Supplementary Materials for
Identification of genomic regions distorting population structure inference in diverse continental groups

Qiuxuan Liu, Degang Wu, Chaolong Wang

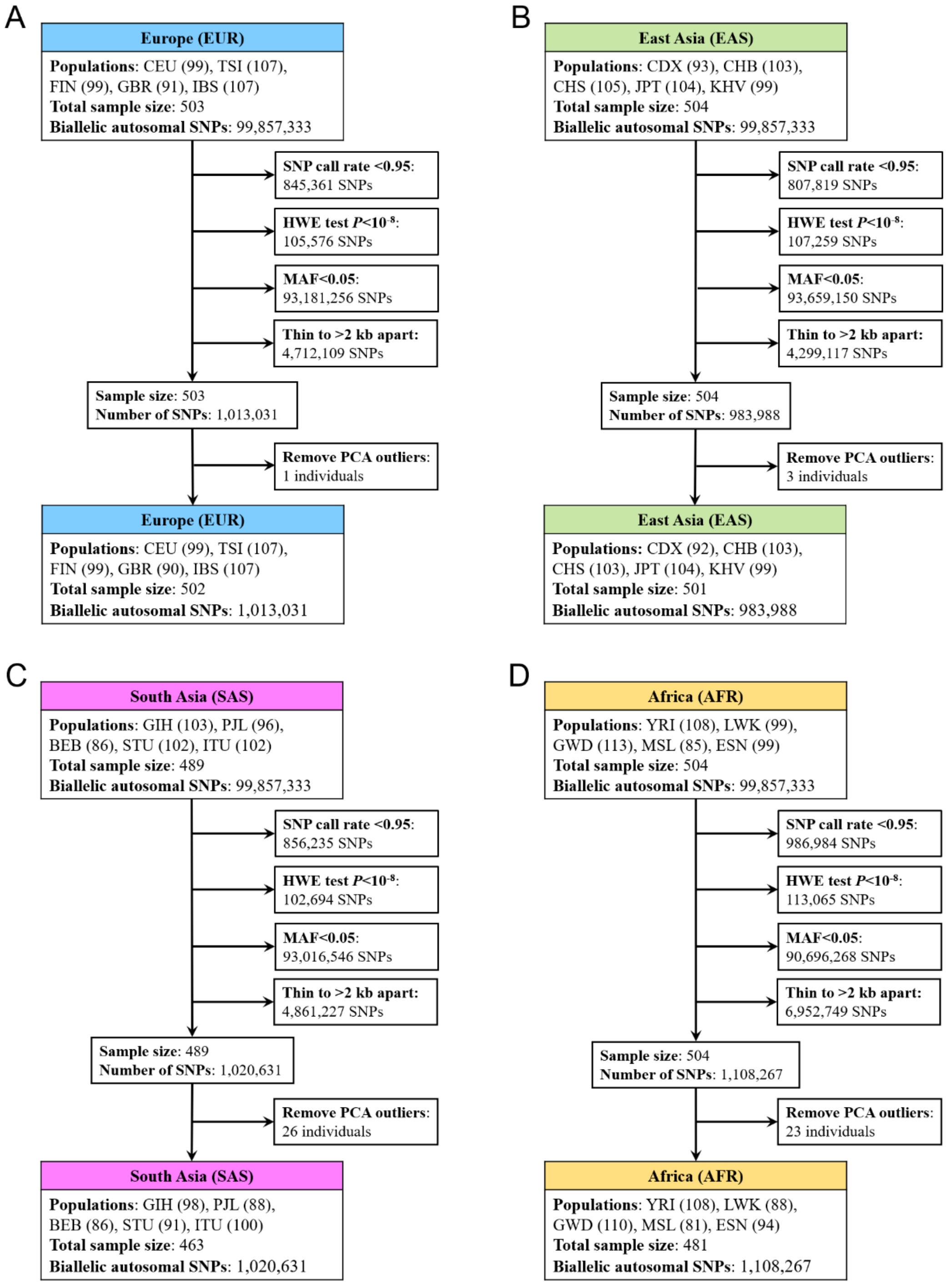

Figure S1. Flowchart of quality controls for each continental group. (A) Europe. (B) East Asia. (C) South Asia. (D) Africa.

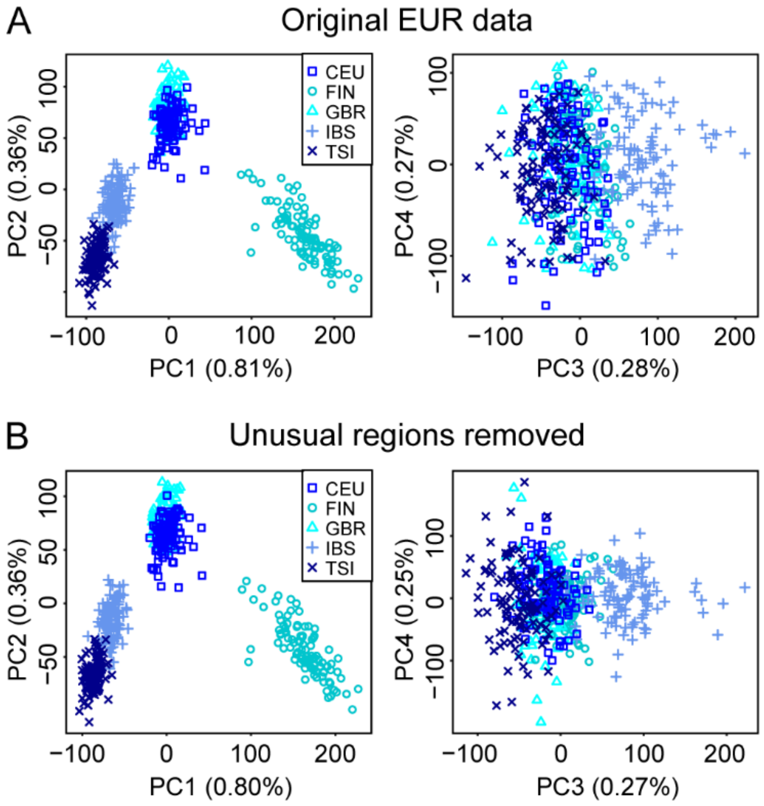

Figure S2. Top 4 PCs of Europeans before and after excluding unusual regions. (A) PC1-PC4 based on the original European data. (B) PC1-PC4 based on the European data with unusual regions removed.

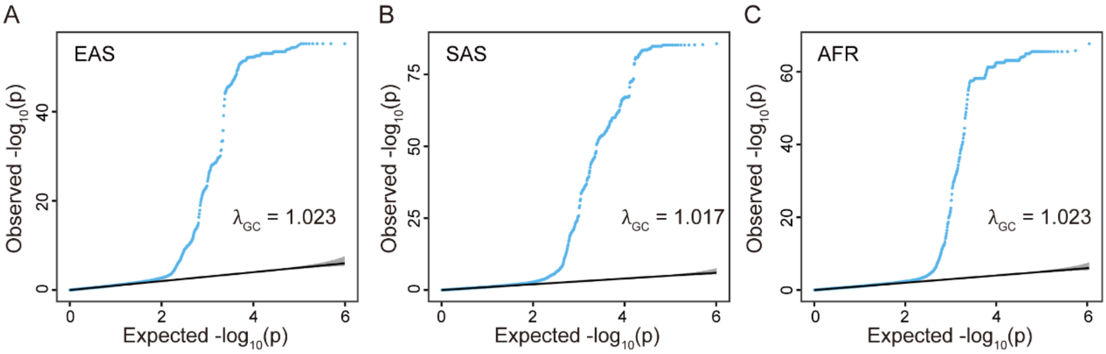

Figure S3. QQ plots of combined signals across significant PCs after iteration 1.

(A) East Asian data. (B) South Asian data. (C) African data.

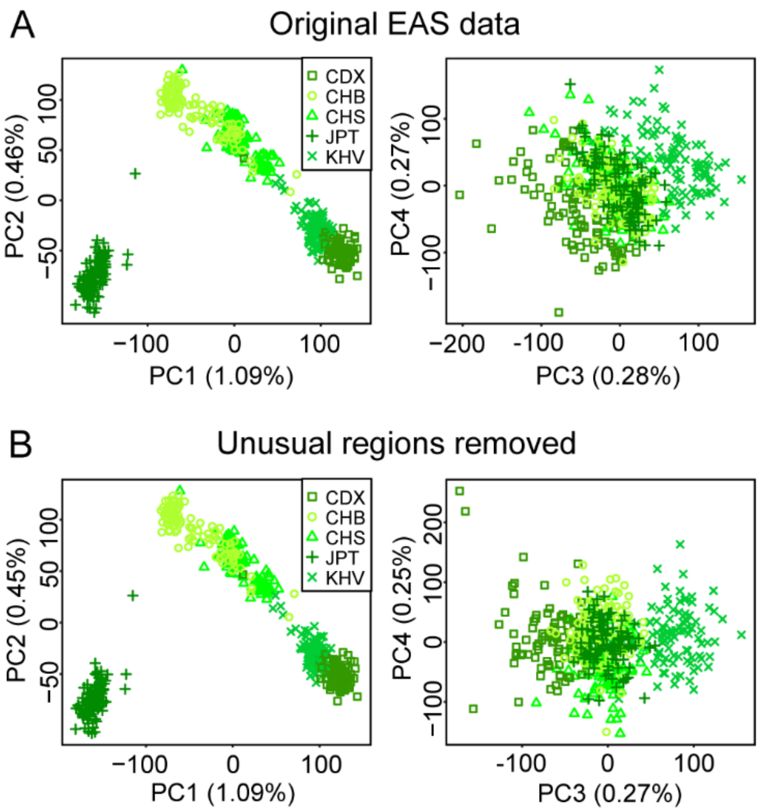

Figure S4. Top 4 PCs of East Asians before and after excluding unusual regions. (A) PC1-PC4 based on the original East Asian data. (B) PC1-PC4 based on the East Asian data with unusual regions removed.

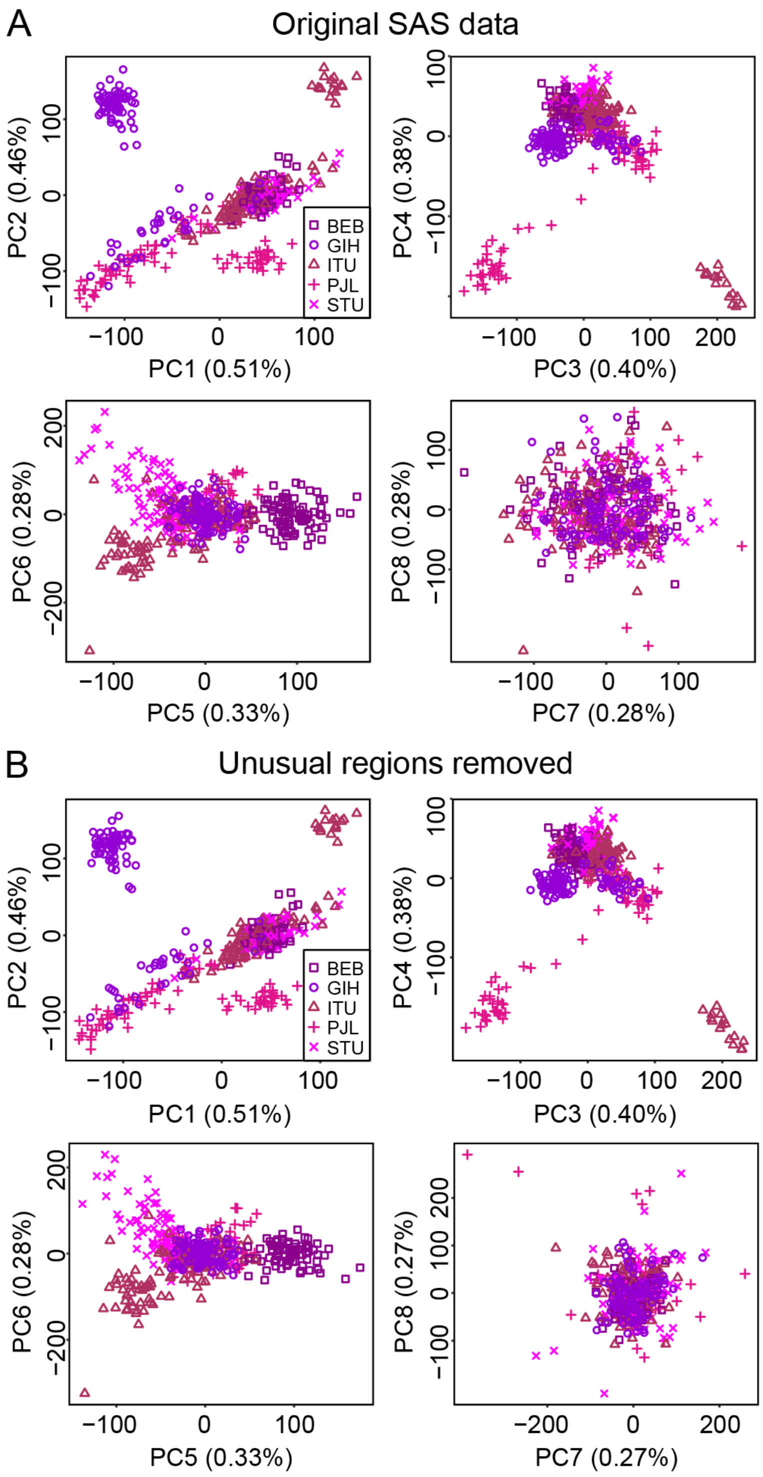

Figure S5. Top 8 PCs of South Asians before and after excluding unusual regions. (A) PC1-PC8 based on the original South Asian data. (B) PC1-PC8 based on the South Asian data with unusual regions removed.

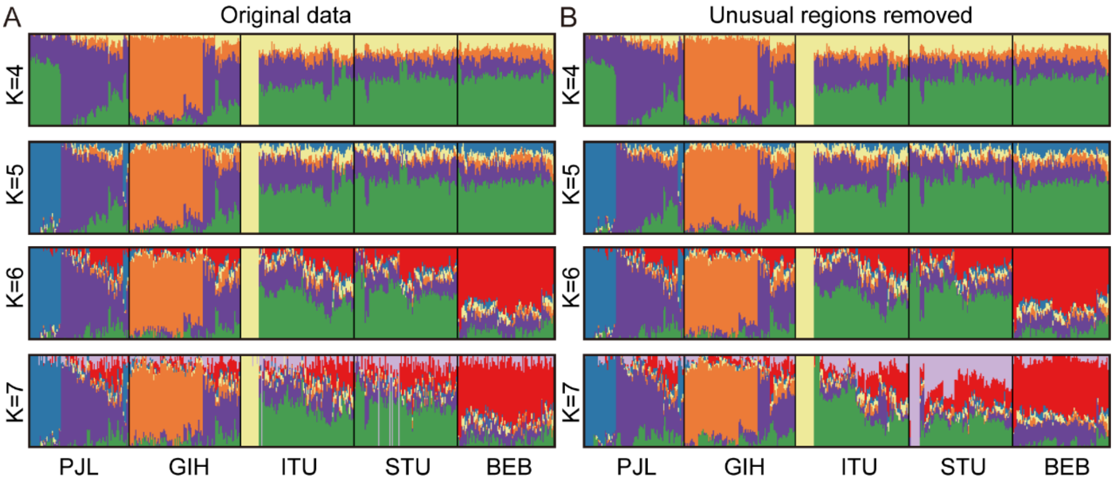

Figure S6. Impacts of unusual genomic regions on the unsupervised ADMIXTURE analyses of South Asian data. (A) Results based on original data. (B) Results based on data excluding unusual regions. We assumed *K* = 4, 5, 6, or 7 ancestral components, indicated by colors, in each analysis. Each vertical bar represents one individual, and the colored segments represent proportions of ancestral components.

Orders of individuals in each panel are the same and the order was determined by hierarchical clustering on the ancestral proportions in panel B (*K* = 7).

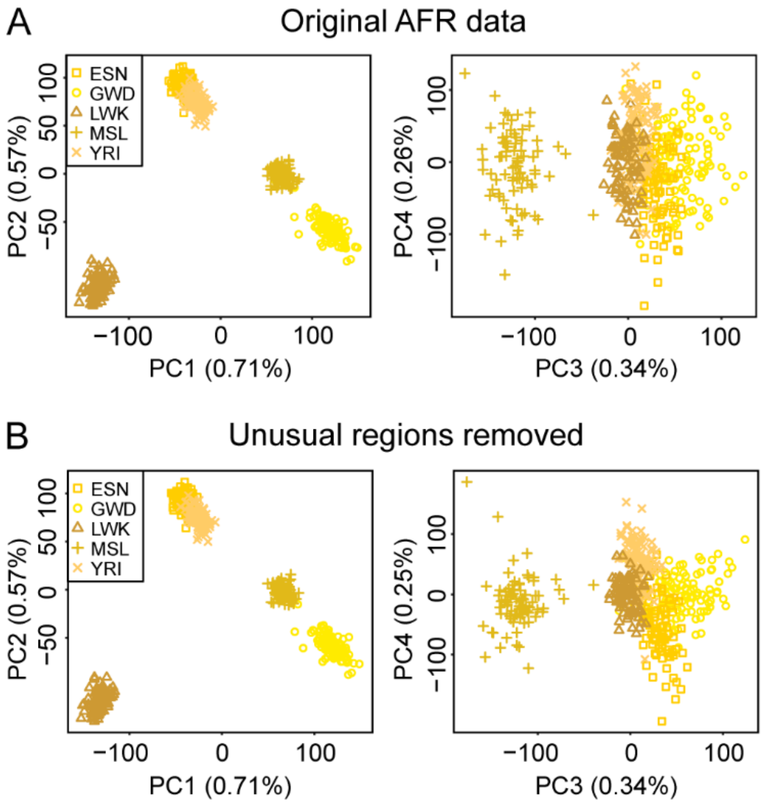

Figure S7. Top 4 PCs of Africans before and after excluding unusual regions. (A) PC1-PC4 based on the original African data. (B) PC1-PC4 based on the African data with unusual regions removed.

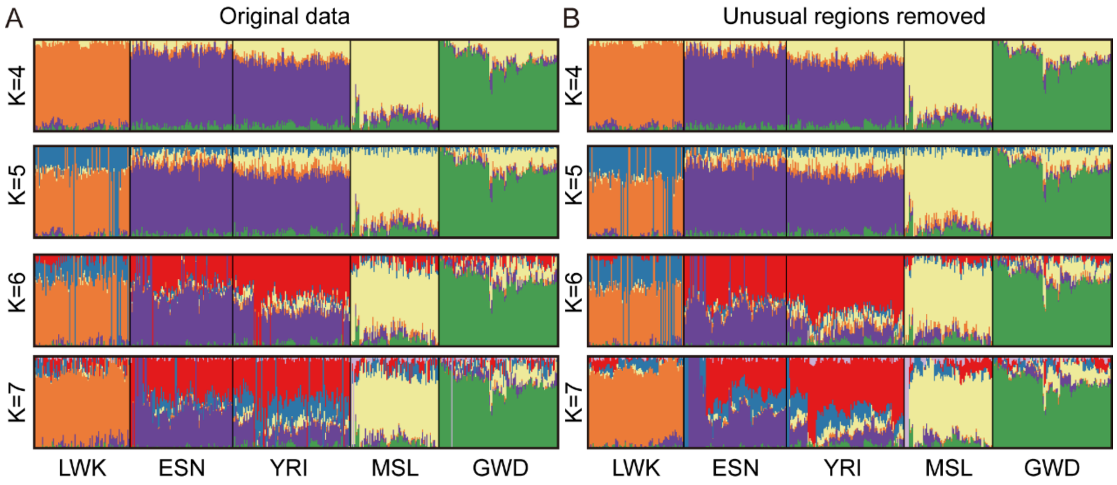

Figure S8. Impacts of unusual genomic regions on the unsupervised ADMIXTURE analyses of African data. (A) Results based on original data. (B) Results based on data excluding unusual regions. We assumed *K* = 4, 5, 6, or 7 ancestral components, indicated by colors, in each analysis. Each vertical bar represents one individual, and the colored segments represent proportions of ancestral components.

Orders of individuals in each panel are the same and the order was determined by hierarchical clustering on the ancestral proportions in panel B (*K* = 7).

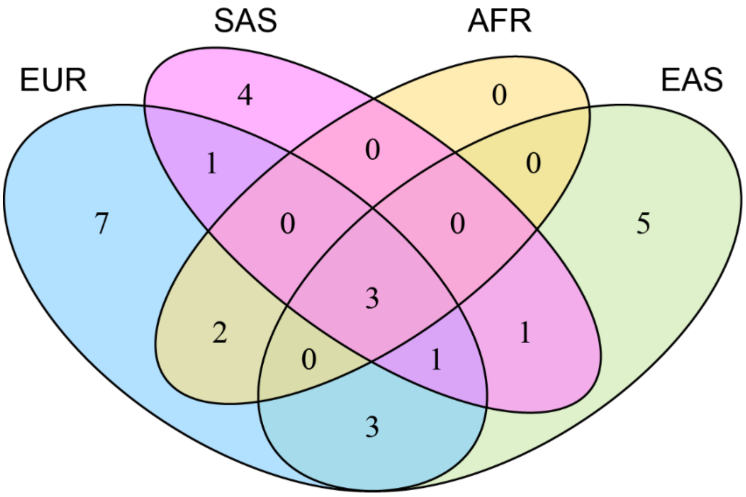

Figure S9. Venn diagram indicating the number of common and exclusive unusual genomic regions identified in four continental groups.

Table S1. List of loci reported to be under natural selection.

Position (Mb)	Candidate genes	Selected trait	Continental group	References
chr6: 25.71-31.58	*HLA* group	Immune response	EUR, EAS	33
chr2: 134.54-136.27	*LCT*	Lactase persistence	EUR	24
chr3: 50.27-51.72	*DOCK3, MAPKAPK3, CISH*	Height	EUR, AFR	36
chr14: 66.17-67.42	*GPHN*	Neural development	EAS	37
chr11: 61.62-61.90	*FADS1, FADS2*	Biosynthesis of LC-PUFA*	EAS	38; 39
chr14: 105.51-105.76	*IGH* group	Immune response	EAS	34
chr15: 28.09-28.32	*OCA2, HERC2*	Eye and skin pigmentation	EUR	28
chr15: 48.10-48.22	*MYEF2, SLC24A5*	Skin pigmentation	SAS	35

* LC-PUFA, long chain polyunsaturated fatty acids.

Table S2. Information of samples and populations included in the analysis.

Continent	Population name	Abbreviation	Final sample size	Number of outliers	Excluded outlier samples
Europe (EUR)	Toscani	TSI	107	0	
	Iberian	IBS	107	0	
	British	GBR	90	1	HG00120
	CEPH	CEU	99	0	
	Finnish	FIN	99	0	
East Asia (EAS)	Dai Chinese	CDX	92	1	HG02380
	Kinh Vietnamese	KHV	99	0	
	Southern Han Chinese	CHS	103	2	HG00475, HG00542
	Han Chinese	CHB	103	0	
	Japanese	JPT	104	0	
South Asia (SAS)	Punjabi	PJL	88	8	HG02648, HG02657, HG02658, HG02684, HG02690, HG02691, HG03228, HG03229
	Gujarati	GIH	98	5	NA20882, NA20891, NA20900, NA21109, NA21135
	Telugu	ITU	100	2	HG03873, HG04070
	Tamil	STU	91	11	HG03692, HG03696, HG03733, HG03750, HG03754, HG03837, HG03896, HG03898, HG03899, HG03955, HG03998
	Bengali	BEB	86	0	
Africa (AFR)	Luhya	LWK	88	11	NA19025, NA19027, NA19042, NA19307, NA19312, NA19331, NA19334, NA19376, NA19384, NA19451, NA19452
	Esan	ESN	94	5	HG03301, HG03343, HG03352, HG03366, HG03372
	Yoruba	YRI	108	0	
	Mende	MSL	81	4	HG03428, HG03464, HG03478, HG03484
	Gambian Mandinka	GWD	110	3	HG02610, HG02624, HG03259