

**Ancestry, Language and Culture**

**Enrico Spolaore and Romain Wacziarg**

**May 2015**

**Online Appendix – Supplementary Empirical Results**

**Part A – Alternative choices of sample and questions.**  
**Metric of cultural distance: Euclidian distance.**

In Part A of this appendix we consider alternative samples and sets of questions. In the baseline included in the paper, we sought to maximize the country coverage. The requirement that the same number of questions be included in the analysis for all country pairs led to the loss of many WVS questions. In the tables that follow we replicate Tables 7, 8 and 9 in the paper considering alternative tradeoffs between country coverage and question coverage, resulting in the use of more questions but fewer countries.

**1. Maximizing the number of questions, but including China.**

**Table 7A1 – Correlations between Genetic Distance and Cultural Distance measures  
(Maximizing # of questions with China, Euclidian distance)**

	CD, all questions	CD, cat. A	CD, cat. B	CD, cat. C	CD, cat. D	CD, cat. E	CD, cat. F	CD, cat. G	CD, binary questions	CD, Non-binary questions
CD, category A	0.911*	1								
CD, category B	0.098*	0.0318	1							
CD, category C	0.739*	0.587*	0.027	1						
CD, category D	0.611*	0.511*	0.024	0.469*	1					
CD, category E	0.851*	0.689*	0.064*	0.515*	0.431*	1				
CD, category F	0.673*	0.467*	0.097*	0.458*	0.403*	0.478*	1			
CD, category G	0.345*	0.299*	0.131*	0.117*	0.277*	0.251*	0.257*	1		
CD, binary questions	0.923*	0.948*	0.051	0.740*	0.535*	0.667*	0.538*	0.242*	1	
CD, non-binary questions	0.953*	0.783*	0.125*	0.658*	0.606*	0.905*	0.707*	0.389*	0.763*	1
Weighted F <sub>ST</sub> gen. dist.	0.172*	0.191*	0.080*	0.078*	0.057*	0.127*	0.121*	0.078*	0.175*	0.151*

(Correlations based on 1,275 observations; \* denotes 5% significance). We are using a matrix of 51 countries and 164 questions. We are trying to maximize the number of questions in our analysis. 51 country combinations gives us a total of 1275 observations. In this combination China is included.

Key for WVS question categories: Category A: Perceptions of Life. Category C: Work. Category D: Family. Category E: Politics and Society. Category F: Religion and Morale. Category G: National Identity.

**Table 8A1 – Cultural Distance and Genetic Distance  
(Maximizing # of questions with China, Euclidian distance)**

	(1)	(2)	(3)	(4)	(5)	(6)
<b>Variable</b>	<b>Total</b>	<b>Total</b>	<b>Binary</b>	<b>Binary</b>	<b>Non-binary</b>	<b>Non-binary</b>
<b>F<sub>ST</sub> genetic distance, weighted</b>	<b>187.893</b> <b>(6.22)***</b>	<b>117.739</b> <b>(3.42)***</b>	<b>89.831</b> <b>(6.33)***</b>	<b>79.399</b> <b>(4.82)***</b>	<b>98.061</b> <b>(5.45)***</b>	<b>38.341</b> <b>(1.88)*</b>
Geodesic Distance, 1000s of km		-3.427 (2.93)***		-2.060 (3.68)***		-1.366 (1.97)**
Absolute difference in longitudes		0.322 (4.27)***		0.187 (5.19)***		0.134 (3.00)***
Absolute difference in latitudes		0.174 (1.43)		0.071 (1.21)		0.103 (1.43)
1 for contiguity		-55.985 (9.37)***		-24.837 (8.68)***		-31.147 (8.80)***
Number of landlocked countries in the pair		-8.496 (3.25)***		-5.486 (4.38)***		-3.009 (1.94)*
Number of island countries in the pair		19.074 (5.21)***		1.974 (1.13)		17.100 (7.88)***
1 if pair shares at least one sea or ocean		-21.332 (4.35)***		-6.286 (2.68)***		-15.046 (5.18)***
Constant	-11.184 (5.05)***	-4.644 (1.48)	-5.164 (4.95)***	-1.480 (0.99)	-6.020 (4.56)***	-3.163 (1.70)*
Adjusted R <sup>2</sup>	0.03	0.17	0.03	0.14	0.02	0.17
# of observations	1,275	1,250	1,275	1,250	1,275	1,250
Standardized beta	0.172	0.108	0.175	0.154	0.151	0.059

(t-statistics in parentheses; \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%)

We are using a matrix of 51 countries and 164 questions. We are trying to maximize the number of questions in our analysis. 51 country combinations gives us a total of 1,275 observations. In this combination China is included.

**Table 9A1 - Cultural Distance and Genetic Distance, by question category  
(Maximizing # of questions with China, Euclidian distance)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variable	Category A	Category B	Category C	Category D	Category E	Category F	Category G
<b>Univariate Specification</b>							
F <sub>ST</sub> genetic distance, weighted	93.152 (6.95)***	2.845 (2.87)***	16.405 (2.80)***	4.579 (2.04)**	43.713 (4.57)***	24.133 (4.34)***	3.067 (2.79)***
Adjusted R <sup>2</sup>	0.04	0.01	0.01	0.00	0.02	0.01	0.01
Standardized beta	0.191	0.080	0.078	0.057	0.127	0.121	0.078
<b>Multivariate Specification</b>							
F <sub>ST</sub> genetic distance, weighted	67.806 (4.41)***	5.129 (4.27)***	22.249 (3.25)***	-0.984 (0.40)	14.652 (1.30)	11.205 (1.69)*	-2.318 (1.77)*
Adjusted R <sup>2</sup>	0.16	0.05	0.11	0.21	0.11	0.09	0.08
Standardized beta	0.139	0.144	0.106	-0.012	0.043	0.056	-0.059

(t-statistics in parentheses; \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%)

We are using a matrix of 51 countries and 164 questions. We are trying to maximize the number of questions in our analysis. 51 country combinations gives us a total of 1,275 observations. In this combination China is included.

The univariate specification is based on 1,275 observations (country pairs). The multivariate specification is based on 1,250 observations. All specifications include an intercept.

The multivariate specification includes the following geographic controls: Geodesic distance, absolute difference in longitudes, absolute difference in latitudes, dummy for contiguity, number of landlocked countries in the pair, number of island countries in the pair, dummy=1 if pair shares at least one sea or ocean.

Key for WVS question categories: Category A: Perceptions of Life. Category C: Work. Category D: Family. Category E: Politics and Society. Category F: Religion and Morale. Category G: National Identity.

2. Maximizing the number of questions, but excluding China (version 1).

Table 7A2 – Correlations between Genetic Distance and Cultural Distance measures  
(Maximizing # of questions without China, version 1, Euclidian distance)

	CD, all questions	CD, cat. A	CD, cat. B	CD, cat. C	CD, cat. D	CD, cat. E	CD, cat. F	CD, cat. G	CD, binary questions	CD, Non-binary questions
CD, category A	0.863*	1								
CD, category B	0.107*	0.006	1							
CD, category C	0.776*	0.615*	0.112*	1						
CD, category D	0.770*	0.547*	0.085*	0.612*	1					
CD, category E	0.825*	0.670*	0.071*	0.559*	0.583*	1				
CD, category F	0.686*	0.372*	0.100*	0.444*	0.548*	0.386*	1			
CD, category G	0.315*	0.230*	0.108*	0.146*	0.300*	0.219*	0.255*	1		
CD, binary questions	0.929*	0.913*	0.068*	0.750*	0.649*	0.635*	0.631*	0.233*	1	
CD, non-binary questions	0.957*	0.739*	0.127*	0.719*	0.789*	0.892*	0.661*	0.347*	0.781*	1
Weighted $F_{ST}$ gen. dist.	0.131*	0.061	0.027	0.166*	0.267*	0.057	0.098*	0.035	0.091*	0.149*

(Correlations based on 990 observations; \* denotes 5% significance). In this version we are using a matrix of 45 countries and 226 questions. We are trying to maximize the number of questions in our analysis. 45 country combinations gives us a total of 990 observations. In this combination China is excluded.

Key for WVS question categories: Category A: Perceptions of Life. Category C: Work. Category D: Family. Category E: Politics and Society. Category F: Religion and Morale. Category G: National Identity.

**Table 8A2 – Cultural Distance and Genetic Distance**  
**(Maximizing # of questions without China, version 1, Euclidian distance)**

<b>Variable</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>(4)</b>	<b>(5)</b>	<b>(6)</b>
	<b>Total</b>	<b>Total</b>	<b>Binary</b>	<b>Binary</b>	<b>Non-binary</b>	<b>Non-binary</b>
<b>F<sub>ST</sub> genetic distance, weighted</b>	267.752 (4.15)***	178.678 (2.62)***	87.038 (2.88)***	51.891 (1.61)	180.714 (4.73)***	126.786 (3.16)***
Geodesic Distance, 1000s of km		-5.570 (2.62)***		-2.781 (2.77)***		-2.789 (2.23)**
Absolute difference in longitudes		0.413 (3.18)***		0.245 (4.00)***		0.167 (2.19)**
Absolute difference in latitudes		0.380 (1.71)*		0.117 (1.12)		0.262 (2.01)**
1 for contiguity		-75.251 (8.61)***		-33.302 (8.06)***		-41.949 (8.15)***
Number of landlocked countries in the pair		-7.310 (1.81)*		-6.590 (3.46)***		-0.720 (0.30)
Number of island countries in the pair		22.254 (3.51)***		1.572 (0.52)		20.682 (5.53)***
1 if pair shares at least one sea or ocean		-29.907 (3.96)***		-9.087 (2.54)**		-20.820 (4.68)***
Constant	-11.198 (3.18)***	-1.575 (0.33)	-3.640 (2.21)**	1.873 (0.82)	-7.558 (3.63)***	-3.449 (1.22)
Adjusted R <sup>2</sup>	0.02	0.14	0.01	0.12	0.02	0.15
# observations	990	967	990	967	990	967
Standardized beta	0.131	0.087	0.091	0.054	0.149	0.105

(t-statistics in parentheses; \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%)

We are using a matrix of 45 countries and 226 questions. We are trying to maximize the number of questions in our analysis. 45 country combinations gives us a total of 990 observations. In this combination China is excluded (version 1).

**Table 9A2 – Cultural Distance and Genetic Distance, by question category  
(Maximizing # of questions without China, version 1, Euclidian distance)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variable	Category A	Category B	Category C	Category D	Category E	Category F	Category G
<b>Univariate Specification</b>							
F <sub>ST</sub> genetic distance, weighted,	49.263 (1.93)*	1.311 (0.84)	58.608 (5.30)***	68.532 (8.69)***	32.067 (1.78)*	55.986 (3.10)***	1.984 (1.11)
Adjusted R <sup>2</sup>	0.00	-0.00	0.03	0.07	0.00	0.01	0.00
Standardized beta	0.061	0.027	0.166	0.266	0.057	0.098	0.035
<b>Multivariate Specification</b>							
F <sub>ST</sub> genetic distance, weighted	6.524 (0.24)	4.062 (2.36)**	64.066 (5.36)***	66.716 (7.96)***	12.201 (0.63)	28.538 (1.44)	-3.430 (1.75)*
Adjusted R <sup>2</sup>	0.14	0.04	0.10	0.18	0.10	0.07	0.06
Standardized beta	0.008	0.083	0.182	0.259	0.022	0.050	-0.061

(t-statistics in parentheses; \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%)

We are using a matrix of 45 countries and 226 questions. We are trying to maximize the number of questions in our analysis. 45 country combinations gives us a total of 990 observations. In this combination China is excluded.

The univariate specification is based on 990 observations (country pairs). The multivariate specification is based on 967 observations.

All specifications include an intercept.

The multivariate specification includes the following geographic controls: Geodesic distance, absolute difference in longitudes, absolute difference in latitudes, dummy for contiguity, number of landlocked countries in the pair, number of island countries in the pair, dummy=1 if pair shares at least one sea or ocean.

Key for WVS question categories: Category A: Perceptions of Life. Category C: Work. Category D: Family. Category E: Politics and Society. Category F: Religion and Morale. Category G: National Identity.

3. Maximizing the number of questions, but excluding China (version 2).

**Table 7A3 – Correlations between Genetic Distance and Cultural Distance measures  
(Maximizing # of questions without China, version 2, Euclidian distance)**

	CD, all questions	CD, cat. A	CD, cat. B	CD, cat. C	CD, cat. D	CD, cat. E	CD, cat. F	CD, cat. G	CD, binary questions	CD, Non-binary questions
CD, category A	0.878*	1								
CD, category B	0.100*	0.006	1							
CD, category C	0.748*	0.603*	0.048	1						
CD, category D	0.617*	0.522*	0.024	0.472*	1					
CD, category E	0.828*	0.669*	0.062*	0.558*	0.444*	1				
CD, category F	0.694*	0.393*	0.119*	0.423*	0.382*	0.403*	1			
CD, category G	0.342*	0.284*	0.154*	0.118*	0.279*	0.229*	0.267*	1		
CD, binary questions	0.938*	0.914*	0.047	0.732*	0.555*	0.645*	0.655*	0.266*	1	
CD, non-binary questions	0.953*	0.758*	0.137*	0.685*	0.608*	0.906*	0.657*	0.374*	0.790*	1
Weighted F <sub>ST</sub> gen. dist.	0.174*	0.127*	-0.008	0.177*	0.089*	0.117*	0.161*	0.089*	0.156*	0.172*

(Correlations based on 1,081 observations; \* denotes 5% significance). In this version we are using a matrix of 47 countries and 164 questions. We are trying to maximize the number of questions in our analysis. 47 country combinations gives us a total of 1,081 observations. In this combination China is excluded.

Key for WVS question categories: Category A: Perceptions of Life. Category C: Work. Category D: Family. Category E: Politics and Society. Category F: Religion and Morale. Category G: National Identity.



**Table 8A3 – Cultural Distance and Genetic Distance**  
**(Maximizing # of questions without China, version 2, Euclidian distance)**

<b>Variable</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>(4)</b>	<b>(5)</b>	<b>(6)</b>
	<b>Total</b>	<b>Total</b>	<b>Binary</b>	<b>Binary</b>	<b>Non-binary</b>	<b>Non-binary</b>
<b>F<sub>ST</sub> genetic distance, weighted</b>	278.916 (5.80)***	145.647 (2.72)***	123.624 (5.19)***	70.541 (2.62)***	155.291 (5.73)***	75.106 (2.49)**
Geodesic Distance, 1000s of km		-6.435 (4.10)***		-3.516 (4.44)***		-2.919 (3.30)***
Absolute difference in longitudes		0.460 (4.61)***		0.281 (5.60)***		0.178 (3.18)***
Absolute difference in latitudes		0.471 (2.85)***		0.213 (2.55)**		0.259 (2.78)***
1 for contiguity		-63.734 (8.97)***		-29.215 (8.16)***		-34.519 (8.63)***
Number of landlocked countries in the pair		-9.827 (3.06)***		-6.826 (4.22)***		-3.002 (1.66)*
Number of island countries in the pair		29.562 (6.39)***		8.912 (3.82)***		20.650 (7.93)***
1 if pair shares at least one sea or ocean		-26.454 (4.41)***		-9.043 (2.99)***		-17.410 (5.15)***
Constant	-14.325 (4.95)***	-4.927 (1.30)	-6.274 (4.37)***	-1.751 (0.91)	-8.050 (4.94)***	-3.176 (1.48)
Adjusted R <sup>2</sup>	0.03	0.19	0.02	0.16	0.03	0.19
# observations	1,081	1,058	1,081	1,058	1,081	1,058
Standardized beta	0.174	0.091	0.156	0.089	0.172	0.083

(t-statistics in parentheses; \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%)

We are using a matrix of 47 countries and 188 questions. We are trying to maximize the number of questions in our analysis. 47 country combinations gives us a total of 1081 observations. In this combination China is excluded (version 2).

**Table 9A3 – Cultural Distance and Genetic Distance, by question category  
(Maximizing # of questions without China, version 2, Euclidian distance)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variable	Category A	Category B	Category C	Category D	Category E	Category F	Category G
<b>Univariate Specification</b>							
F <sub>ST</sub> genetic distance, weighted,	84.957 (4.20)***	-0.359 (0.27)	47.729 (5.91)***	9.016 (2.93)***	56.265 (3.88)***	76.791 (5.36)***	4.518 (2.93)***
Adjusted R <sup>2</sup>	0.02	-0.00	0.03	0.01	0.01	0.03	0.01
Standardized beta	0.127	-0.008	0.177	0.089	0.117	0.161	0.089
<b>Multivariate Specification</b>							
F <sub>ST</sub> genetic distance, weighted	25.567 (1.13)	3.975 (2.51)**	51.135 (5.49)***	0.855 (0.26)	21.984 (1.32)	44.913 (2.68)***	-2.782 (1.54)
Adjusted R <sup>2</sup>	0.16	0.05	0.13	0.23	0.12	0.11	0.08
Standardized beta	0.038	0.091	0.190	0.008	0.046	0.094	-0.055

(t-statistics in parentheses; \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%)

We are using a matrix of 47 countries and 188 questions. We are trying to maximize the number of questions in our analysis. 47 country combinations gives us a total of 1,081 observations. In this combination China is excluded.

The univariate specification is based on 1,081 observations (country pairs). The multivariate specification is based on 1,058 observations.

All specifications include an intercept.

The multivariate specification includes the following geographic controls: Geodesic distance, absolute difference in longitudes, absolute difference in latitudes, dummy for contiguity, number of landlocked countries in the pair, number of island countries in the pair, dummy=1 if pair shares at least one sea or ocean.

Key for WVS question categories: Category A: Perceptions of Life. Category C: Work. Category D: Family. Category E: Politics and Society. Category F: Religion and Morale. Category G: National Identity.

**Part B – Alternative metrics of cultural distance: Manhattan Distance.**  
**Sample choice: maximizing the number of countries available.**

The three tables that follow redo Tables 7, 8 and 9 in the main paper, using Manhattan distance as a measure of cultural distance from the WVS.

**Table 7B – Correlations between Genetic Distance and Cultural Distance measures**  
**(Maximizing # of countries, Manhattan distance)**

	<b>CD, all questions</b>	<b>CD, cat. A</b>	<b>CD, cat. C</b>	<b>CD, cat. D</b>	<b>CD, cat. E</b>	<b>CD, cat. F</b>	<b>CD, cat. G</b>	<b>CD, binary questions</b>	<b>CD, Non-binary questions</b>
CD, category A	0.879*	1							
CD, category C	0.644*	0.481*	1						
CD, category D	0.620*	0.530*	0.375*	1					
CD, category E	0.802*	0.598*	0.284*	0.392*	1				
CD, category F	0.729*	0.544*	0.394*	0.399*	0.476*	1			
CD, category G	0.443*	0.327*	0.311*	0.217*	0.308*	0.287*	1		
CD, binary questions	0.870*	0.837*	0.816*	0.584*	0.514*	0.544*	0.315*	1	
CD, non-binary questions	0.949*	0.787*	0.448*	0.560*	0.878*	0.749*	0.466*	0.670*	1
Weighted F <sub>ST</sub> gen. dist.	0.269*	0.245*	0.075*	0.026	0.298*	0.216*	0.129*	0.147*	0.311*

(Correlations based on 2,628 observations; \* denotes 5% significance). We are using a matrix of 73 countries and 98 questions. We are trying to maximize the number of countries in our analysis. 73 country combinations gives us a total of 2,628 observations.

Key for WVS question categories: Category A: Perceptions of Life. Category C: Work. Category D: Family. Category E: Politics and Society. Category F: Religion and Morale. Category G: National Identity.

**Table 8B – Cultural Distance and Genetic Distance  
(Maximizing # of countries, Manhattan distance)**

<b>Variable</b>	<b>(1) Total</b>	<b>(2) Total</b>	<b>(3) Binary</b>	<b>(4) Binary</b>	<b>(5) Non- binary</b>	<b>(6) Non- binary</b>
<b>F<sub>ST</sub> genetic distance, weighted</b>	143.659 (14.30)***	137.954 (12.18)***	33.291 (7.60)***	29.873 (6.02)***	110.368 (16.75)***	108.081 (14.53)***
Geodesic Distance, 1000s of km		-1.071 (2.14)**		-0.334 (1.52)		-0.737 (2.24)**
Absolute difference in longitudes		0.071 (1.96)**		0.019 (1.21)		0.052 (2.18)**
Absolute difference in latitudes		0.107 (2.14)**		0.037 (1.67)*		0.071 (2.14)**
1 for contiguity		-34.388 (9.92)***		-13.900 (9.16)***		-20.488 (9.00)***
Number of landlocked countries in the pair		-8.808 (7.32)***		-4.027 (7.65)***		-4.781 (6.05)***
Number of island countries in the pair		-3.252 (2.07)**		-2.704 (3.93)***		-0.548 (0.53)
1 if pair shares at least one sea or ocean		-12.103 (5.25)***		-2.999 (2.98)***		-9.103 (6.02)***
Constant	-10.549 (10.84)***	-3.685 (2.50)**	-2.368 (5.58)***	0.702 (1.09)	-8.182 (12.83)***	-4.386 (4.53)***
Adjusted R <sup>2</sup>	0.07	0.14	0.02	0.08	0.10	0.16
# observations	2,628	2,513	2,628	2,513	2,628	2,513
Standardized beta	0.269	0.254	0.147	0.130	0.311	0.300

(t-statistics in parentheses; \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%)

We are using a matrix of 73 countries and 98 questions. We are trying to maximize the number of countries in our analysis. 73 country combinations gives us a total of 2,628 observations.

**Table 9B – Cultural Distance and Genetic Distance, by question category  
(Maximizing # of countries, Manhattan distance)**

	(1)	(3)	(4)	(5)	(6)	(7)
Variable	Category A	Category C	Category D	Category E	Category F	Category G
<b>Univariate Specification</b>						
F <sub>ST</sub> genetic distance, weighted,	47.646 (12.93)***	9.019 (3.85)***	1.455 (1.34)	57.263 (15.99)***	24.171 (11.33)***	4.105 (6.67)***
Adjusted R <sup>2</sup>	0.06	0.01	0.00	0.09	0.05	0.02
Standardized beta	0.245	0.075	0.026	0.298	0.216	0.129
<b>Multivariate Specification</b>						
F <sub>ST</sub> genetic distance, weighted	47.659 (11.56)***	7.000 (2.62)***	1.406 (1.15)	54.871 (13.50)***	24.842 (10.09)***	2.175 (3.11)***
Adjusted R <sup>2</sup>	0.14	0.05	0.08	0.14	0.08	0.04
Standardized beta	0.241	0.058	0.025	0.283	0.218	0.069

(t-statistics in parentheses; \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%)

We are using a matrix of 73 countries and 98 questions. We are trying to maximize the number of countries in our analysis. 73 country combinations gives us a total of 2,628 observations

The univariate specification is based on 2,628 observations (country pairs). The multivariate specification is based on 2,513 observations. All specifications include an intercept.

The multivariate specification includes the following geographic controls: Geodesic distance, absolute difference in longitudes, absolute difference in latitudes, dummy for contiguity, number of landlocked countries in the pair, number of island countries in the pair, dummy=1 if pair shares at least one sea or ocean.

Key for WVS question categories: Category A: Perceptions of Life. Category C: Work. Category D: Family. Category E: Politics and Society. Category F: Religion and Morale. Category G: National Identity.

**Part C – Alternative metrics of cultural distance: F<sub>ST</sub> Cultural Distance.**  
**Sample choice: maximizing the number of countries available.**

The three tables that follow redo Tables 7, 8 and 9 in the main paper, using the F<sub>ST</sub> functional form to compute cultural distance from the WVS.

**Table 7C – Correlations between Genetic Distance and Cultural Distance measures**  
**(Maximizing # of countries, F<sub>ST</sub> distance)**

	<b>CD, all questions</b>	<b>CD, cat. A</b>	<b>CD, cat. C</b>	<b>CD, cat. D</b>	<b>CD, cat. E</b>	<b>CD, cat. F</b>	<b>CD, cat. G</b>	<b>CD, binary questions</b>	<b>CD, Non-binary questions</b>
CD, category A	0.903*	1							
CD, category C	0.750*	0.586*	1						
CD, category D	0.675*	0.622*	0.440*	1					
CD, category E	0.736*	0.550*	0.368*	0.410*	1				
CD, category F	0.775*	0.566*	0.509*	0.395*	0.556*	1			
CD, category G	0.411*	0.296*	0.277*	0.159*	0.366*	0.387*	1		
CD, binary questions	0.926*	0.930*	0.805*	0.6538*	0.515*	0.588*	0.274*	1	
CD, non-binary questions	0.916*	0.727*	0.571*	0.588*	0.850*	0.847*	0.490*	0.697*	1
Weighted F <sub>ST</sub> gen. dist.	0.132*	0.070*	0.001	-0.047*	0.235*	0.222*	0.103*	0.026	0.222*

(Correlations based on 2,628 observations; \* denotes 5% significance). We are using a matrix of 73 countries and 98 questions. We are trying to maximize the number of countries in our analysis. 73 country combinations gives us a total of 2,628 observations.

Key for WVS question categories: Category A: Perceptions of Life. Category C: Work. Category D: Family. Category E: Politics and Society. Category F: Religion and Morale. Category G: National Identity.

**Table 8C – Cultural Distance and Genetic Distance  
(Maximizing # of countries, F<sub>ST</sub> distance)**

Variable	(1) Total	(2) Total	(3) Binary	(4) Binary	(5) Non- binary	(6) Non- binary
<b>F<sub>ST</sub> genetic distance, weighted</b>	0.041 (6.80)***	0.038 (5.50)***	0.013 (1.34)	0.015 (1.41)	0.057 (11.66)***	0.050 (9.18)***
Geodesic Distance, 1000s of km		0.000 (1.30)		-0.001 (1.08)		0.001 (3.75)***
Absolute difference in longitudes		-0.000 (1.49)		0.000 (0.21)		-0.000 (3.13)***
Absolute difference in latitudes		-0.000 (1.86)*		0.000 (0.32)		-0.000 (3.99)***
1 for contiguity		-0.016 (7.62)***		-0.025 (7.36)***		-0.011 (6.69)***
Number of landlocked countries in the pair		-0.008 (10.47)***		-0.009 (7.92)***		-0.007 (11.64)***
Number of island countries in the pair		-0.001 (1.15)		-0.005 (3.08)***		0.001 (1.18)
1 if pair shares at least one sea or ocean		-0.006 (4.09)***		-0.008 (3.46)***		-0.005 (4.14)***
Constant	0.036 (61.73)***	0.042 (47.11)***	0.048 (51.85)***	0.057 (39.74)***	0.030 (62.79)***	0.034 (47.81)***
Adjusted R <sup>2</sup>	0.02	0.09	0.00	0.05	0.05	0.13
# observations	2,628	2,513	2,628	2,513	2,628	2,513
Standardized beta	0.132	0.119	0.026	0.031	0.222	0.193

(t-statistics in parentheses; \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%)

We are using a matrix of 73 countries and 98 questions. We are trying to maximize the number of countries in our analysis. 73 country combinations gives us a total of 2,628 observations.

**Table 9C – Cultural Distance and Genetic Distance, by question category  
(Maximizing # of countries, F<sub>ST</sub> distance)**

	(1)	(3)	(4)	(5)	(6)	(7)
Variable	Category A	Category C	Category D	Category E	Category F	Category G
	<b>Univariate Specification</b>					
F <sub>ST</sub> genetic distance, weighted,	0.029 (3.60)***	0.001 (0.07)	-0.024 (2.42)**	0.055 (12.39)***	0.129 (11.66)***	0.029 (5.30)***
Adjusted R2	0.00	-0.00	0.00	0.05	0.05	0.01
Standardized beta	0.070	0.001	-0.047	0.235	0.222	0.103
	<b>Multivariate Specification</b>					
F <sub>ST</sub> genetic distance, weighted	0.029 (3.19)***	0.003 (0.22)	-0.022 (1.98)**	0.050 (10.12)***	0.112 (9.00)***	0.013 (2.22)**
Adjusted R2	0.06	0.04	0.08	0.12	0.12	0.04
Standardized beta	0.069	0.005	-0.043	0.214	0.190	0.049

(t-statistics in parentheses; \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%)

We are using a matrix of 73 countries and 98 questions. We are trying to maximize the number of countries in our analysis. 73 country combinations gives us a total of 2,628 observations

The univariate specification is based on 2,628 observations (country pairs). The multivariate specification is based on 2513 observations.

All specifications include an intercept.

The multivariate specification includes the following geographic controls: Geodesic distance, absolute difference in longitudes, absolute difference in latitudes, dummy for contiguity, number of landlocked countries in the pair, number of island countries in the pair, dummy=1 if pair shares at least one sea or ocean.

Key for WVS question categories: Category A: Perceptions of Life. Category C: Work. Category D: Family. Category E: Politics and Society. Category F: Religion and Morale. Category G: National Identity.



**Part D – Simple Correlations between Various Metrics of Cultural Distance, by question type and category**

This section explores the simple correlations between Euclidian, Manhattan and F<sub>ST</sub> measures of cultural distance for the baseline sample of countries and set of questions used in the main paper. Note that for binary questions Manhattan and Euclidian distances are the same.

**Table D1. Correlation matrices between Euclidian, Manhattan and F<sub>ST</sub> distances for various question types**

**Panel A – Total Index of Cultural Distance (all questions together)**

	<b>Euclidian</b>	<b>Manhattan</b>
Manhattan	0.9999	1
F <sub>ST</sub>	0.8734	0.8731

**Panel B – Breakdown by question type**

	<b>Euclidian</b>	<b>Manhattan</b>	<b>Euclidian</b>	<b>Manhattan</b>
	<b>Binary Distance</b>		<b>Non-binary Distance</b>	
Manhattan	1	1	0.9998	1
F <sub>ST</sub>	0.9172	0.9172	0.8173	0.8166

**Panel C – Breakdown by Question Category**

	<b>Euclidian</b>	<b>Manhattan</b>	<b>Euclidian</b>	<b>Manhattan</b>
	<b>A Category Distance</b>		<b>C Category Distance</b>	
Manhattan	1	1	0.9998	1
F <sub>ST</sub>	0.8611	0.8614	0.9122	0.9118
	<b>D Category Distance</b>		<b>E Category Distance</b>	
Manhattan	0.9997	1	0.9997	1
F <sub>ST</sub>	0.8371	0.8360	0.8416	0.8400
	<b>F Category Distance</b>		<b>G Category Distance</b>	
Manhattan	1	1	0.9889	1
F <sub>ST</sub>	0.7050	0.7051	0.7701	0.7602

Sample: 73 countries, i.e. 2,628 observations