

Supplementary materials

Trajectory of human migration: Insights from autosomal and non-autosomal variant clustering patterns

Samayeta Sarkar Tuli^{1, †}, Joyatry Sarker^{1, †}, Mrinmoy Saha Roddur^{2,3}, Anik Biswas¹, Reefa Nawar¹, Tahmina Akter¹, Md. Wahid Murad¹, Abu Ashfaqur Sajib^{1,*}

Author affiliations:

¹Department of Genetic Engineering and Biotechnology, University of Dhaka, Dhaka, Bangladesh

²Department of Computer Science and Engineering, Southeast University, Dhaka, Bangladesh

³University of Illinois Urbana-Champaign, Urbana, IL-61820, USA

[†]These authors contributed equally.

***Corresponding author**

Abu Ashfaqur Sajib

Email: abu.sajib@du.ac.bd

Postal address: Department of Genetic Engineering & Biotechnology, University of Dhaka, Dhaka 1000, Bangladesh.

Supplementary Table 1. List of populations with their geographic distribution

Superpopulation (Continent)	Population	Size	Geographic description
African (AFR)	ACB	96	African Caribbean in Barbados
	ASW	61	African Ancestry in Southwest US
	ESN	99	Esan in Nigeria
	GWD	113	Gambian in Western Division, The Gambia
	LWK	99	Luhya in Webuye, Kenya
	MSL	85	Mende in Sierra Leone
	YRI	108	Yoruba in Ibadan, Nigeria
East Asian (EAS)	CDX	93	Chinese Dai in Xishuangbanna, China
	CHB	103	Han Chinese in Beijing, China
	CHS	105	Southern Han Chinese, China
	JPT	104	Japanese in Tokyo, Japan
	KHV	99	Kinh in Ho Chi Minh City, Vietnam
European (EUR)	CEU	99	Utah residents with Northern and Western European ancestry
	FIN	99	Finnish in Finland
	GBR	91	British in England and Scotland
	IBS	107	Iberian populations in Spain
	TSI	107	Toscani in Italy
South Asian (SAS)	BEB	86	Bengali in Bangladesh
	GIH	103	Gujarati Indian in Houston, TX
	ITU	102	Indian Telugu in the UK
	PJL	96	Punjabi in Lahore, Pakistan
	STU	102	Sri Lankan Tamil in the UK

Supplementary Table 2. Python code for rearranging the columns of the data files

```
import sys
import numpy
import pandas
import re
ref_col = ['#CHROM', 'POS', 'ID', 'REF', 'ALT', 'YRI_TOTAL_CNT', 'YRI_ALT_CNT', 'YRI_FRQ', 'LWK_TOTAL_CNT',
'LWK_ALT_CNT', 'LWK_FRQ', 'GWD_TOTAL_CNT', 'GWD_ALT_CNT', 'GWD_FRQ', 'MSL_TOTAL_CNT', 'MSL_ALT_CNT',
'MSL_FRQ', 'ESN_TOTAL_CNT', 'ESN_ALT_CNT', 'ESN_FRQ', 'ASW_TOTAL_CNT', 'ASW_ALT_CNT', 'ASW_FRQ',
'ACB_TOTAL_CNT', 'ACB_ALT_CNT', 'ACB_FRQ', 'MXL_TOTAL_CNT', 'MXL_ALT_CNT', 'MXL_FRQ', 'PUR_TOTAL_CNT',
'PUR_ALT_CNT', 'PUR_FRQ', 'CLM_TOTAL_CNT', 'CLM_ALT_CNT', 'CLM_FRQ', 'PEL_TOTAL_CNT', 'PEL_ALT_CNT',
'PEL_FRQ', 'CHB_TOTAL_CNT', 'CHB_ALT_CNT', 'CHB_FRQ', 'JPT_TOTAL_CNT', 'JPT_ALT_CNT', 'JPT_FRQ',
'CHS_TOTAL_CNT', 'CHS_ALT_CNT', 'CHS_FRQ', 'CDX_TOTAL_CNT', 'CDX_ALT_CNT', 'CDX_FRQ', 'KHV_TOTAL_CNT',
'KHV_ALT_CNT', 'KHV_FRQ', 'CEU_TOTAL_CNT', 'CEU_ALT_CNT', 'CEU_FRQ', 'TSI_TOTAL_CNT', 'TSI_ALT_CNT', 'TSI_FRQ',
'FIN_TOTAL_CNT', 'FIN_ALT_CNT', 'FIN_FRQ', 'GBR_TOTAL_CNT', 'GBR_ALT_CNT', 'GBR_FRQ', 'IBS_TOTAL_CNT',
'IBS_ALT_CNT', 'IBS_FRQ', 'GIH_TOTAL_CNT', 'GIH_ALT_CNT', 'GIH_FRQ', 'PJL_TOTAL_CNT', 'PJL_ALT_CNT', 'PJL_FRQ',
'BEB_TOTAL_CNT', 'BEB_ALT_CNT', 'BEB_FRQ', 'STU_TOTAL_CNT', 'STU_ALT_CNT', 'STU_FRQ', 'ITU_TOTAL_CNT',
'ITU_ALT_CNT', 'ITU_FRQ', 'ALL_TOTAL_CNT', 'ALL_ALT_CNT', 'ALL_FRQ']
#Only for TSV
data = pandas.read_csv(sys.argv[1], sep='\t', index_col=False)
cols = data.columns.tolist()
new_col = []
for i in ref_col:
    if i in cols:
        new_col.append(i)
data = data[new_col]
frq_check = [i for i in cols if re.search(r".+FRQ",i)]
for i in frq_check:
    data[i] = pandas.to_numeric(data[i], errors='coerce')
output = data[data.all(axis='columns') == False]
output.to_csv(sys.argv[2], sep = '\t', index=None)
```

Supplementary Table 3. An example of the data format is provided ([Excel file](#)).

Supplementary Table 4. Chromosomal position of centromeres.

Chromosome #	Start, bp	End, bp
Chr1	121535434	124535434
Chr2	92326171	95326171
Chr3	90504854	93504854
Chr4	49660117	52660117
Chr5	46405641	49405641
Chr6	58830166	61830166
Chr7	58054331	61054331
Chr8	43838887	46838887
Chr9	47367679	50367679
Chr10	39254935	42254935
Chr11	51644205	54644205
Chr12	34856694	37856694
Chr13	16000000	19000000
Chr14	16000000	19000000
Chr15	17000000	20000000
Chr16	35335801	38335801

Chr17	22263006	25263006
Chr18	15460898	18460898
Chr19	24681782	27681782
Chr20	26369569	29369569
Chr21	11288129	14288129
Chr22	13000000	16000000
ChrX	58632012	61632012
ChrY	10104553	13104553
