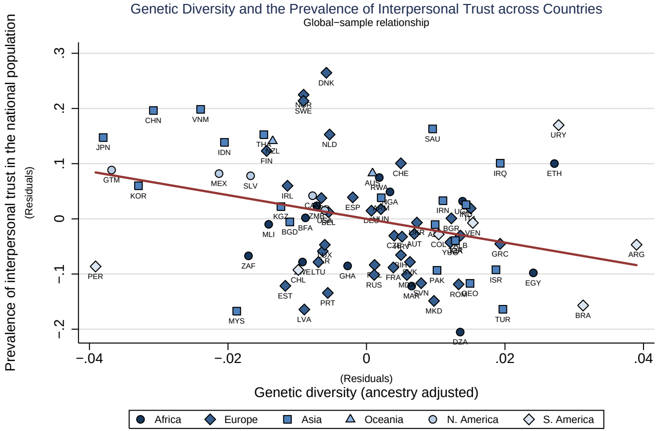


# Genetic Diversity & Cultural Fragmentation



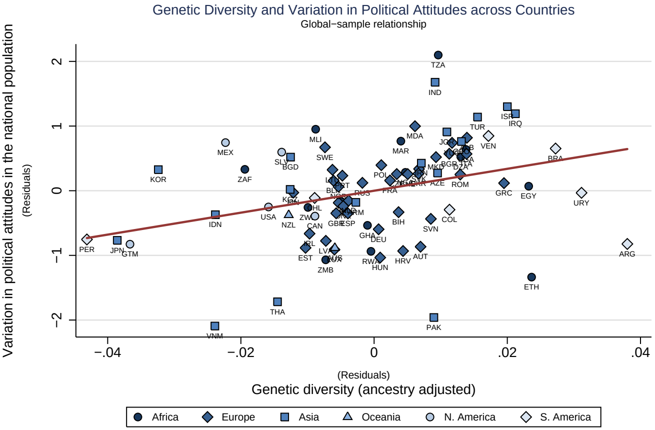

Relationship in the global sample; conditional on baseline geographical controls and continental fixed effects
Slope coefficient = 5.431; (robust) standard error = 1.798; t−statistic = 3.021; partial R−squared = 0.054; observations = 147

# Genetic Diversity & Trust



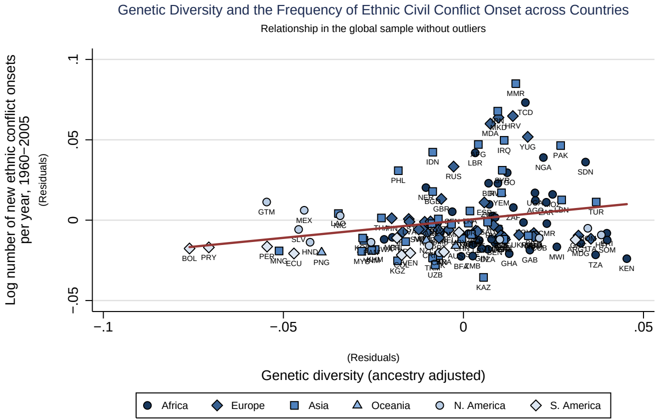

Relationship in the global sample; conditional on baseline geographical controls and regional fixed effects
Slope coefficient = −2.151; (robust) standard error = 0.756; t−statistic = −2.845; partial R−squared = 0.105; observations = 84

# Genetic Diversity & Variation in Political Preferences




Relationship in the global sample; conditional on baseline geographical controls and regional fixed effects
Slope coefficient = 16.963; (robust) standard error = 5.954; t−statistic = 2.849; partial R−squared = 0.111; observations = 81

# Genetic Diversity & Ethnic Civil Conflict




Relationship in the global sample with influential outliers eliminated; conditional on baseline geographical controls

Slope coefficient = 0.220; (robust) standard error = 0.067; t−statistic = 3.260; partial R−squared = 0.048; observations = 140

# The Benefits of Diversity – Evidence

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Patent Applications 1980-2000 | | Scientific Articles 1980-2000 | |
| Predicted Diversity (Ancestry Adjusted) | 0.851** | 0.673** | 2.290*** | 1.816*** |
| | (0.343) | (0.312) | (0.576) | (0.541) |
| Log Years since NR (Ancestry Adjusted) | -0.021 | -0.007 | -0.091* | -0.076 |
| | (0.041) | (0.042) | (0.048) | (0.048) |
| Log % of Arable Land | -0.003 | -0.000 | 0.009 | 0.007 |
| | (0.013) | (0.013) | (0.016) | (0.015) |
| Log Absolute Latitude | 0.010 | 0.017 | 0.045* | 0.055** |
| | (0.015) | (0.013) | (0.024) | (0.023) |
| Social Infrastructure | 0.241*** | 0.177** | 0.685*** | 0.548*** |
| | (0.078) | (0.072) | (0.117) | (0.127) |
| Ethnic Fractionalization | 0.003 | -0.008 | 0.095 | 0.073 |
| | (0.059) | (0.060) | (0.096) | (0.096) |
| % of Population of European Descent | 0.029 | -0.042 | 0.042 | -0.040 |
| | (0.067) | (0.062) | (0.090) | (0.081) |
| % of Population at Risk of Contracting Malaria | 0.031 | 0.035 | 0.102* | 0.131*** |
| | (0.049) | (0.043) | (0.055) | (0.048) |
| Terrain Ruggedness | -0.086 | -0.060 | -0.349* | -0.269 |
| | (0.098) | (0.095) | (0.177) | (0.169) |
| % of Population Living in Tropical Zones | -0.021 | 0.004 | 0.018 | 0.049 |
| | (0.036) | (0.031) | (0.058) | (0.055) |
| Mean Distance to Nearest Waterway | -0.037 | -0.031 | 0.105*** | 0.118*** |
| | (0.044) | (0.044) | (0.038) | (0.035) |
| Years of Schooling | | 0.020*** | | 0.032*** |
| | | (0.007) | | (0.008) |
| Observations | 77 | 77 | 93 | 93 |
| R-squared | 0.74 | 0.77 | 0.80 | 0.82 |

## Conclusions: Genetic Diversity to Comparative Development

- The migration of humans out of Africa 70,000-90,000 BP affected
  - The distribution of genetic diversity across the globe
  - Comparative economic development
    - Accounts for 16% of the variation in the income per capita across countries

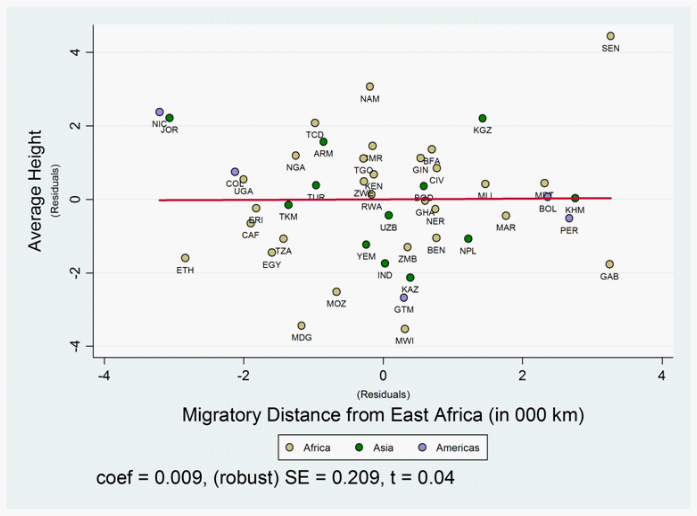
## Conclusions: Costs and Benefits of Diversity

- Diversity adversely affects the cohesiveness of society, increasing the incidence of:
  - Mistrust (Ashraf-Galor, AER 2013)
  - Civil conflicts (Arbatli-Ashraf-Galor, 2015)
  - Ethnic fractionalization (Ashraf-Galor, AER P&P 2013)
- Diversity enhances innovations and knowledge creation

## Policy Implications

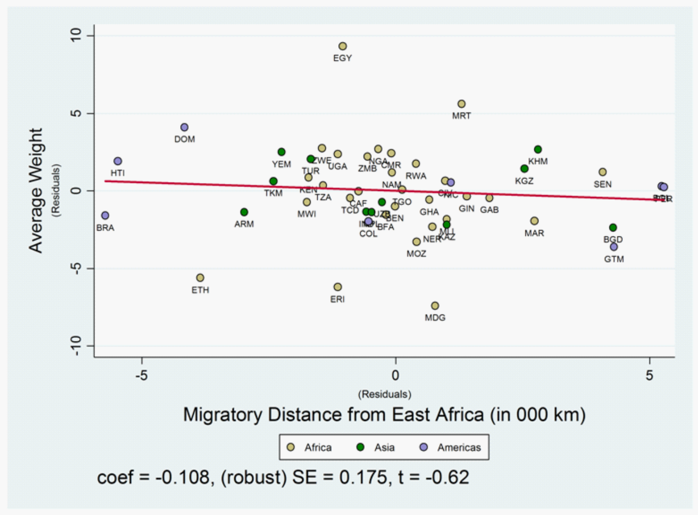
- Education policy
  - In overly-diverse societies:
    - Education geared towards: social cohesiveness & tolerance
    - $\Longrightarrow$ Mitigating the cost of diversity
  - In overly-homogeneous societies:
    - cultivation of cultural diversity
    - $\Longrightarrow$ substitute for low genetic diversity
- Optimal level of cultural assimilation

# Migratory Distance from East Africa and Height



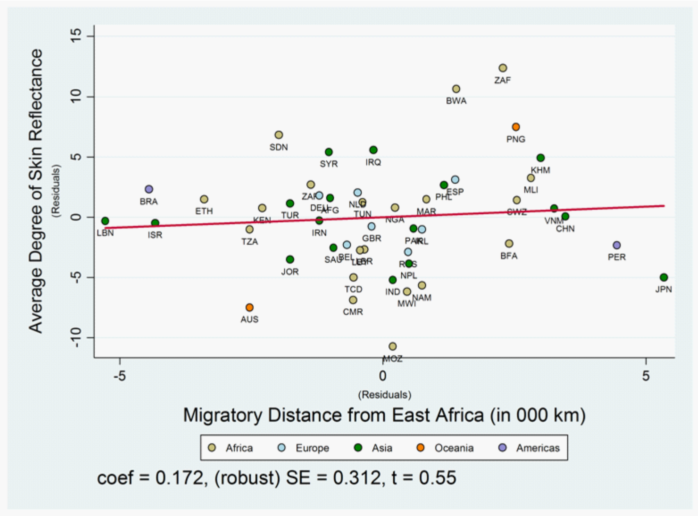

Accounting for distance from the equator.

# Migratory Distance from East Africa and Weight



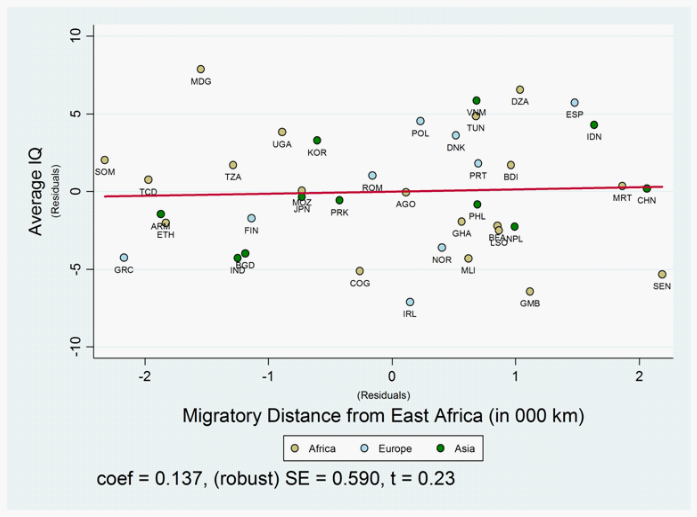

Accounting for distance from the equator.

# Migratory Distance from East Africa and Skin Reflectance



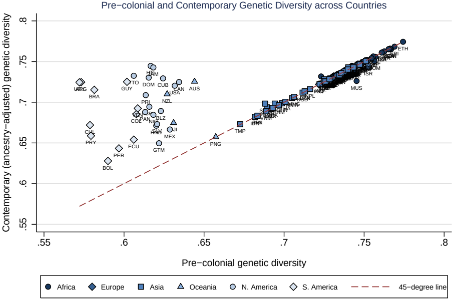

Accounting for distance from the equator.

# Migratory Distance from East Africa and IQ




Accounting for distance from the equator.

# The Impact of Post-1500 Migrations on Genetic Diversity




Correlation in the global sample = 0.750; correlation in the Old−World sample = 0.993

# Theoretical Foundations of the Hump-Shaped Effect of Diversity

$$y = (1 - \alpha\omega)A(z,\omega)f(x) \equiv y(\omega); \qquad \alpha \in (0,1)$$

- $y \equiv$ output per capita
- $A(z,\omega) \equiv$ technological level
- $\omega \in [0,1] \equiv$ degree of diversity
- $z \equiv$ institutional, geographical, and human capital factors
- $f(x) \equiv$ production function
- $x \equiv$ inputs per capita

# Theoretical Foundations of the Hump-Shaped Effect of Diversity

- Diversity and TFP growth

$$A(z,\omega) > 0,\ A_\omega(z,\omega) > 0,\ \ A_{\omega\omega}(z,\omega) < 0$$

$$\lim_{\omega \longrightarrow 0} A_\omega(z,\omega) = \infty;\ \lim_{\omega \longrightarrow 1} A_\omega(z,\omega) = 0$$

- For instance:

$$A(z,\omega) = z \int_0^\omega \omega_i^\theta di \qquad \theta \in (0,1)$$

## Theoretical Foundations of the Hump-Shaped Effect of Diversity

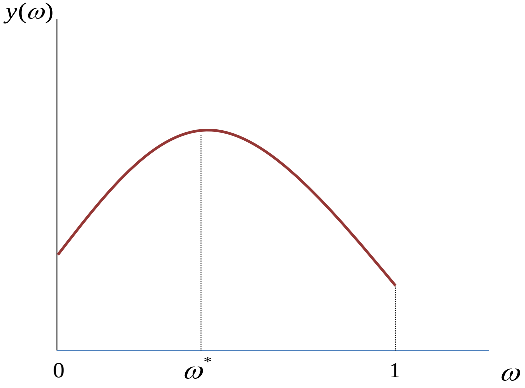
- Properties of $y(\omega)$

$$y'(\omega) = [(1-\alpha\omega)A_{\omega}(z,\omega) - \alpha A(z,\omega)]f(x)$$

$$y''(\omega) = [(1-\alpha\omega)A_{\omega\omega}(z,\omega) - 2\alpha A_{\omega}(z,\omega)]f(x) < 0$$

$$\lim_{\omega \longrightarrow 0} y'(\omega) > 0; \quad \lim_{\omega \longrightarrow 1} y'(\omega) < 0$$

# The Optimal Level of Genetic Diversity



# An Rise in the Optimal Diversity – Faster Technological Progress