Supplementary Material of "Anisotropic isolation by distance: the main orientations of human genetic differentiation"

- Figure S1: First and second components of the multidimensional scaling analysis for each continent.
- Figure S2: Projections of the first principal components for each continent.
- Figure S3: Correspondence between azimuths/bearings (θ) and cardinal directions.
- Figure S4: F_{st} as a function of the distance computed along the orientation of maximum and minimum differentiation.
- Figure S5: Prediction of average F_{st} for pairs of populations separated by 0 km and by $1,000$ km along the axis of maximum differentiation.
- Figure S6: Grid used for the isolation by distance *ms* simulations.
- Figure S7: Angular distributions of the angles of maximum differentiation found with the three methods under anisotropic isolation by distance models
- Figure S8: Boxplots of the individual localization errors for each population from the 4 continents
- Figure S9: Mean error of the localization method in N-S and E-W directions
- Figure S10: F_{st} of Sub-Saharan populations as a function of the distance computed along the orientation of maximum and minimum differentiation.
- Figure S11: F_{st} of Native American populations, typed with microsatellites, as a function of the distance computed along the orientation of maximum and minimum differentiation.
- Table S1: Orientations of maximum differentiation obtained after different transformations of the dependent and independent variables
- Table S2: Orientations of maximum differentiation after having moved the geographical coordinates by 500 km
- Table S3: Orientations of maximum differentiation obtained with the different methods
- Table S4: Median errors of geographic localization based on the SNP data sets
- Table S5: The different subdivisions of Europe
- Table S6: The different subdivisions of Africa
- Table S7: The different subdivisions of Asia

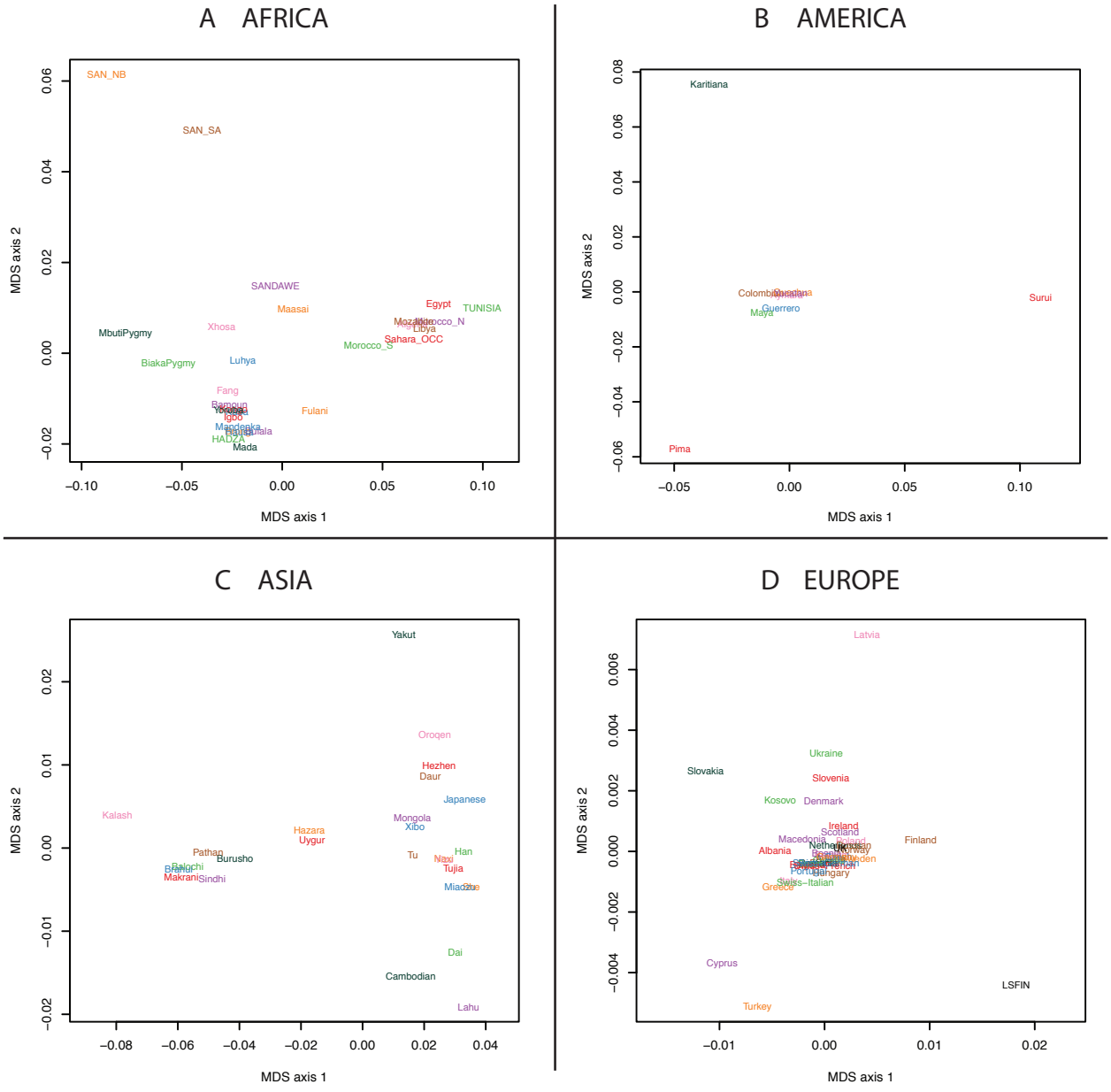

Figure 1: First and second components of multidimensional scaling analysis (MDS). MDS was applied separately in each continent.

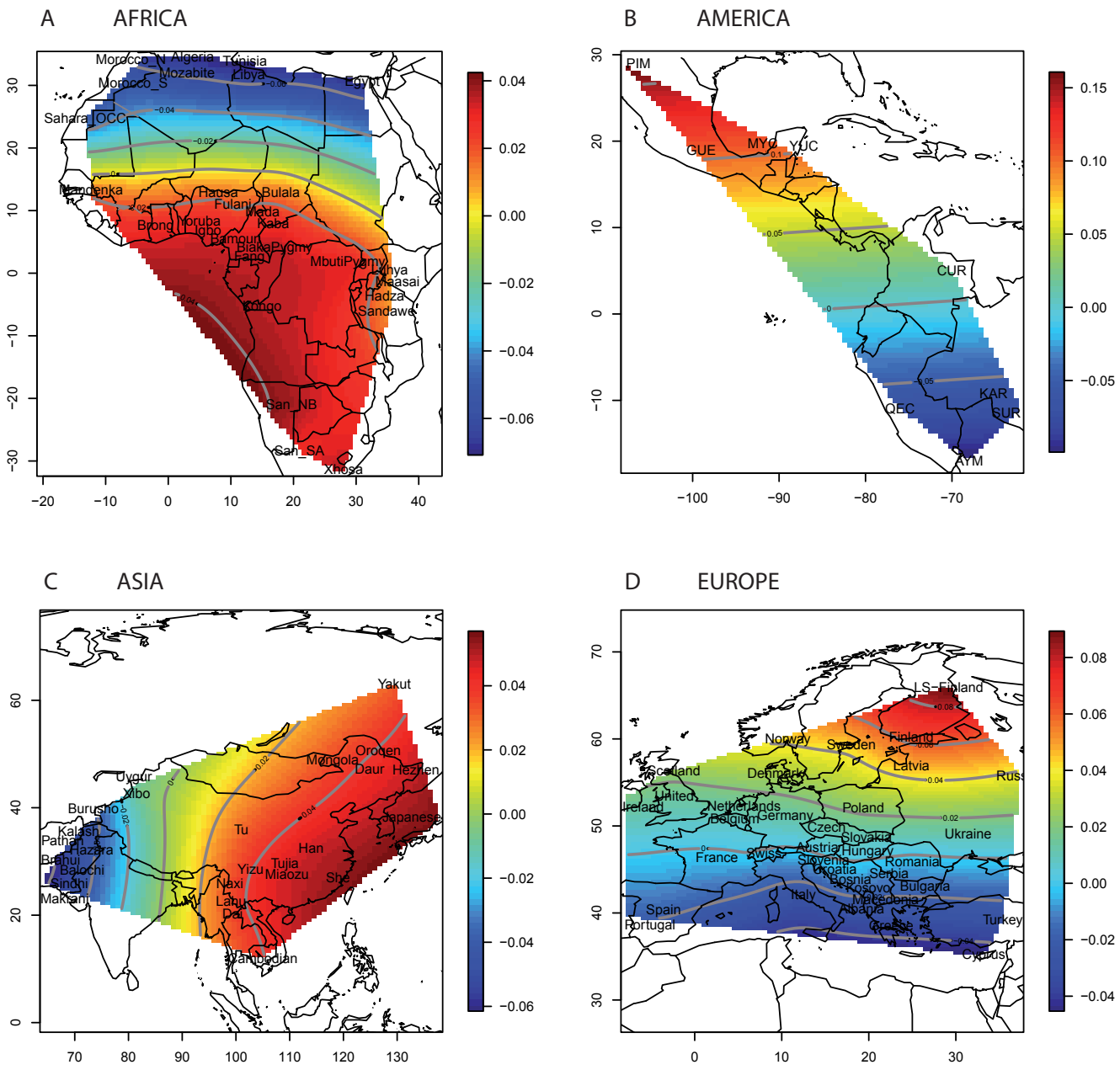

Figure 2: Spatial interpolation of the first component of principal component analysis (PCA). PCA was applied separately in each continent. Spatial interpolation was performed using the *krig* function with a trend surface of degree 2.

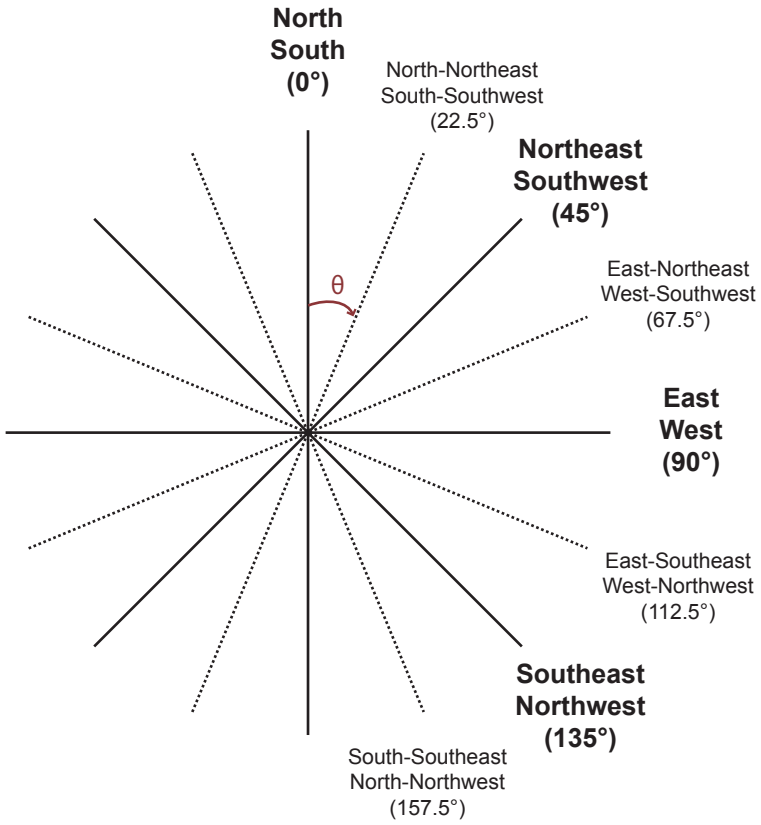

Figure 3: Correspondence between azimuths/bearings (θ) and compass directions.

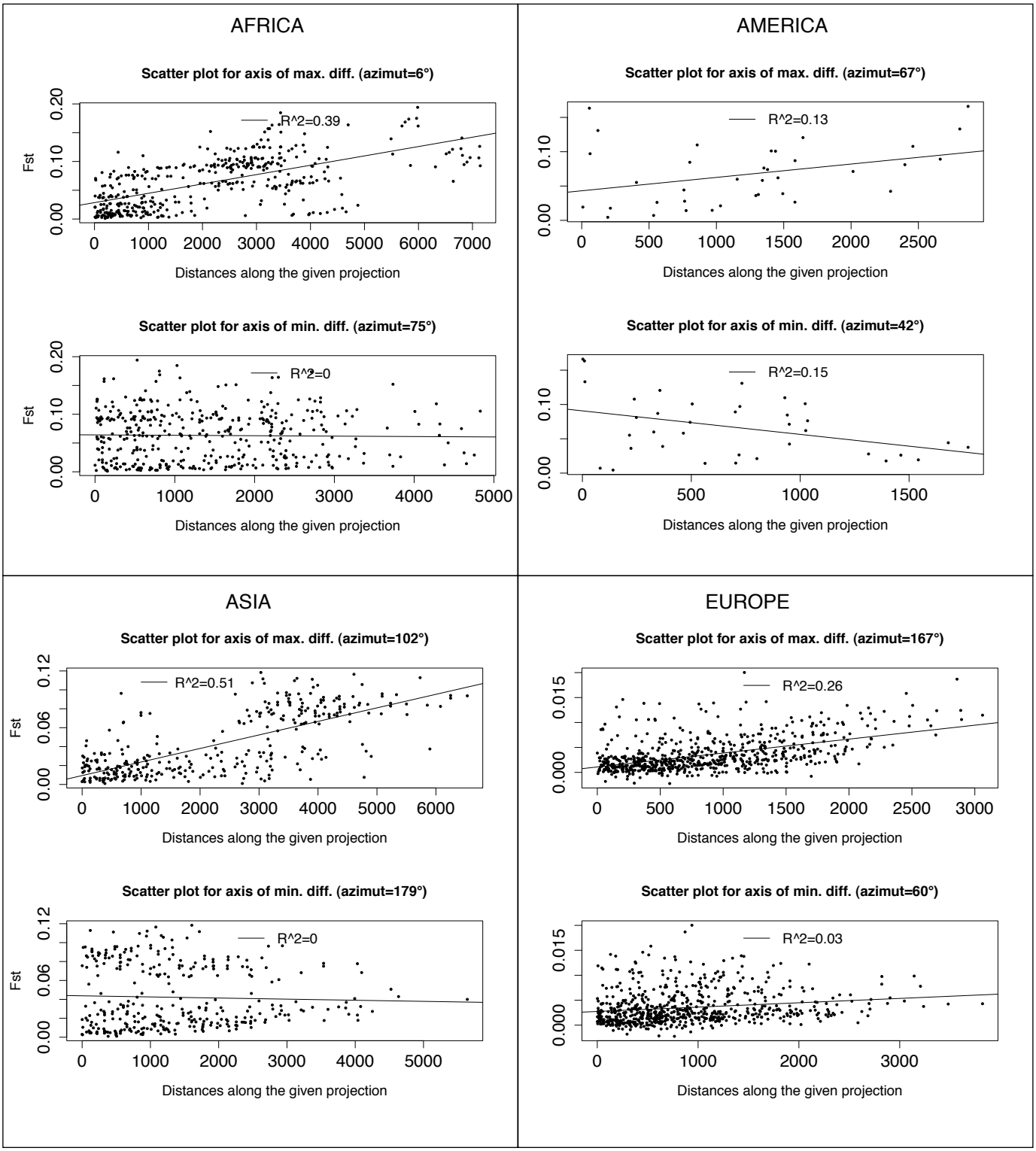

Figure 4: F_{st} as a function of the distance computed along the orientation of maximum and minimum differentiation. Here the orientation of maximum (resp. minimum) differentiation is the orientation that maximizes the correlation between F_{st} and the orientational distances d_θ computed along the different orientations.

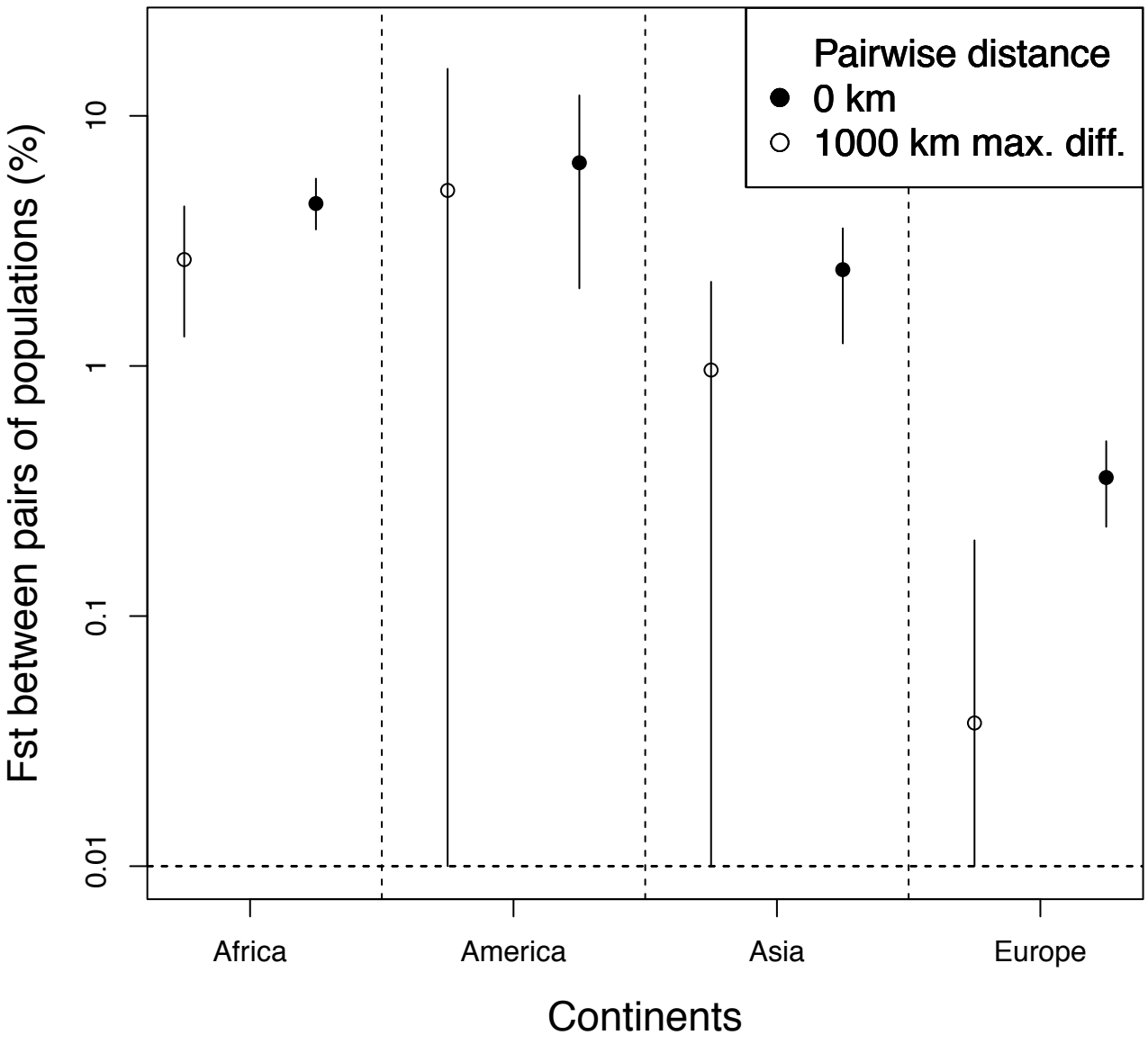

Figure 5: Prediction of average F_{st} for pairs of populations separated by 0 km and by $1,000$ km along the axis of maximum differentiation. The F_{st} was regressed with the equation (1) of the main text.

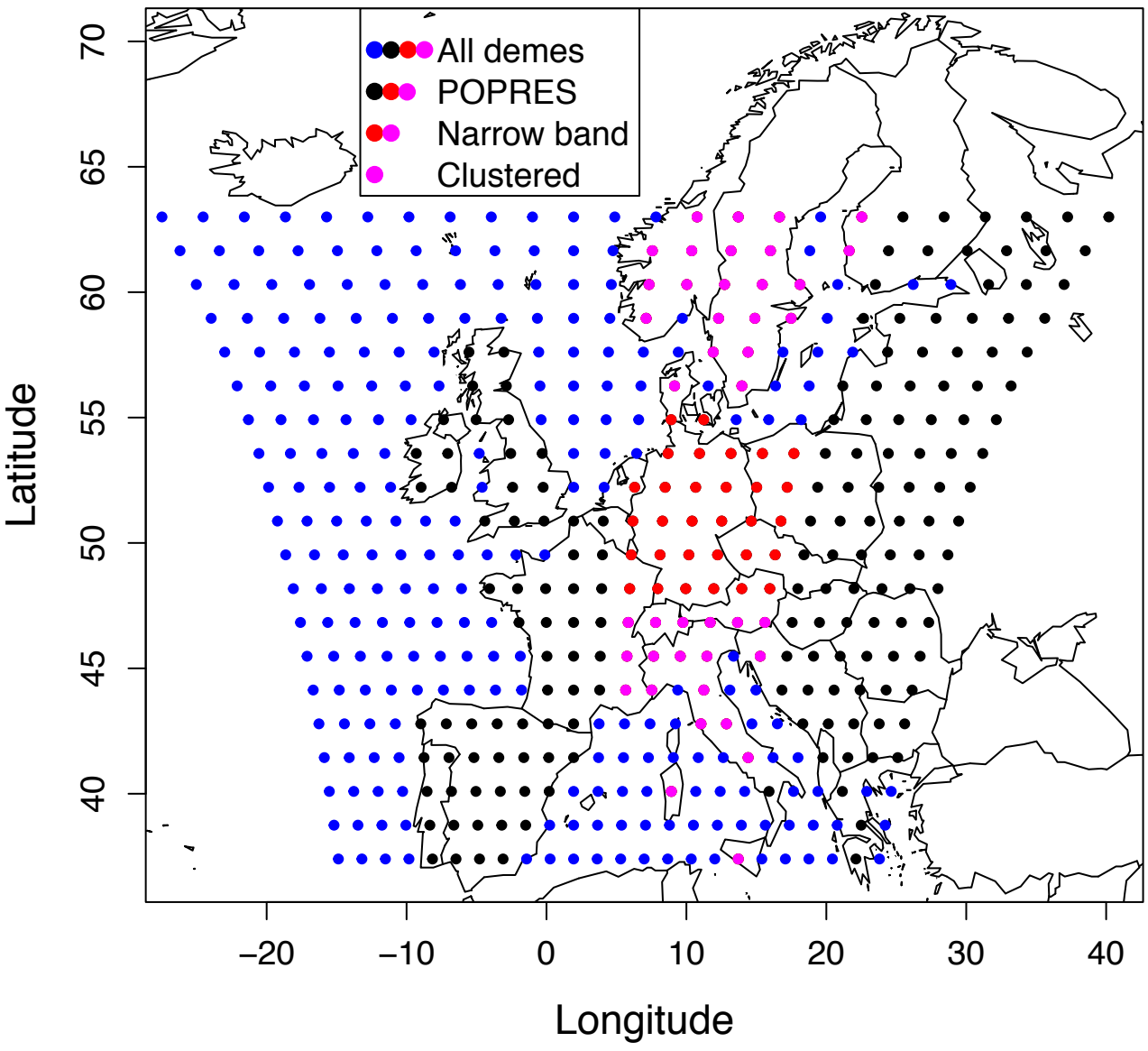

Figure 6: Grid used for the isolation by distance *ms* simulations and the four sampling scheme that was considered to pick $n = 38$ populations.

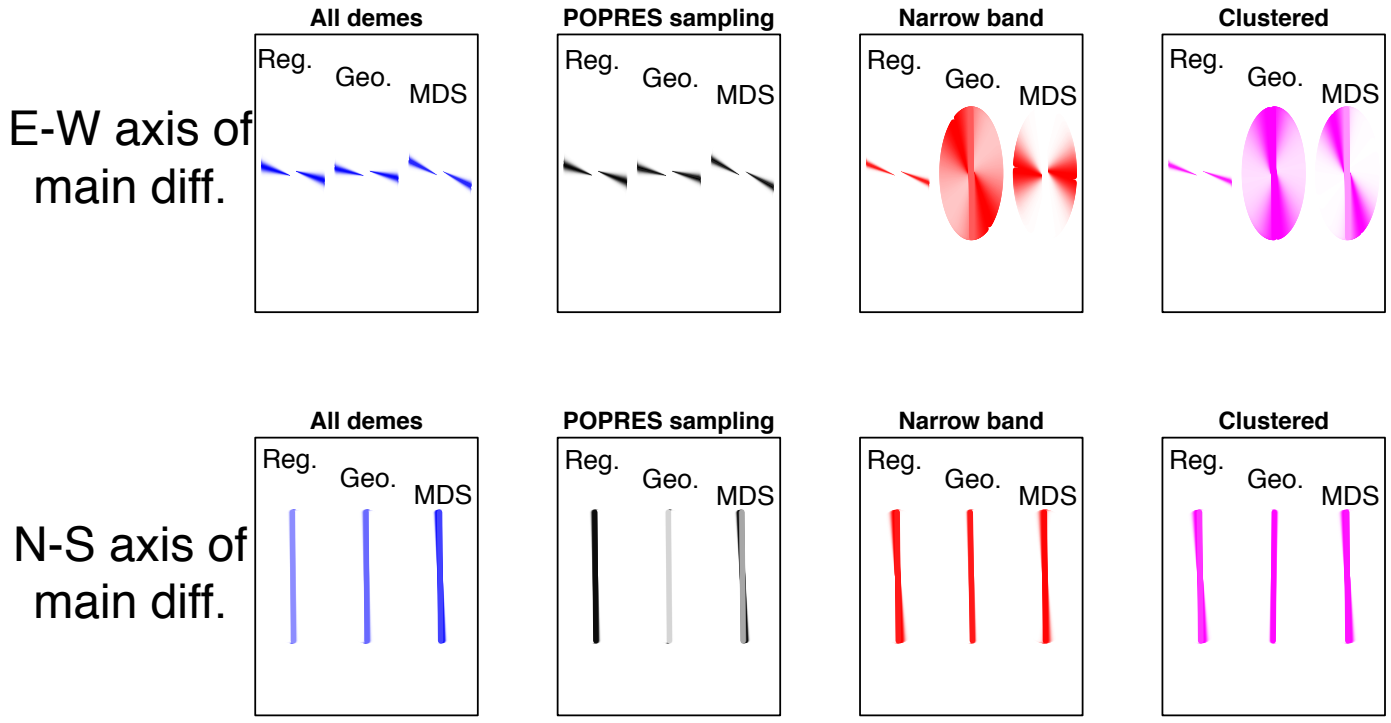

Figure 7: Angular distribution of the angles of maximum differentiation found with the MDS, regression, and geometric methods under anisotropic isolation by distance models. A total of 100 IBD simulations was performed for each scenario and the quadrants show the distribution of the 100 estimated angles of maximum differentiation. Simulations were performed under anisotropic isolation by distance assuming a 20 (N-S) $\times$ 24 (E-W) grid with $4Nm = 1$ under the direction of maximum genetic differentiation and $4Nm = 5$ in the other direction. The grid used for the isolation by distance simulations is shown in Figure S6.

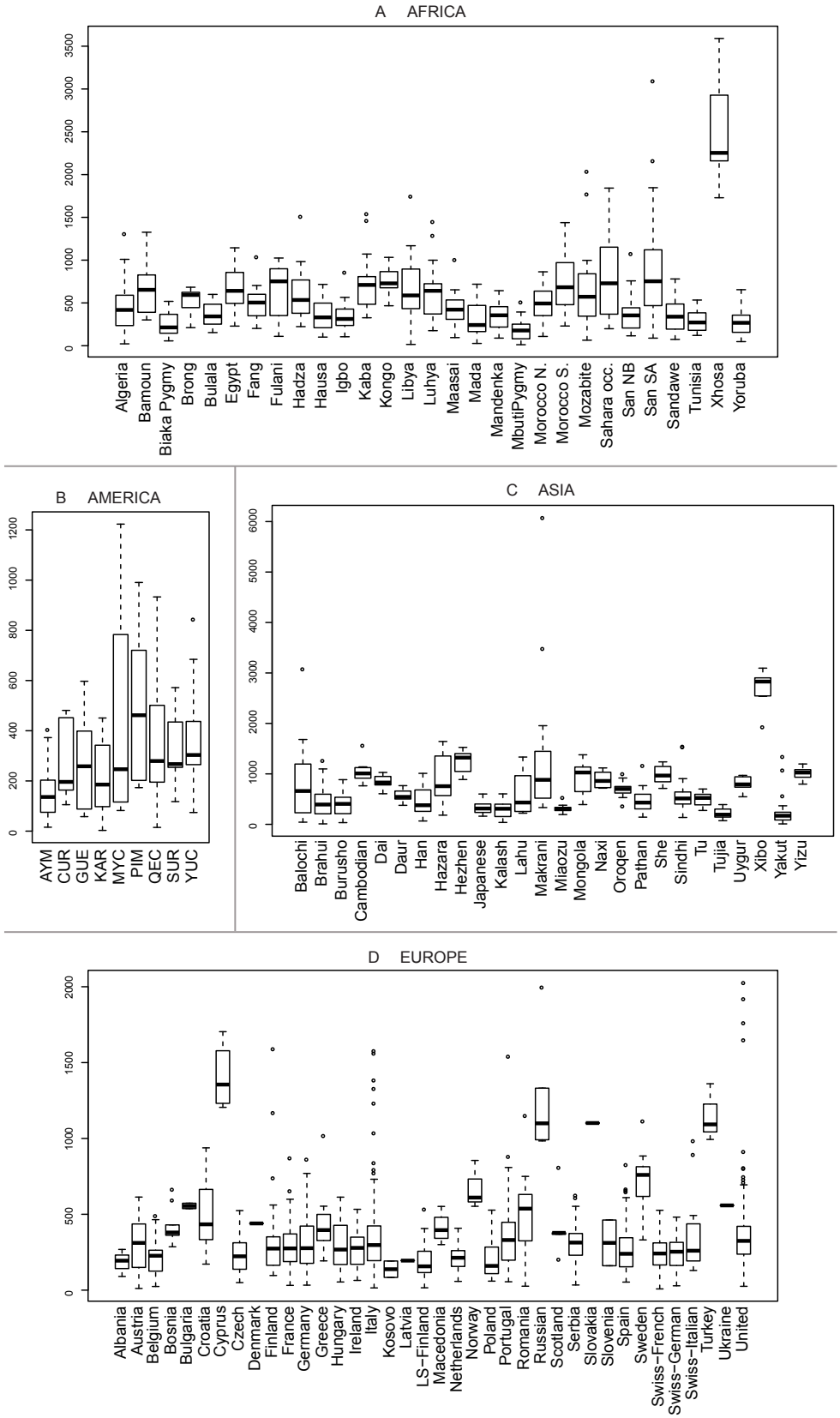

Figure 8: Boxplots of the individual localization errors for each population from the 4 continents The errors were computed using SNP data.

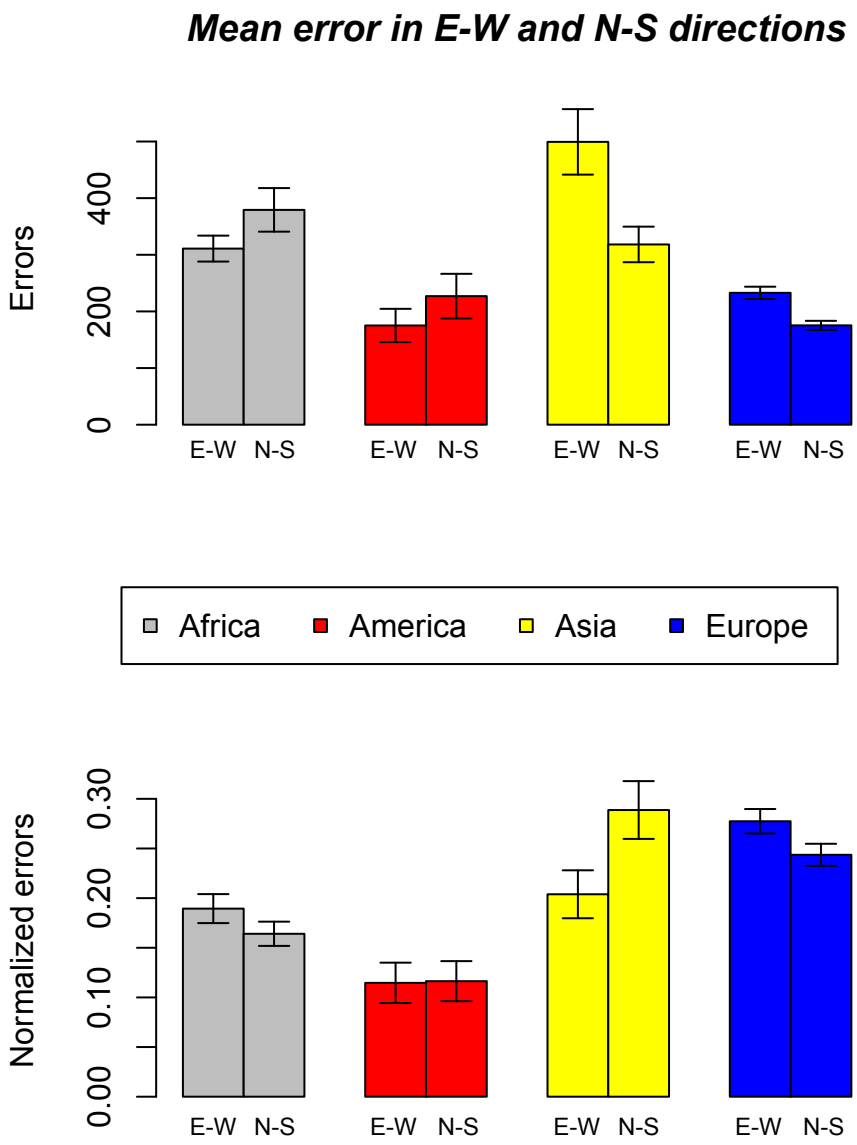

Figure 9: Mean error of the localization method in N-S (lat.) and E-W (long.) direction. Upper panel shows the errors in km and lower panel shows the error after rescaling to account for the intra-continental sampling.

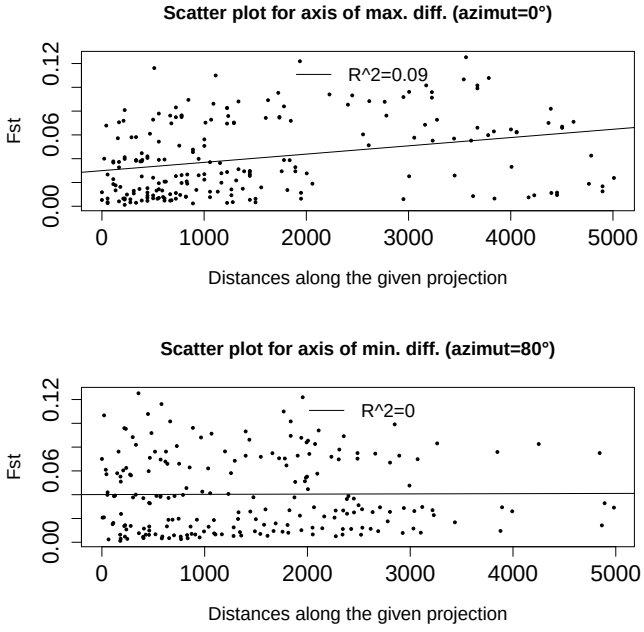

Figure 10: F_{st} of the Sub-Saharan populations as a function of the distance computed along the orientation of maximum and minimum differentiation. Here the orientation of maximum (resp. minimum) differentiation is the orientation that maximizes the correlation between F_{st} and the orientational distances d_θ computed along the different orientations.

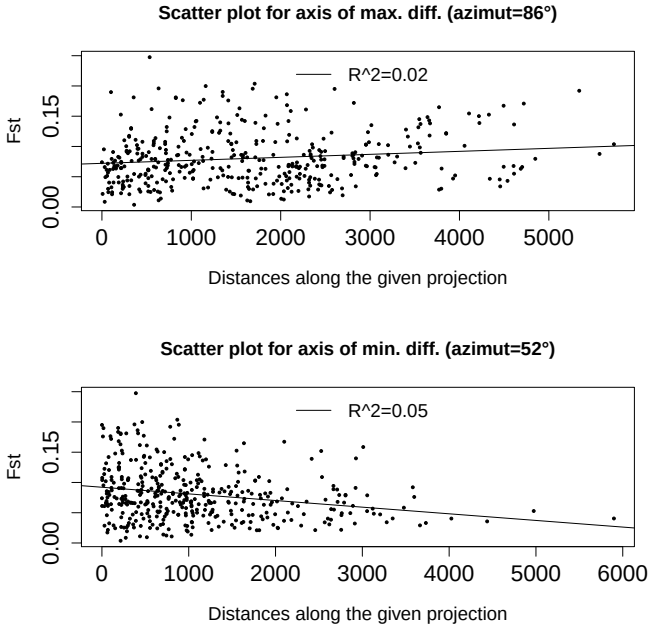

Figure 11: F_{st} of the Native American populations, typed with microsatellites, as a function of the distance computed along the orientation of maximum and minimum differentiation. Here the orientation of maximum (resp. minimum) differentiation is the orientation that maximizes the correlation between F_{st} and the orientational distances d_θ computed along the different orientations.

Continent	Region	Regression			Geometric		
		F_{st}	$\log F_{st}/(1-F_{st})$	$\log F_{st}/(1-F_{st})$ and $\log d$	F_{st}	$\log F_{st}/(1-F_{st})$	$\log F_{st}/(1-F_{st})$ and $\log d$
Africa	All	9	10	0	6	2	177
	Western	9	9	5	7	7	175
	Eastern	46	51	58	54	54	38
	Sub-Sahara	25	35	10	3	173	176
	North Sahara	79	77	75	21	21	21
Asia	All	102	111	97	102	95	90
	Eastern	177	169	0	6	8	3
	Western	66	49	62	86	71	66
America	All	92	69	15	67	67	1
	All (microsat.)	139	133	136	86	85	88
Europe	All	167	161	165	167	161	0
	$n_{\text{indiv}} > 2$[1]	163	161	158	164	161	173
	$n_{\text{indiv}} > 2$, no SEandME[2,3]	15	20	34	12	48	68
	$n_{\text{indiv}} > 2$, no SE[2]	161	166	157	164	168	162
	$n_{\text{indiv}} > 2$, no ME[3]	177	167	170	0	179	5
	$n_{\text{indiv}} > 2$, no Fennoscandia	156	153	161	155	152	122
	$n_{\text{indiv}} > 2$, SEandME[2,3]	129	157	156	137	155	167
	$n_{\text{indiv}} > 2$, Western	154	75	119	173	3	161
	$n_{\text{indiv}} > 2$, Southern	148	162	160	155	152	158
	$n_{\text{indiv}} > 2$, Northern	11	176	129	16	15	136
	$n_{\text{indiv}} > 2$, Middle	152	129	9	164	7	21
	$n_{\text{indiv}} > 2$, Eastern	159	157	155	158	160	153
	$n_{\text{indiv}} > 2$, Central	70	11	26	82	80	96

[1] $n_{\text{indiv}} > 2$ correspond to the 31 populations with a sample size larger than 2
[2] SE stands for Southeast
[3] ME stands for Middle East (Cyprus and Turkey here)

Table 1: Orientations of maximum differentiation obtained when regressing F_{st} on d, $\log F_{st}/(1-F_{st})$ on d, and $\log F_{st}/(1-F_{st})$ on $\log d$.

Continent	Region	Regression[1]	Geometric[1]
Africa	All	8 (177-18)	5 (176-14))
	Western	6 (175-16)	4 (172-16)
	Eastern	37 (1-73)	44(15-74)
	Sub-Sahara	15 (162-49)	10 (153-46)
	North Sahara	77 (11-143)	173 (88-77)
Asia	All	100 (91-110)	101 (93-110)
	Eastern	4 (148-39)	10 (167-32)
	Western	54 (166-121)	66 (171-142)
America	All	97 (47-148)	72 (36-110)
	All (microsat.)	135 (118-152)	90 (77-102)
Europe	All	163 (143-3)	162 (143-2)
	$n_{\text{indiv}} > 2$[2]	159 (144-174)	159 (147-171)
	$n_{\text{indiv}} > 2$, no SEandME[3,4]	22 (168-57)	23 (167-58)
	$n_{\text{indiv}} > 2$, no SE[3]	154 (138-171)	156 (139-173)
	$n_{\text{indiv}} > 2$, no ME[4]	1(160-23)	2 (162-22)
	$n_{\text{indiv}} > 2$, no Fennoscandia	139 (114-164)	139 (114-164)
	$n_{\text{indiv}} > 2$, SEandME[3,4]	120 (64-176)	120 (66-174)
	$n_{\text{indiv}} > 2$, Western	176 (130-43	178 (132-45))
	$n_{\text{indiv}} > 2$, Southern	135 (105-165)	141 (114-168)
	$n_{\text{indiv}} > 2$, Northern	47 (145-128	17 (117-97))
	$n_{\text{indiv}} > 2$, Middle	178 (112-65)	176 (100-73)
	$n_{\text{indiv}} > 2$, Eastern	160 (141-0)	160 (141-177)
	$n_{\text{indiv}} > 2$, Central	91 (22-160)	91 (20-163)

[1] The 95% confidence intervals are given in parenthesis and should be read clockwise

[2] $n_{\text{indiv}} > 2$ correspond to the 31 populations with a sample size larger than 2

[3] SE stands for Southeast

[4] ME stands for Middle East (Cyprus and Turkey here)

Table 2: Median angle of maximal differentiation after having moved the geographical coordinates along a rhumb line of 500 km with angles chosen uniformly between 0° and 360°. A total of 100 orientations of maximum differentiation were calculated and an interval containing 95% of the replicates is provided in parenthesis (should be read clockwise). No transformation of F_{st} or distances were considered.

.

Continent	Region	Pop. number	MDS Angle	Regression Angle[1]	Regression Mantel test[2]	Regression Partial Mantel test[3]	Geometric Angle [1]	Geometric R^2 [4]
Africa	All	29	9	9*** (160-33)	$< 10^{-4}$	$< 10^{-4}$	6 (164-26)	0.4
	Western	18	NA	9*** (166-28)	$< 10^{-4}$	$< 10^{-4}$	7(162-31)	0.74
	Eastern	11	NA	46 (164-111)	0.093	0.371	54 (179-112)	0.20
	Sub-Sahara	21	NA	19 (130-86)	0.091	0.358	0 (103-70)	0.09
	North Sahara	8	NA	79 (19-127)	0.95	0.921	21 (108-74)	0.01
Asia	All	26	109	102*** (84-121)	$< 10^{-4}$	$< 10^{-4}$	102 (80-124)	0.51
	Eastern	16	NA	177 (131-43)	$< 10^{-4}$	0.060	6 (149-41)	0.51
	Western	10	NA	66 (172-145)	0.022	0.400	86 (38-152)	0.25
America	All	9	67	92 (22-167)	0.133	0.33	67 (10-118)	0.13
	All (microsat.)	29	NA	139* (106-163)	0.93	0.05	86 (54-123)	0.02
Europe	All	38	3	167* (140-14)	$< 10^{-4}$	0.012	167 (138-14)	0.26
	$n_{\text{indiv}} > 2$[5]	30	NA	163*** (137-9)	$< 10^{-4}$	2×10^{-4}	164 (137-9)	0.53
	$n_{\text{indiv}} > 2$, no SEandME[6,7]	21	NA	15 (139-71)	$< 10^{-4}$	0.178	12 (146-71)	0.34
	$n_{\text{indiv}} > 2$, no SE[6]	23	NA	161*** (127-18)	$< 10^{-4}$	5×10^{-4}	164 (125-23)	0.56
	$n_{\text{indiv}} > 2$, no ME[7]	28	NA	177** (145-25)	$< 10^{-4}$	0.006	0 (147-28)	0.38
	$n_{\text{indiv}} > 2$, no Fennoscandia		NA	156** (130-7)	$< 10^{-4}$	$< 10^{-4}$	155 (127-2)	0.62
	$n_{\text{indiv}} > 2$, SEandME[6,7]	9	NA	129 (69-39)	4×10^{-4}	0.798	137 (86-9)	0.82
	$n_{\text{indiv}} > 2$, Western	8	NA	154 (100-65)	0.002	0.677	173 (121-60)	0.61
	$n_{\text{indiv}} > 2$, Southern	13	NA	148** (119-5)	0.011	0.007	155 (129-2)	0.61
	$n_{\text{indiv}} > 2$, Northern	8	NA	11 (117-81)	0.34	0.802	16 (105-73)	0.14
	$n_{\text{indiv}} > 2$, Middle	8	NA	152 (100-65)	0.023	0.675	164 (84-64)	0.37
	$n_{\text{indiv}} > 2$, Eastern	14	NA	159*** (137-3)	$< 10^{-4}$	$< 10^{-4}$	158 (142-171)	0.88
	$n_{\text{indiv}} > 2$, Central	9	NA	70(165-150)	$< 10^{-4}$	0.937	82 (26-159)	0.50

[1] The 95% confidence intervals are given in parenthesis and should be read clockwise
[2] The Mantel test assesses whether there is an effect of the geographic distances on F_{st}.
[3] The partial Mantel test assesses whether there is an additional effect of the bearings between two populations when regressing F_{st}.
[4] The coefficient R^2 is obtained when regressing F_{st} by the distance computed along the direction of maximum differentiation
[5] $n_{\text{indiv}} > 2$ correspond to the 31 populations with a sample size larger than 2
[6] SE stands for Southeast
[7] ME stands for Middle East (Cyprus and Turkey here)

Table 3: Orientations of maximum differentiation obtained with the different methods. A total of $10,000$ permutations was performed to provide the P values.

Continent	Number of PC	Error	E-W error	N-S error	Rel. error	Rel. E-W error	Rel. N-S error
Africa	2	1180	910	480	0.38	0.62	0.27
Africa	$K_{\text{opt}} = 52$	430	250	260	0.15	0.14	0.12
America	2	470	330	240	0.19	0.23	0.12
America	$K_{\text{opt}} = 17$	250	140	180	0.10	0.09	0.10
Asia	2	1090	390	520	0.39	0.16	0.54
Asia	$K_{\text{opt}} = 34$	510	350	200	0.17	0.14	0.19
Europe	2	470	380	230	0.41	0.47	0.33
Europe	$K_{\text{opt}} = 17$	280	190	140	0.24	0.23	0.19

Table 4: Median errors (in kilometers) of geographic localization based on SNP and microsatellite data. The relative errors are computed with respect to a naive localizer which assigns each individual to a population that is chosen at random among the sampled populations.

	n	All pops	BigPops	North	Central	South	West	Mid	East	no SEandCT	no ME	no SE	no Fennoscandia	SEandCT
Swiss-German	84	x	x		x			x		x	x	x	x	
Germany	71	x	x		x			x		x	x	x	x	
Netherlands	17	x	x		x			x		x	x	x	x	
Austria	14	x	x		x			x		x	x	x	x	
Hungary	19	x	x		x				x	x	x	x	x	
Czech	11	x	x		x				x	x	x	x	x	
Slovakia	1	x												
Cyprus	4	x	x			x			x			x	x	x
Turkey	4	x	x			x			x			x	x	x
Sweden	10	x	x	x				x		x	x	x		
Norway	3	x	x	x				x		x	x	x		
Denmark	1	x												
Poland	22	x	x	x					x	x	x	x	x	
Russian	6	x	x	x					x	x	x	x	x	
Finland	41	x	x	x					x	x	x	x		
Latvia	1	x												
Ukraine	1	x												
LS-Finland	41	x												
United	200	x	x	x			x			x	x	x	x	
Ireland	61	x	x	x			x			x	x	x	x	
Scotland	5	x	x	x			x			x	x	x	x	
Italy	219	x	x			x		x		x	x	x	x	
Swiss-Italian	13	x	x			x		x		x	x	x	x	
Serbia	44	x	x			x			x		x		x	x
Romania	14	x	x			x			x		x		x	x
Bosnia	9	x	x			x			x		x		x	x
Croatia	8	x	x			x			x		x		x	x
Greece	8	x	x			x			x		x		x	x
Macedonia	4	x	x			x			x		x		x	x
Albania	3	x	x			x			x		x		x	x
Bulgaria	2	x												
Kosovo	2	x												
Slovenia	2	x												
Spain	136	x	x			x	x			x	x	x	x	
Portugal	128	x	x			x	x			x	x	x	x	
Swiss-French	125	x	x		x		x			x	x	x	x	
France	91	x	x		x		x			x	x	x	x	
Belgium	43	x	x		x		x			x	x	x	x	

Table 5: The different subdivisions of Europe.

	all	West	East	North Sahara	SubSaharan
Morocco N	x	x		x	
Algeria	x	x		x	
Tunisia	x	x		x	
Mozabite	x	x		x	
Libya	x	x		x	
Morocco S	x	x		x	
Egypt	x		x	x	
Sahara OCC	x	x		x	
Bulala	x		x		x
Hausa	x	x			x
Mandenka	x	x			x
Fulani	x	x			x
Mada	x	x			x
Kaba	x	x			x
Yoruba	x	x			x
Brongx	x			x	
Igbo	x	x			x
Bamoun	x	x			x
Biaka Pygmy	x		x		x
Fang	x	x			x
Mbuti Pygmy	x		x		x
Luhya	x		x		x
Maasai	x		x		x
Hadza	x		x		x
Sandawe	x		x		x
Kongo	x	x			x
San NB	x		x		x
San SA	x		x		x
Xhosa	x		x		x

Table 6: The different subdivisions of Africa.

	All	West Asia	East Asia
Makrani	x	x	
Balochi	x	x	
Brahui	x	x	
Sindhi	x	x	
Hazara	x	x	
Pathan	x	x	
Kalash	x	x	
Burusho	x	x	
Uygur	x	x	
Xibo	x	x	
Naxi	x		x
Dai	x		x
Lahu	x		x
Tu	x		x
Yizu	x		x
Cambodian	x		x
Miaozu	x		x
Tujia	x		x
Han	x		x
She	x		x
Mongola	x		x
Daur	x		x
Oroqen	x		x
Yakut	x		x
Hezhen	x		x
Japanese	x		x

Table 7: The different subdivisions of Asia.

.