HGGA, Volume *5*

Supplemental information

Distinct positions of genetic and oral histories: Perspectives from India

Arjun Biddanda, Esha Bandyopadhyay, Constanza de la Fuente Castro, David Witonsky, Jose A. Urban Aragon, Nagarjuna Pasupuleti, Hannah M. Moots, Renée Fonseca, Suzanne Freilich, Jovan Stanisavic, Tabitha Willis, Anoushka Menon, Mohammed S. Mustak, Chinnappa Dilip Kodira, Anjaparavanda P. Naren, Mithun Sikdar, Niraj Rai, and Maanasa Raghavan

Figure S1: *f_4* ratio model. Representation of the model used for estimating the relative proportions of ANI- and ASI-related genetic ancestries in the Southwest Indian populations.

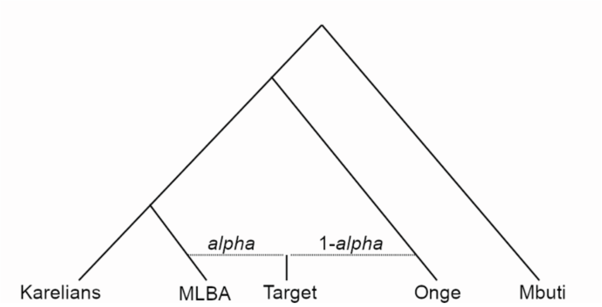


Figure S2: *D*-statistics for the f_4 ratio test. *D*-statistics to evaluate the closest present-day population to Central Steppe MLBA (ANI proxy) for estimating α.

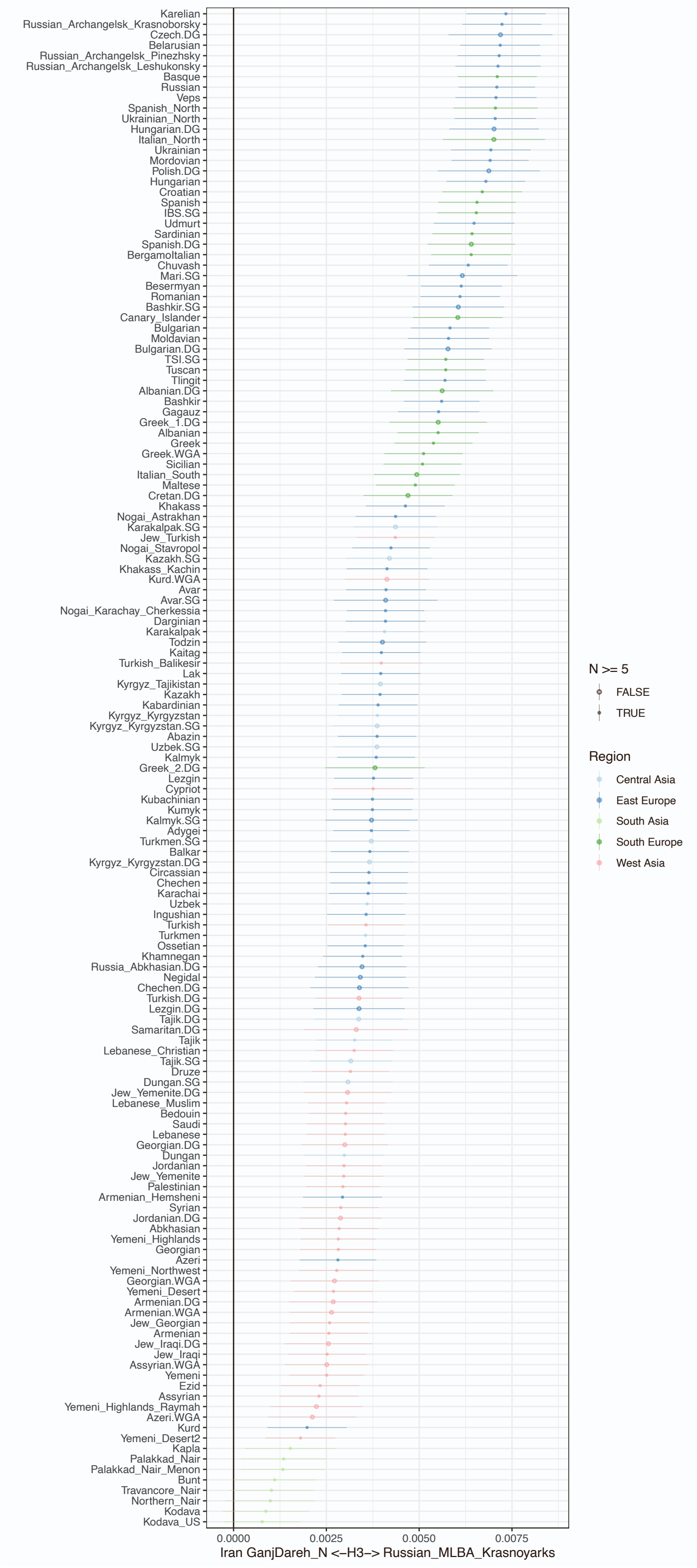


Figure S3: Pairwise qpWave analysis to test for cladality of source populations with respect to the outgroups. This test was run using ADMIXTOOLS2 (*maxmiss* 0.2) and implemented to evaluate if outgroups were able to differentiate source populations. Outgroup populations contribute to the differentiation of the pair when p-values are lower than 0.05.

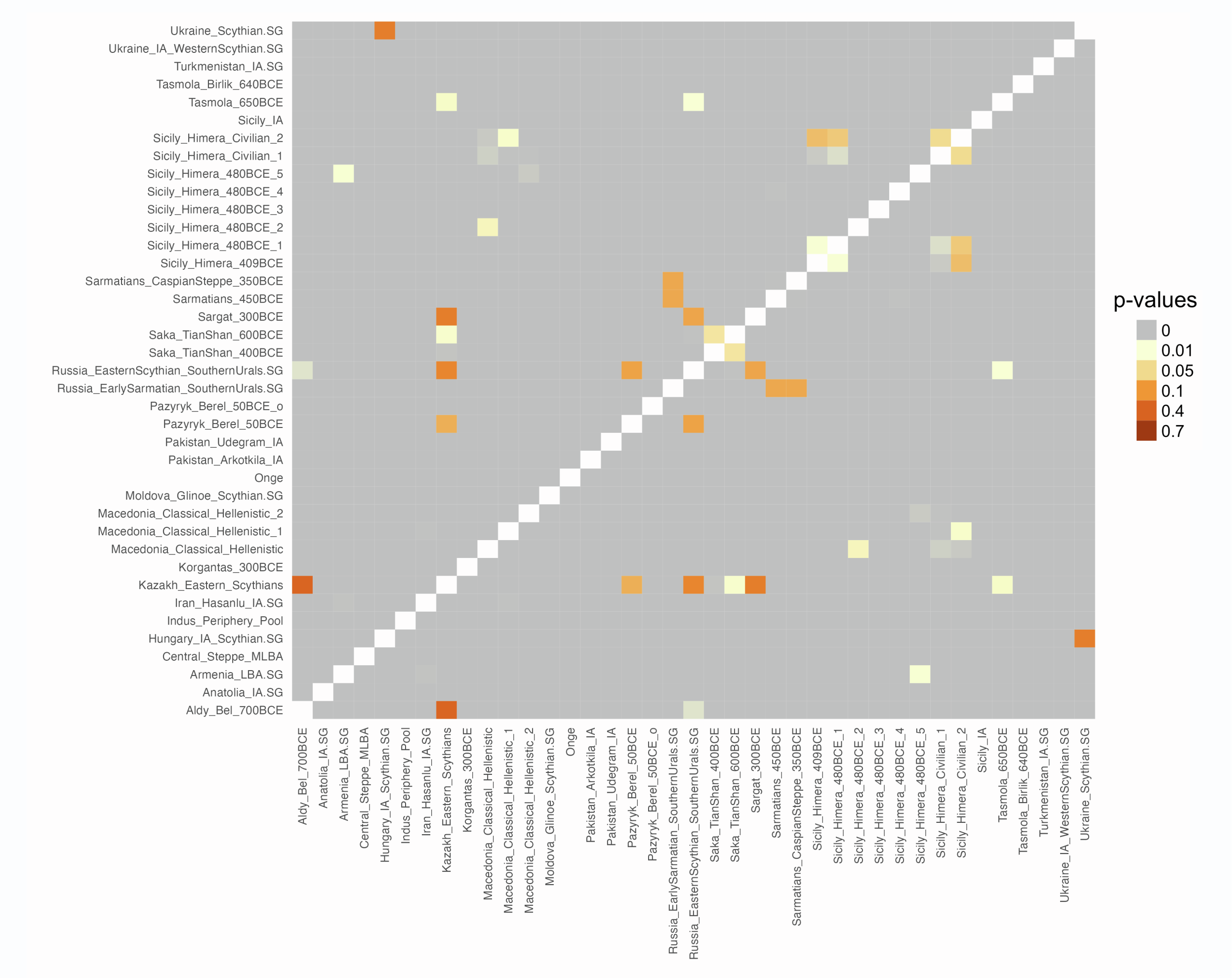


Figure S4: PCA with pseudo-haploid calls.

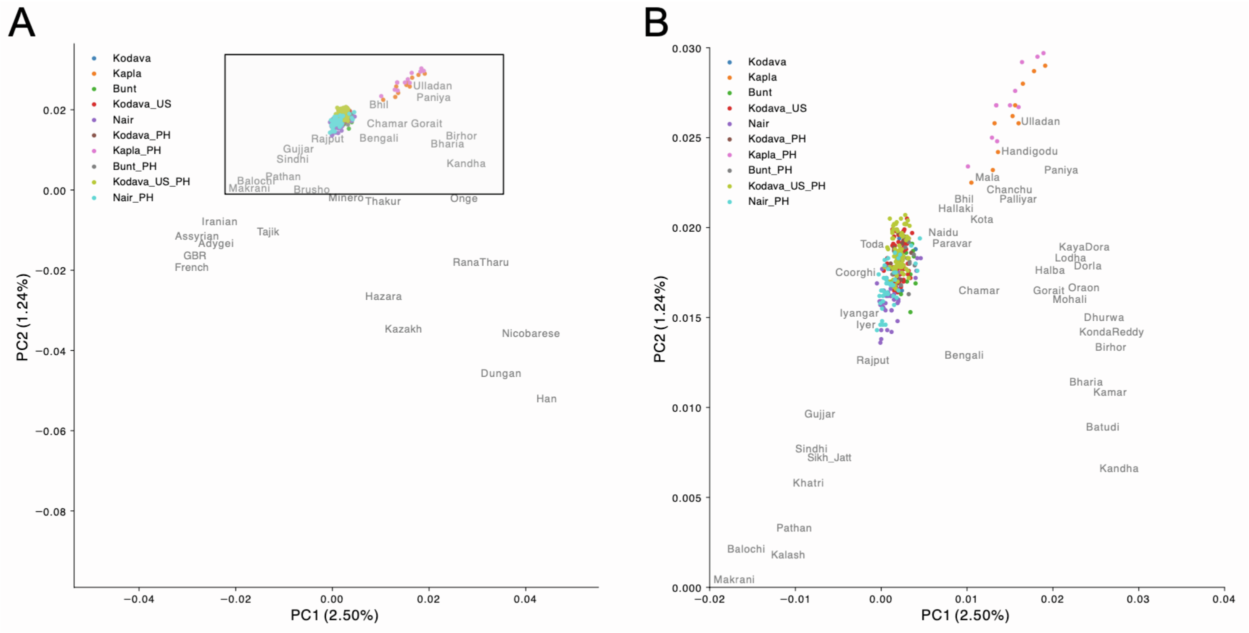


Figure S5A: Extended ADMIXTURE results across K=6 to K=11.

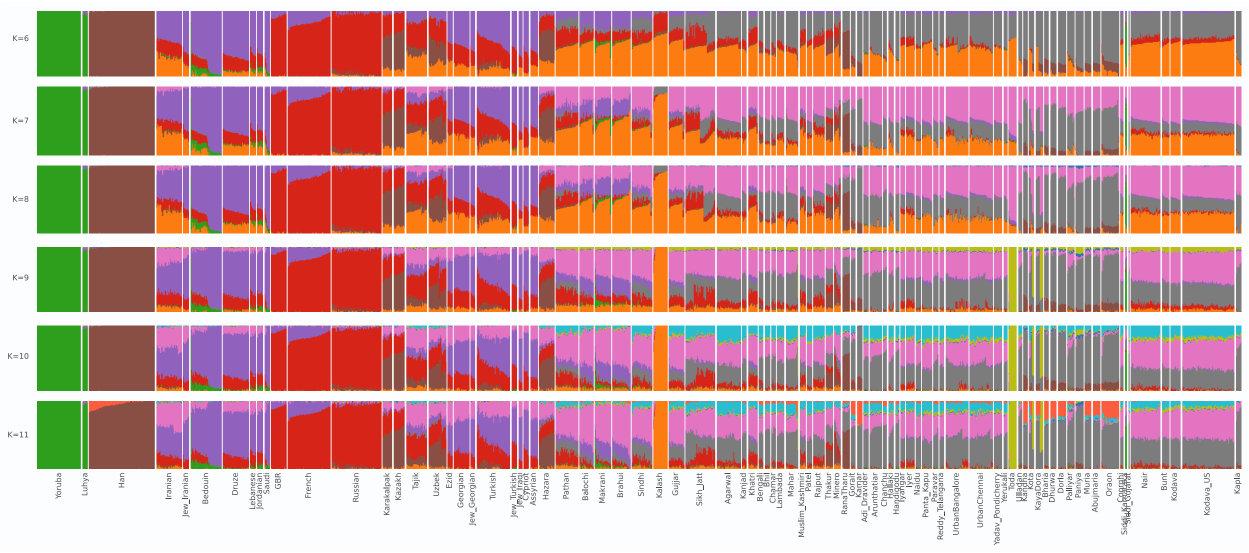


Figure S5B: Cross-validation error of ADMIXTURE from K=6 to K=11, minimized in K=7.

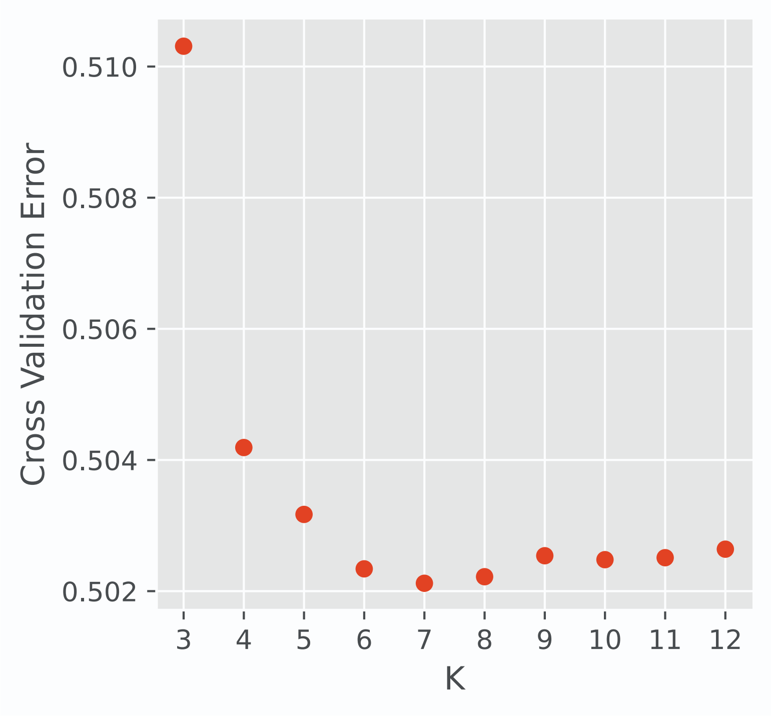


Figure S6: ADMIXTURE with pseudo-haploid calls.

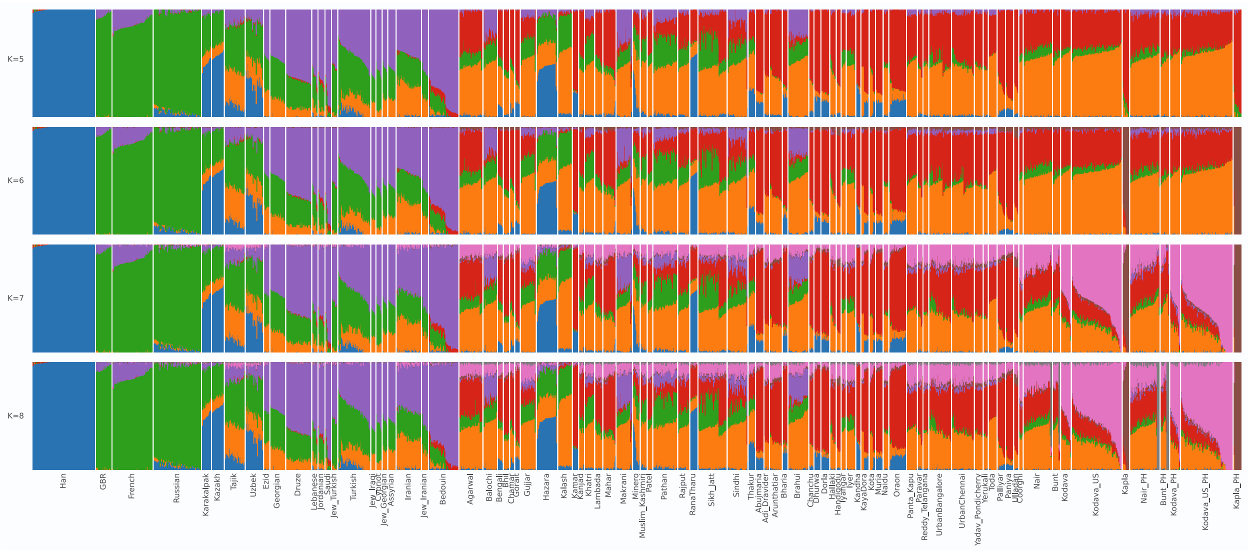


Figure S7A: PCA investigating genetic variation in Kapla. Dispersion of the eight Kapla individuals along the ANI-ASI cline with respect to other select South Indian populations with higher ASI component like the Ulladan, Paniya and Vysya. For comparison, Gujjar is shown as a representative of a population with higher ANI component.

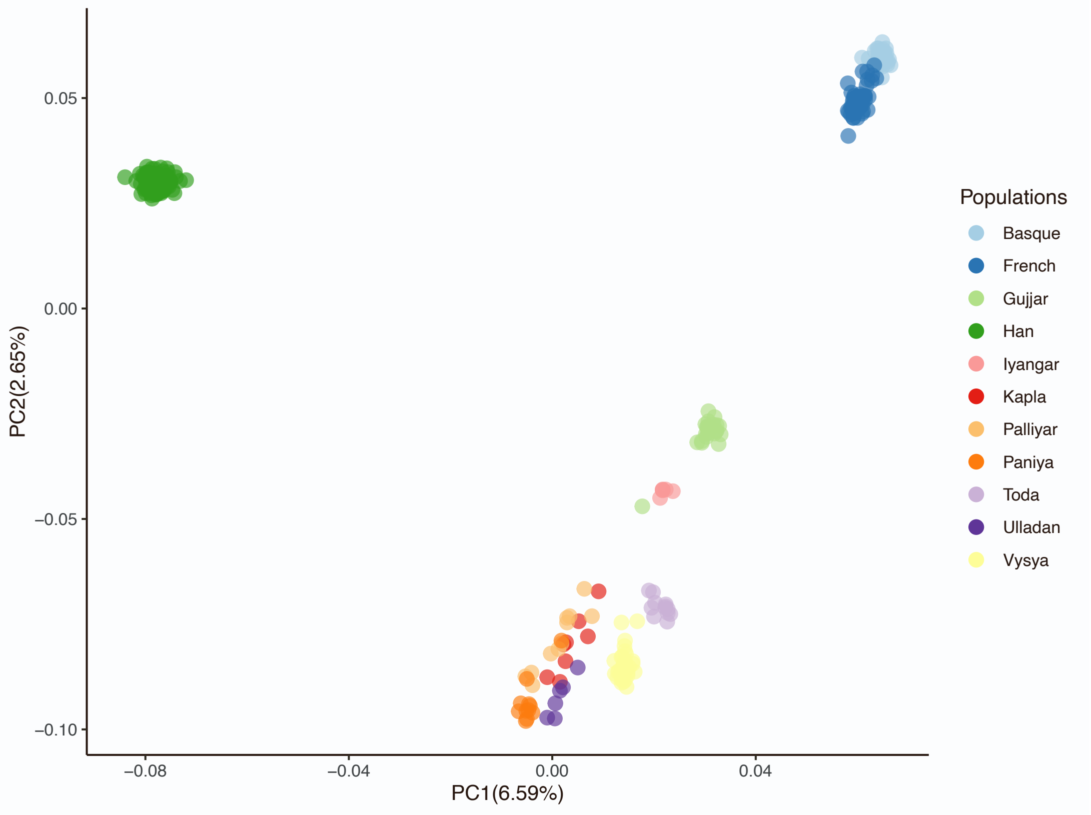


Figure S7B: PCA to investigate the genetic relationship between Kapla and Siddi/African populations.

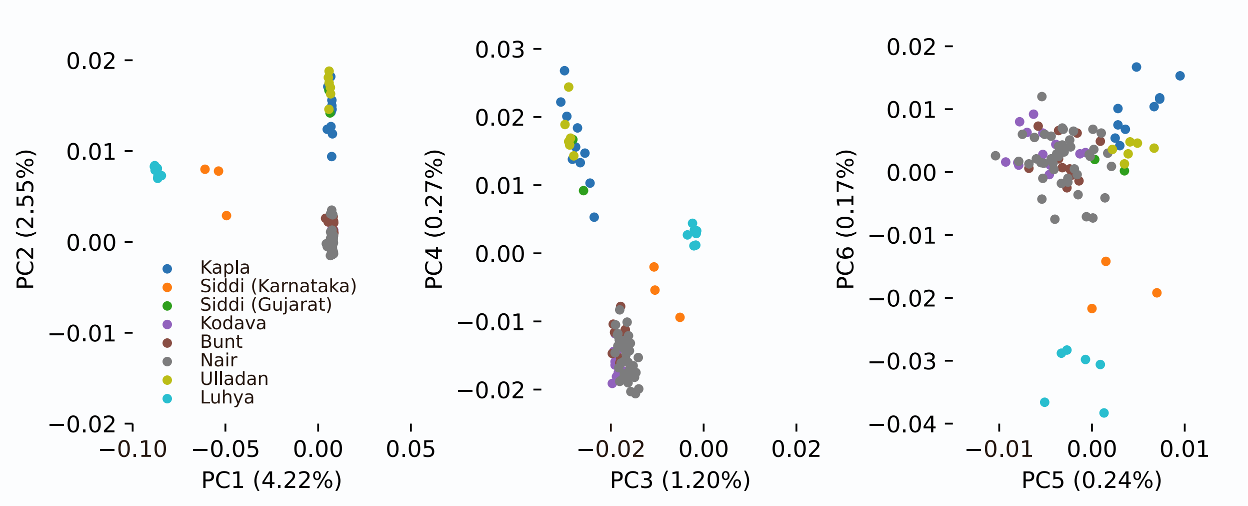


Figure S8: Outgroup-f_3 statistic of the form f_3(Target, X; Mbuti). Target corresponds to one of the Southwest Indian populations sequenced in this study and X to a set of Eurasian populations from the dataset described in Methods.

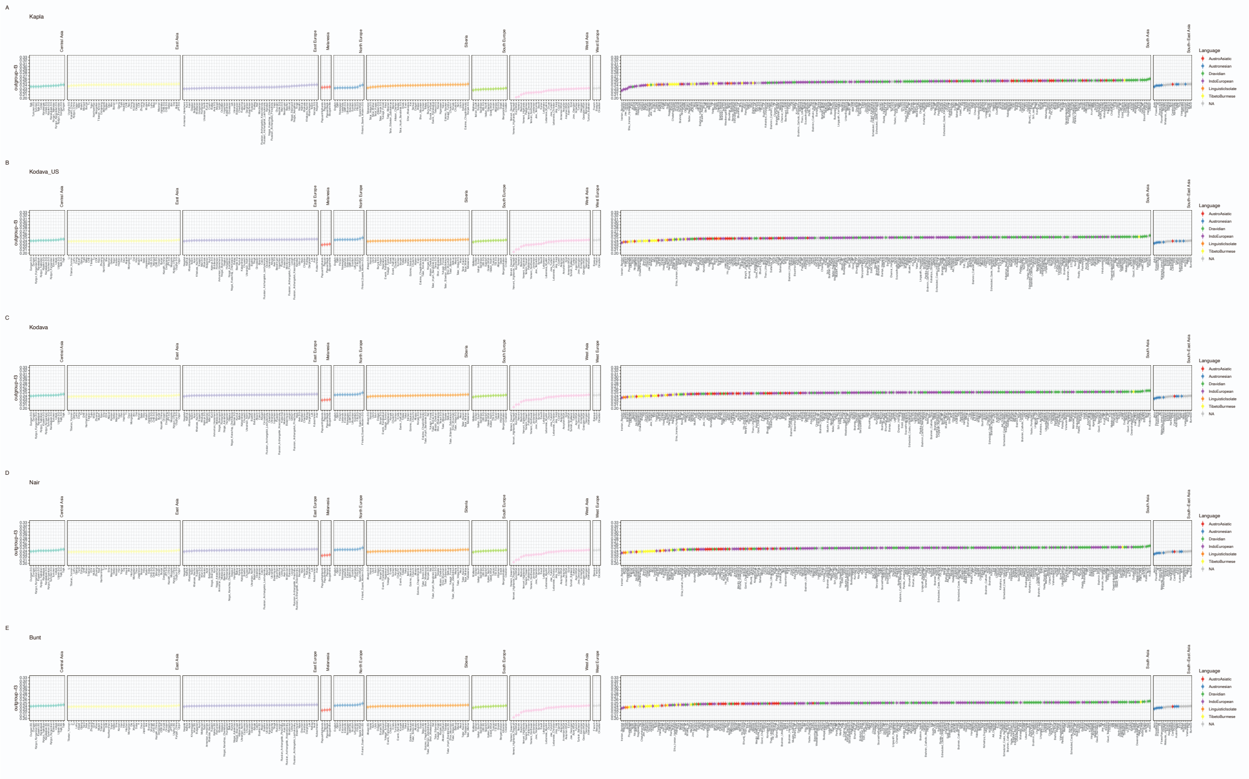


Figure S9: Outgroup-*f*3 statistic of the form *f*3(Target, X; Mbuti) using pseudo-haploid calls. Target corresponds to one of the Southwest Indian populations sequenced in this study and X to a set of Eurasian populations from the dataset described in Methods.

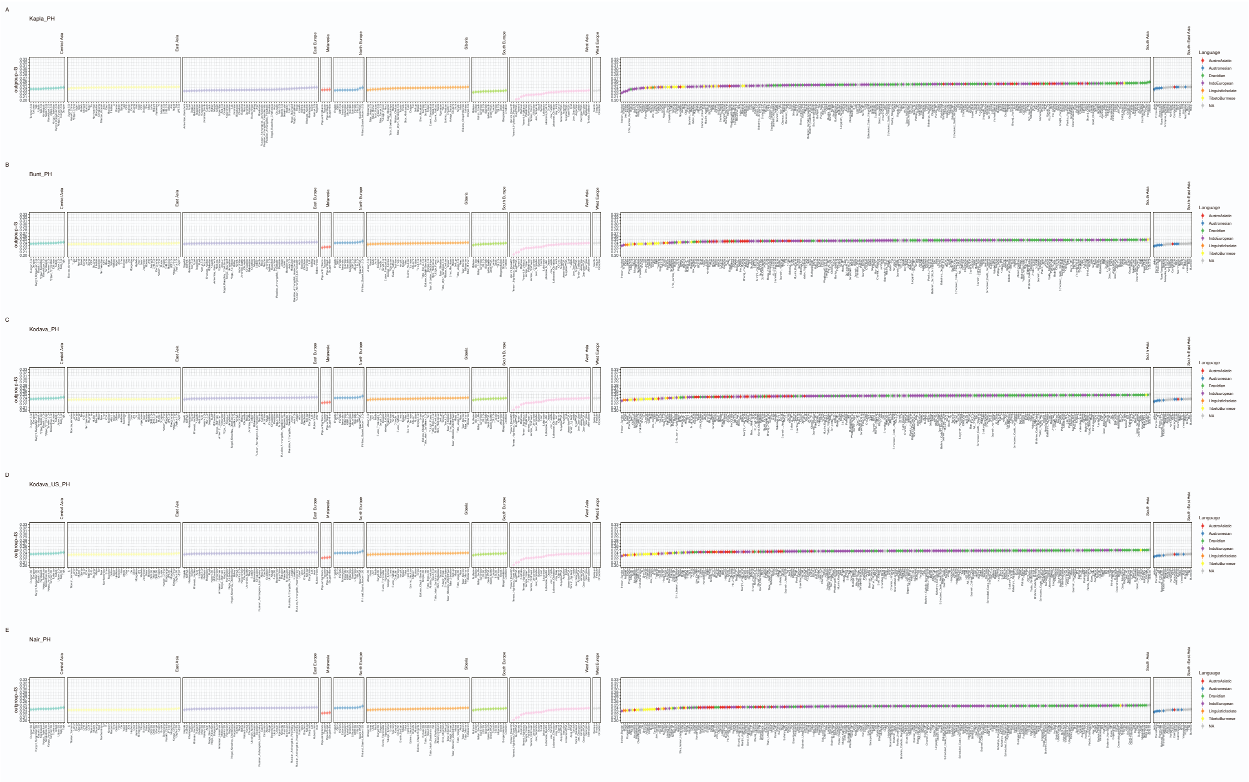
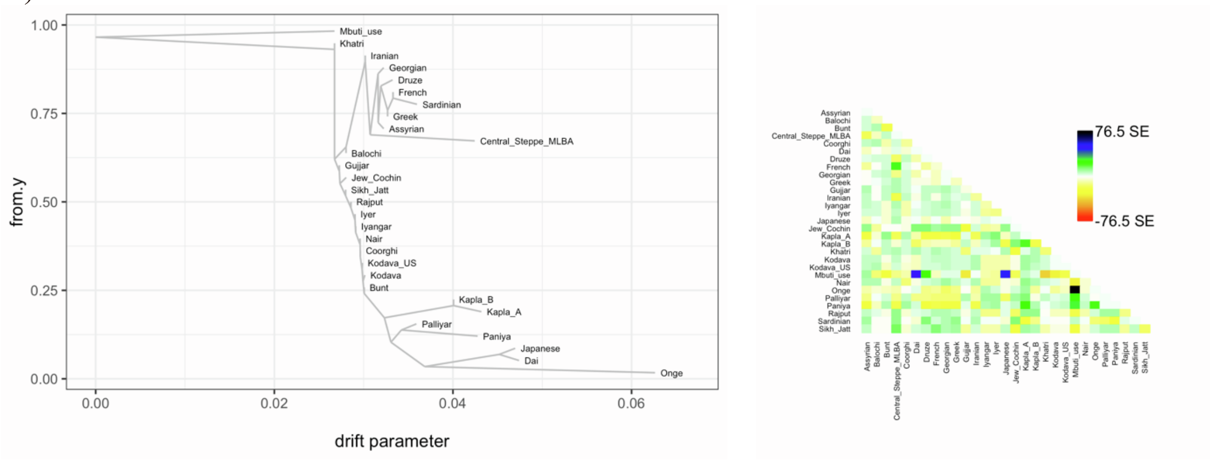
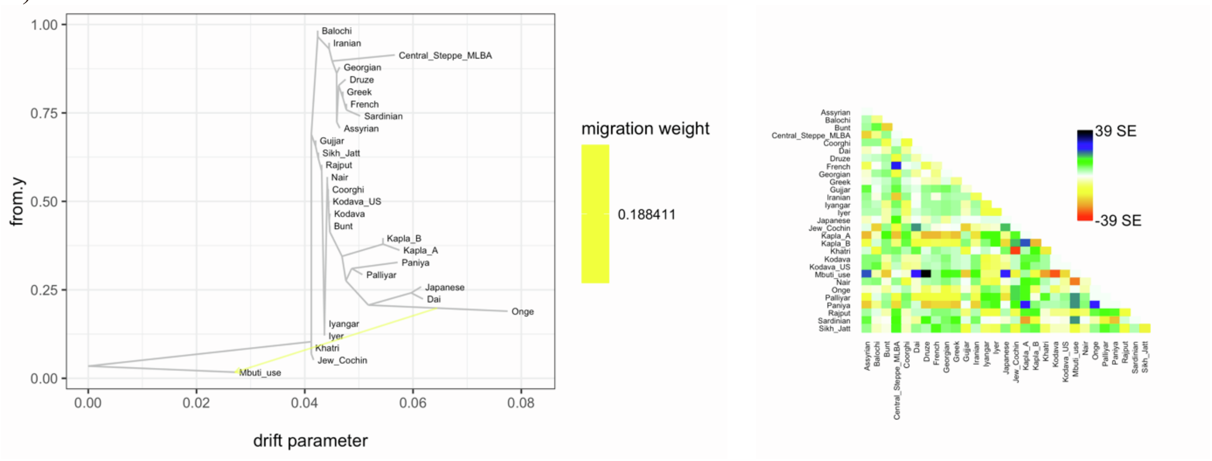


Figure S10A-K: Treemix results. Maximum Likelihood trees (left panel) and residuals (right panel) of allele frequency estimates from Treemix with up to 10 migration edges (m).

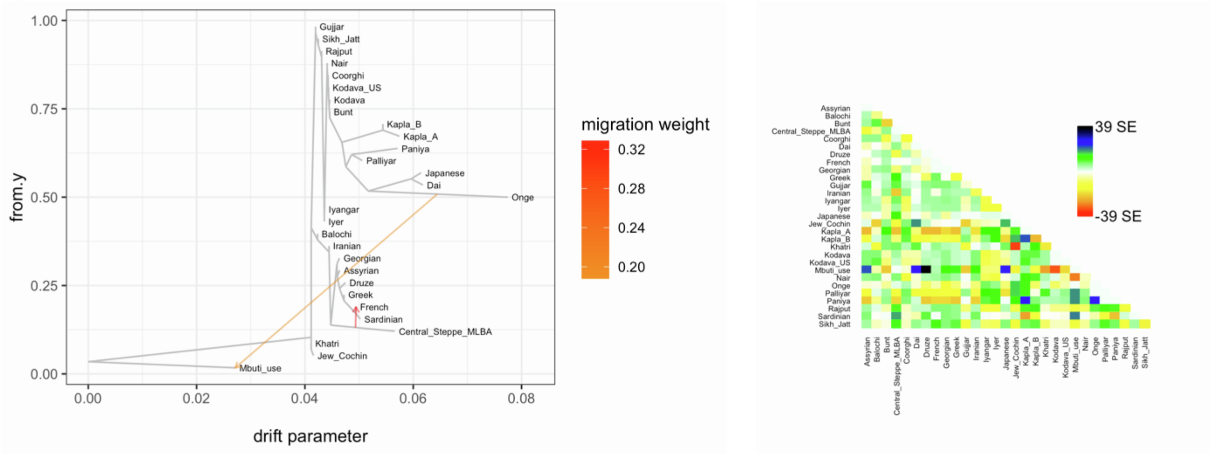
A) m=0



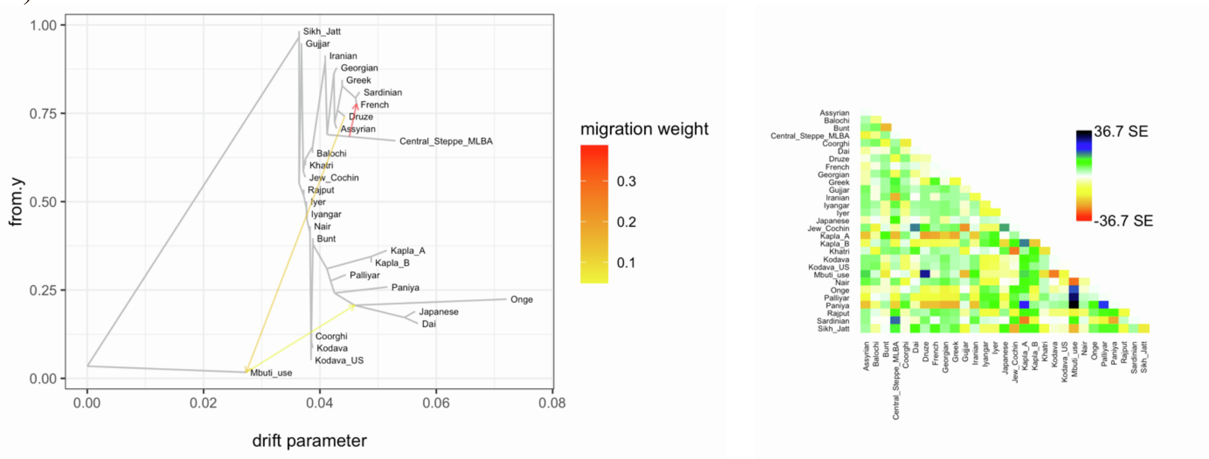
B) m=1



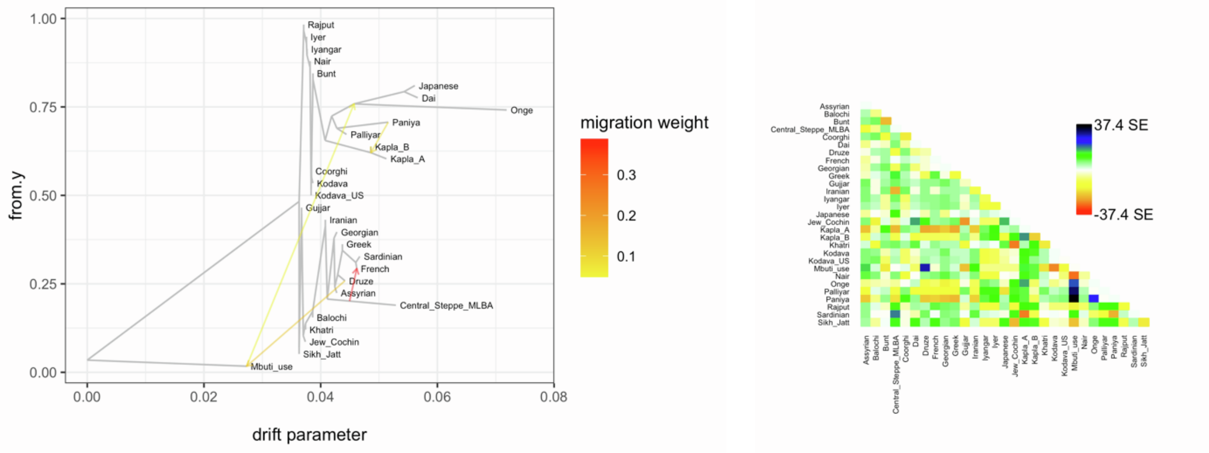
C) m=2



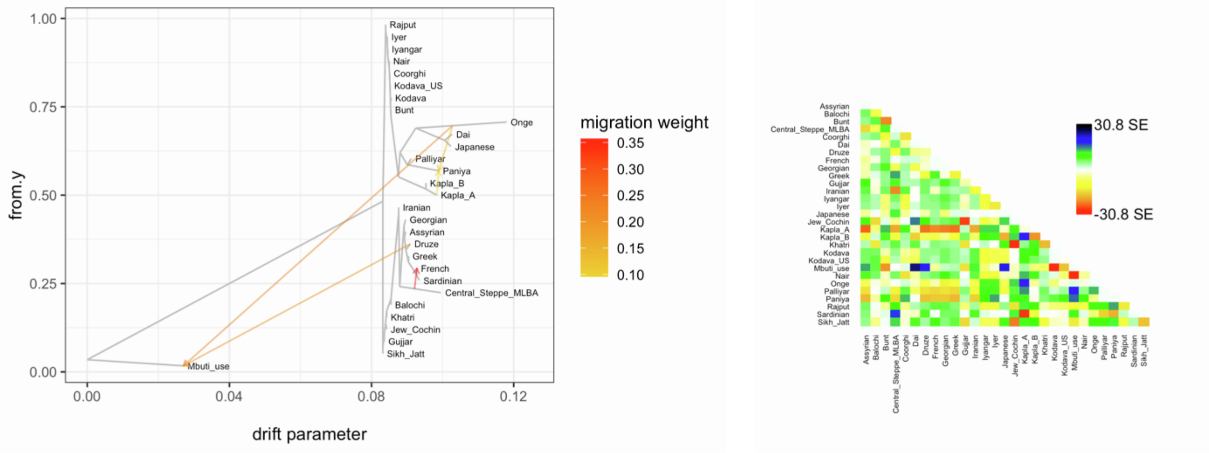
D) m=3


Sikh_Jatt
Gujjar
Iranian
Georgian
Greek
Sardinian
French
Druze
Assyrian
Central_Steppe_MLBA
Balochi
Khatri
Jew_Cochin
Rajput
Iyer
Iyangar
Nair
Bunt
Kapla_A
Kapla_B
Palliyar
Paniya
Onge
Japanese
Dai
Coorghi
Kodava
Kodava_US
Mbuti_use
from.y
drift parameter
1.00
0.75
0.50
0.25
0.00
0.00
0.02
0.04
0.06
0.08
migration weight
0.3
0.2
0.1
36.7 SE
-36.7 SE

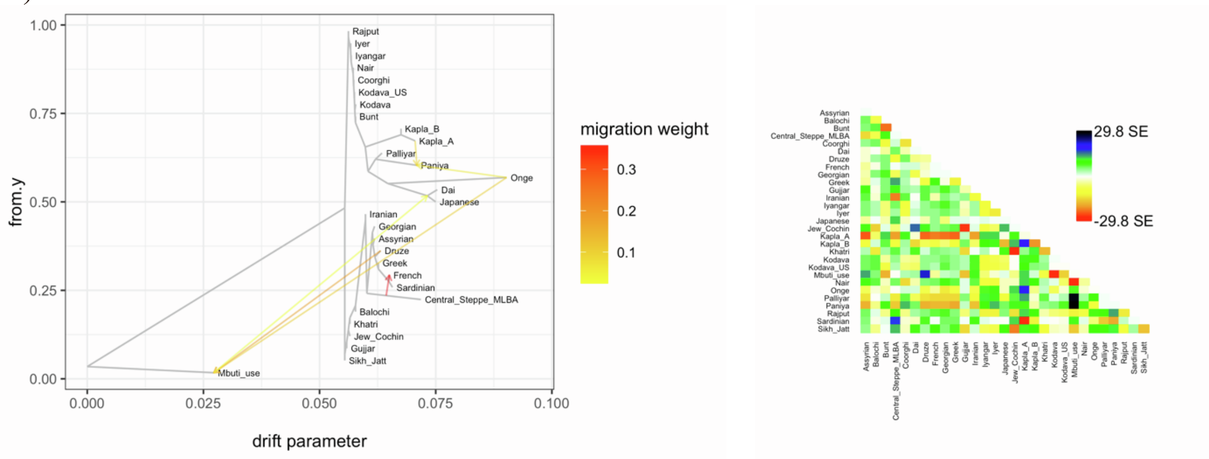
E) m=4


Rajput
Iyer
Iyangar
Nair
Bunt
Japanese
Dai
Onge
Paniya
Palliyar
Kapla_B
Kapla_A
Coorghi
Kodava
Kodava_US
Gujjar
Iranian
Georgian
Greek
Sardinian
French
Druze
Assyrian
Central_Steppe_MLBA
Balochi
Khatri
Jew_Cochin
Sikh_Jatt
Mbuti_use
from.y
drift parameter
1.00
0.75
0.50
0.25
0.00
0.00
0.02
0.04
0.06
0.08
migration weight
0.3
0.2
0.1
37.4 SE
-37.4 SE

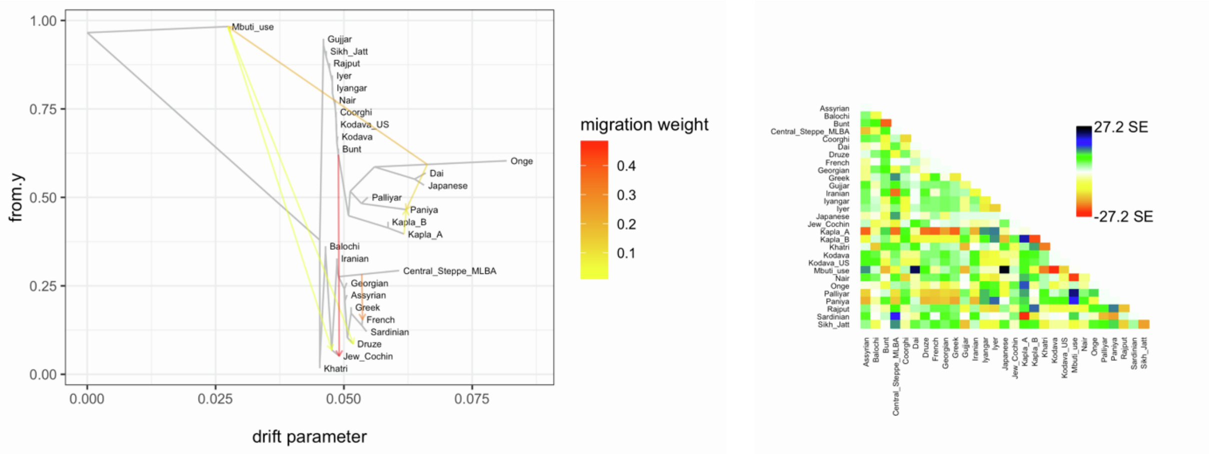
F) m=5


Rajput
Iyer
Iyangar
Nair
Coorghi
Kodava_US
Kodava
Bunt
Onge
Dai
Japanese
Palliyar
Paniya
Kapla_B
Kapla_A
Iranian
Georgian
Assyrian
Druze
Greek
French
Sardinian
Central_Steppe_MLBA
Balochi
Khatri
Jew_Cochin
Gujjar
Sikh_Jatt
Mbuti_use
from.y
drift parameter
1.00
0.75
0.50
0.25
0.00
0.00
0.04
0.08
0.12
migration weight
0.35
0.30
0.25
0.20
0.15
0.10
30.8 SE
-30.8 SE

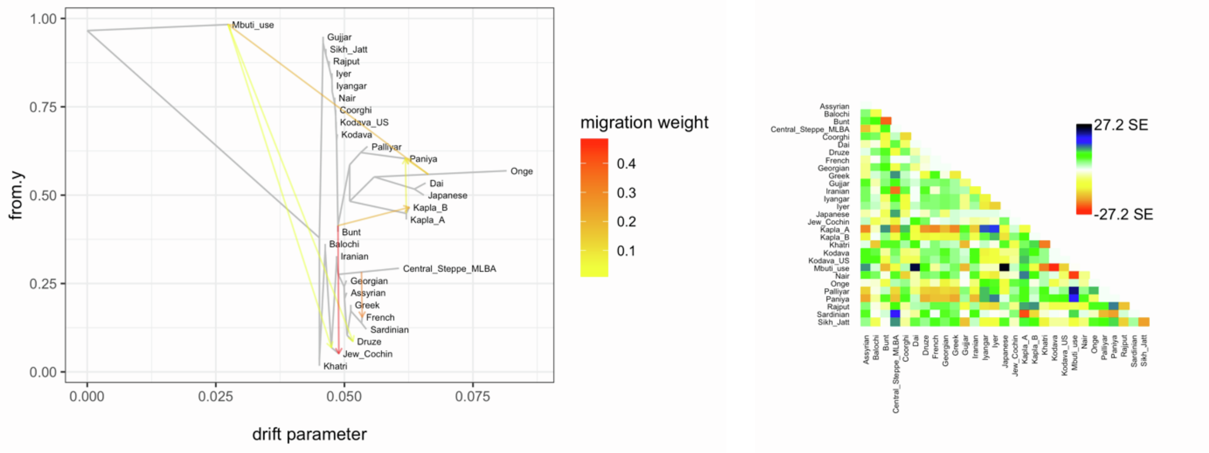
G) m=6


1.00
0.75
0.50
0.25
0.00
from.y
0.000
0.025
0.050
0.075
0.100
drift parameter
Rajput
Iyer
Iyangar
Nair
Coorghi
Kodava_US
Kodava
Bunt
Kapla_B
Kapla_A
Palliyar
Paniya
Onge
Dai
Japanese
Iranian
Georgian
Assyrian
Druze
Greek
French
Sardinian
Central_Steppe_MLBA
Balochi
Khatri
Jew_Cochin
Gujjar
Sikh_Jatt
Mbuti_use
migration weight
0.3
0.2
0.1
Assyrian
Balochi
Bunt
Central_Steppe_MLBA
Coorghi
Dai
Druze
French
Georgian
Greek
Gujjar
Iranian
Iyangar
Iyer
Japanese
Jew_Cochin
Kapla_A
Kapla_B
Khatri
Kodava
Kodava_US
Mbuti_use
Nair
Onge
Palliyar
Paniya
Rajput
Sardinian
Sikh_Jatt
29.8 SE
-29.8 SE

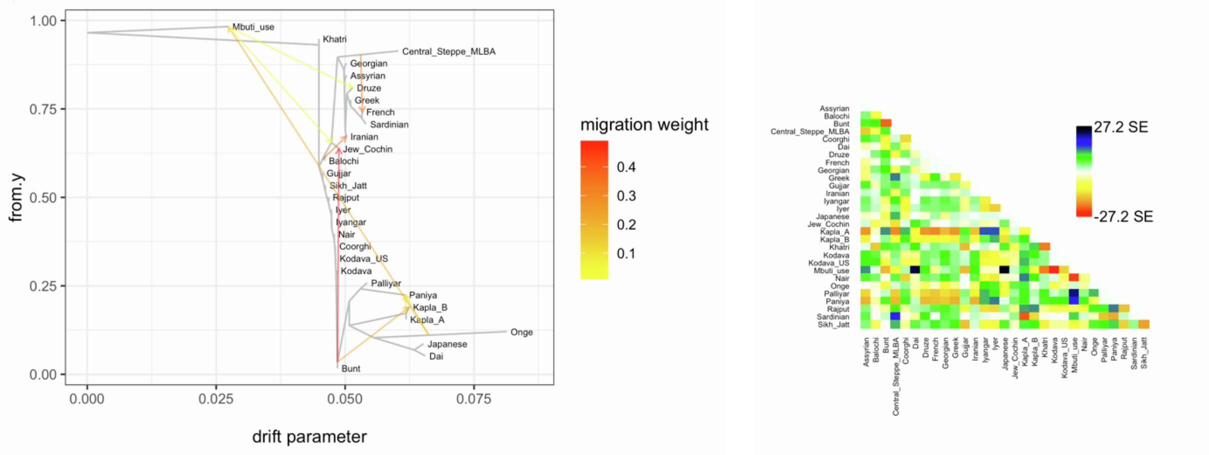
H) m=7


1.00
0.75
0.50
0.25
0.00
from.y
0.000
0.025
0.050
0.075
drift parameter
Mbuti_use
Gujjar
Sikh_Jatt
Rajput
Iyer
Iyangar
Nair
Coorghi
Kodava_US
Kodava
Bunt
Onge
Dai
Japanese
Palliyar
Paniya
Kapla_B
Kapla_A
Balochi
Iranian
Central_Steppe_MLBA
Georgian
Assyrian
Greek
French
Sardinian
Druze
Jew_Cochin
Khatri
migration weight
0.4
0.3
0.2
0.1
Assyrian
Balochi
Bunt
Central_Steppe_MLBA
Coorghi
Dai
Druze
French
Georgian
Greek
Gujjar
Iranian
Iyangar
Iyer
Japanese
Jew_Cochin
Kapla_A
Kapla_B
Khatri
Kodava
Kodava_US
Mbuti_use
Nair
Onge
Palliyar
Paniya
Rajput
Sardinian
Sikh_Jatt
27.2 SE
-27.2 SE

I) m=8


1.00
0.75
0.50
0.25
0.00
from.y
0.000
0.025
0.050
0.075
drift parameter
Mbuti_use
Gujjar
Sikh_Jatt
Rajput
Iyer
Iyangar
Nair
Coorghi
Kodava_US
Kodava
Palliyar
Paniya
Onge
Dai
Japanese
Kapla_B
Kapla_A
Bunt
Balochi
Iranian
Central_Steppe_MLBA
Georgian
Assyrian
Greek
French
Sardinian
Druze
Jew_Cochin
Khatri
migration weight
0.4
0.3
0.2
0.1
Assyrian
Balochi
Bunt
Central_Steppe_MLBA
Coorghi
Dai
Druze
French
Georgian
Greek
Gujjar
Iranian
Iyangar
Iyer
Japanese
Jew_Cochin
Kapla_A
Kapla_B
Khatri
Kodava
Kodava_US
Mbuti_use
Nair
Onge
Palliyar
Paniya
Rajput
Sardinian
Sikh_Jatt
27.2 SE
-27.2 SE

J) m=9


from.y
drift parameter
1.00
0.75
0.50
0.25
0.00
0.000
0.025
0.050
0.075
Mbuti_use
Khatri
Central_Steppe_MLBA
Georgian
Assyrian
Druze
Greek
French
Sardinian
Iranian
Jew_Cochin
Balochi
Gujjar
Sikh_Jatt
Rajput
Iyer
Iyangar
Nair
Coorghi
Kodava_US
Kodava
Palliyar
Paniya
Kapla_B
Kapla_A
Onge
Japanese
Dai
Bunt
migration weight
0.4
0.3
0.2
0.1
27.2 SE
-27.2 SE
Assyrian
Balochi
Bunt
Central_Steppe_MLBA
Coorghi
Dai
Druze
French
Georgian
Greek
Gujjar
Iranian
Iyangar
Iyer
Japanese
Jew_Cochin
Kapla_A
Kapla_B
Khatri
Kodava
Kodava_US
Mbuti_use
Nair
Onge
Palliyar
Paniya
Rajput
Sardinian
Sikh_Jatt

K) m=10

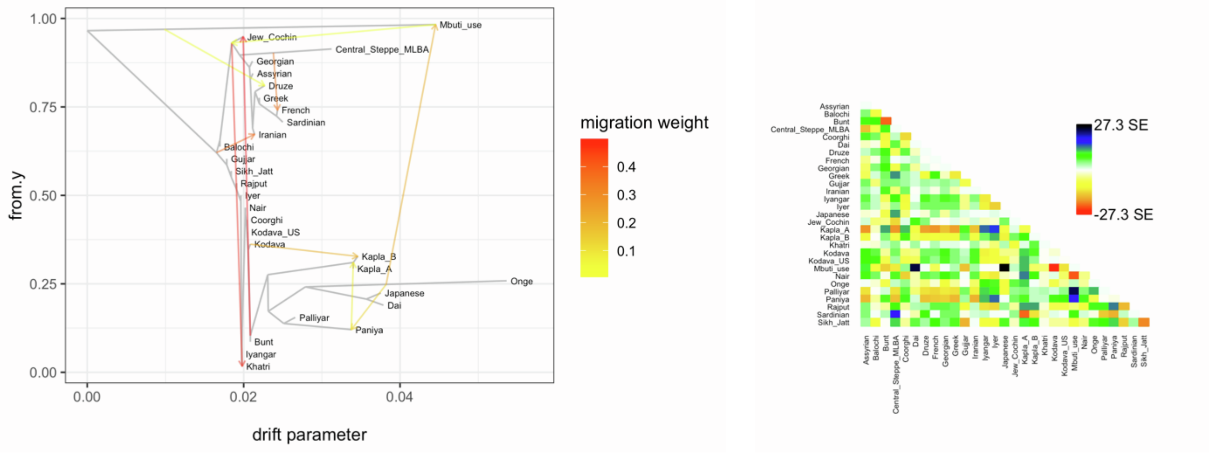

from.y
drift parameter
1.00
0.75
0.50
0.25
0.00
0.00
0.02
0.04
Mbuti_use
Jew_Cochin
Central_Steppe_MLBA
Georgian
Assyrian
Druze
Greek
French
Sardinian
Iranian
Balochi
Gujjar
Sikh_Jatt
Rajput
Iyer
Nair
Coorghi
Kodava_US
Kodava
Kapla_B
Kapla_A
Onge
Japanese
Dai
Palliyar
Paniya
Bunt
Iyangar
Khatri
migration weight
0.4
0.3
0.2
0.1
27.3 SE
-27.3 SE
Assyrian
Balochi
Bunt
Central_Steppe_MLBA
Coorghi
Dai
Druze
French
Georgian
Greek
Gujjar
Iranian
Iyangar
Iyer
Japanese
Jew_Cochin
Kapla_A
Kapla_B
Khatri
Kodava
Kodava_US
Mbuti_use
Nair
Onge
Palliyar
Paniya
Rajput
Sardinian
Sikh_Jatt

Figure S11: ***f₄ ratio* results.** Estimation of the proportion of ANI-related genetic ancestry in South Asians using Central Steppe MLBA (α).

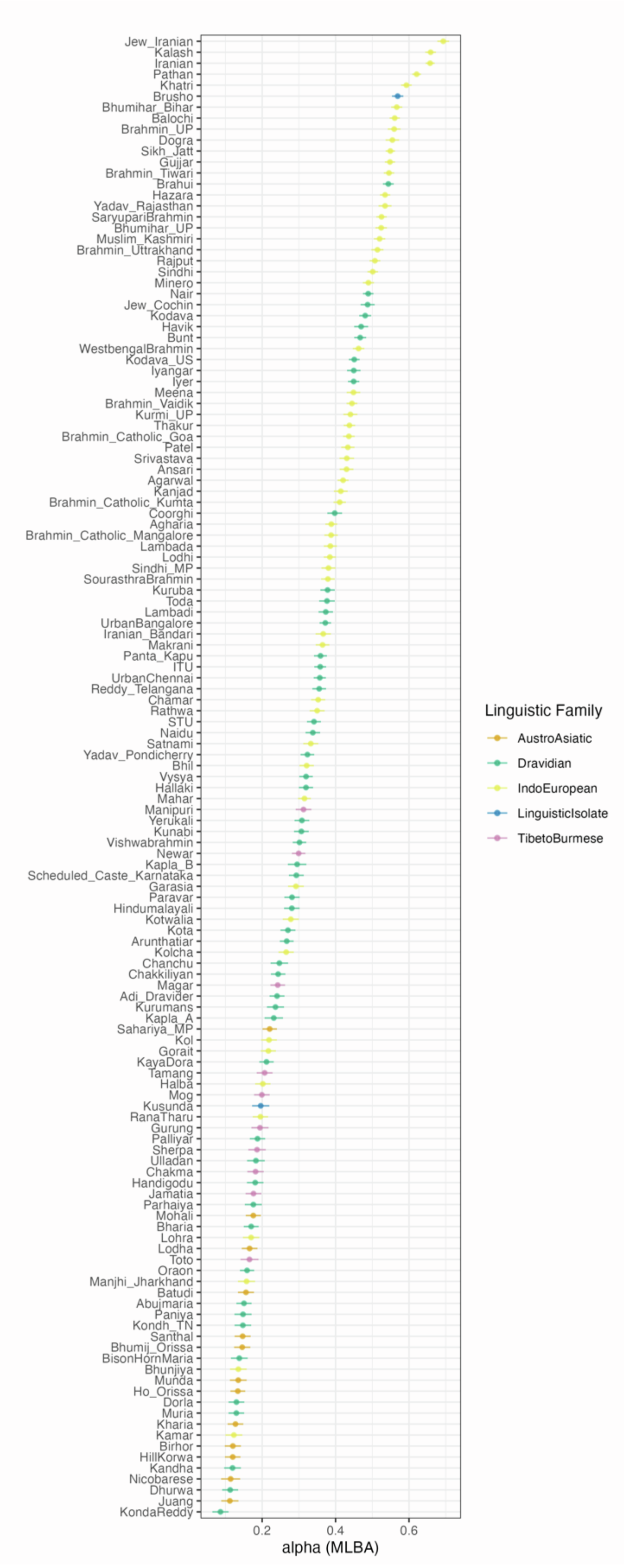


Figure S12: *D*-statistic with source populations from qpAdm model. *D*-statistic of the form *D*(South Asia, H2; Source; Mbuti), where H2 is one of the study populations and Source is one of Central Steppe MLBA (A), Indus Periphery (B) or Onge (C).

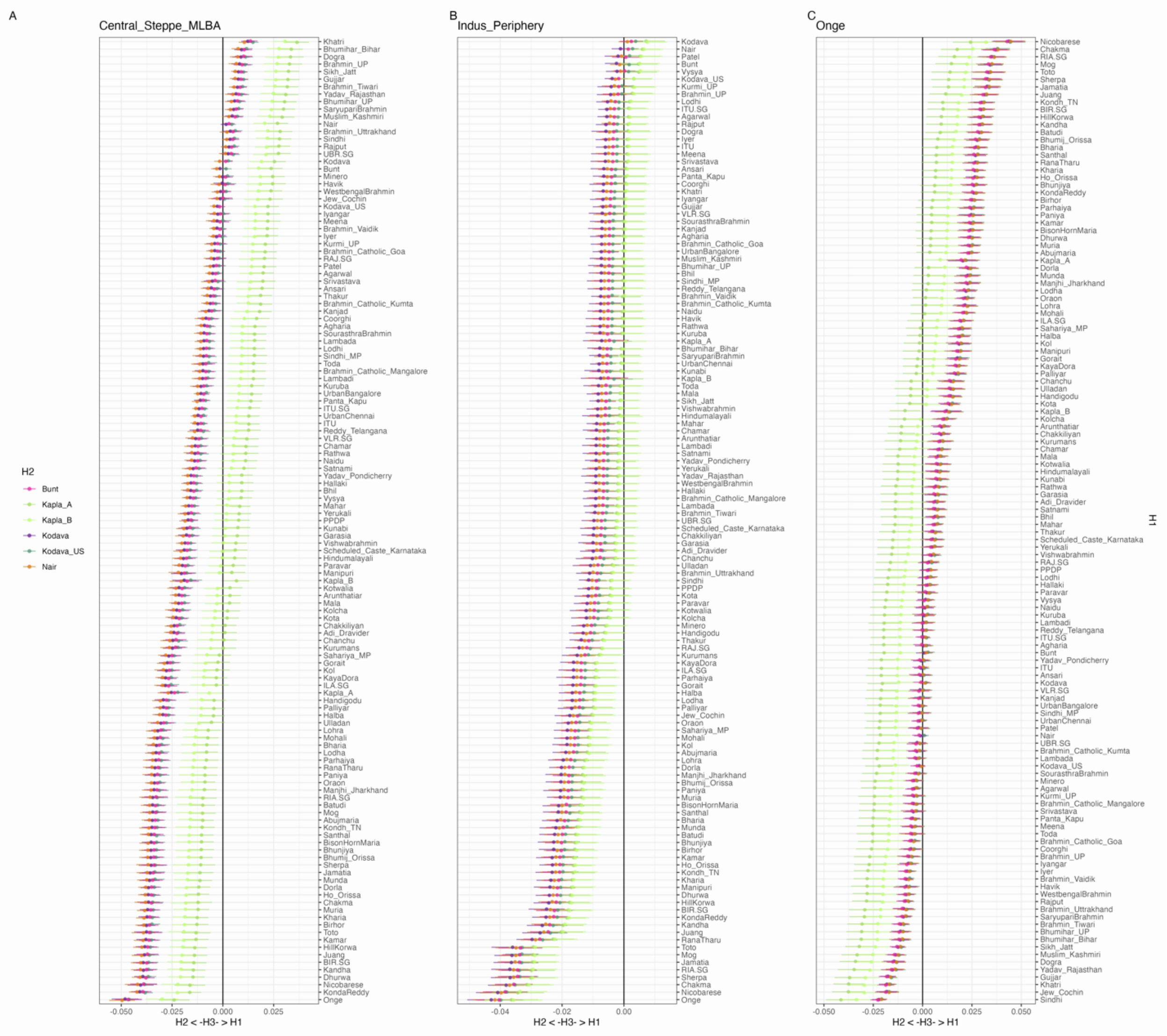


Figure S13: *D*-statistic with source populations from qpAdm model using pseudo-haploid calls. *D*-statistic of the form *D*(South Asia, H2; Source; Mbuti), where H2 is one of the study populations and Source is one of Central Steppe MLBA (A), Indus Periphery (B) or Onge (C).

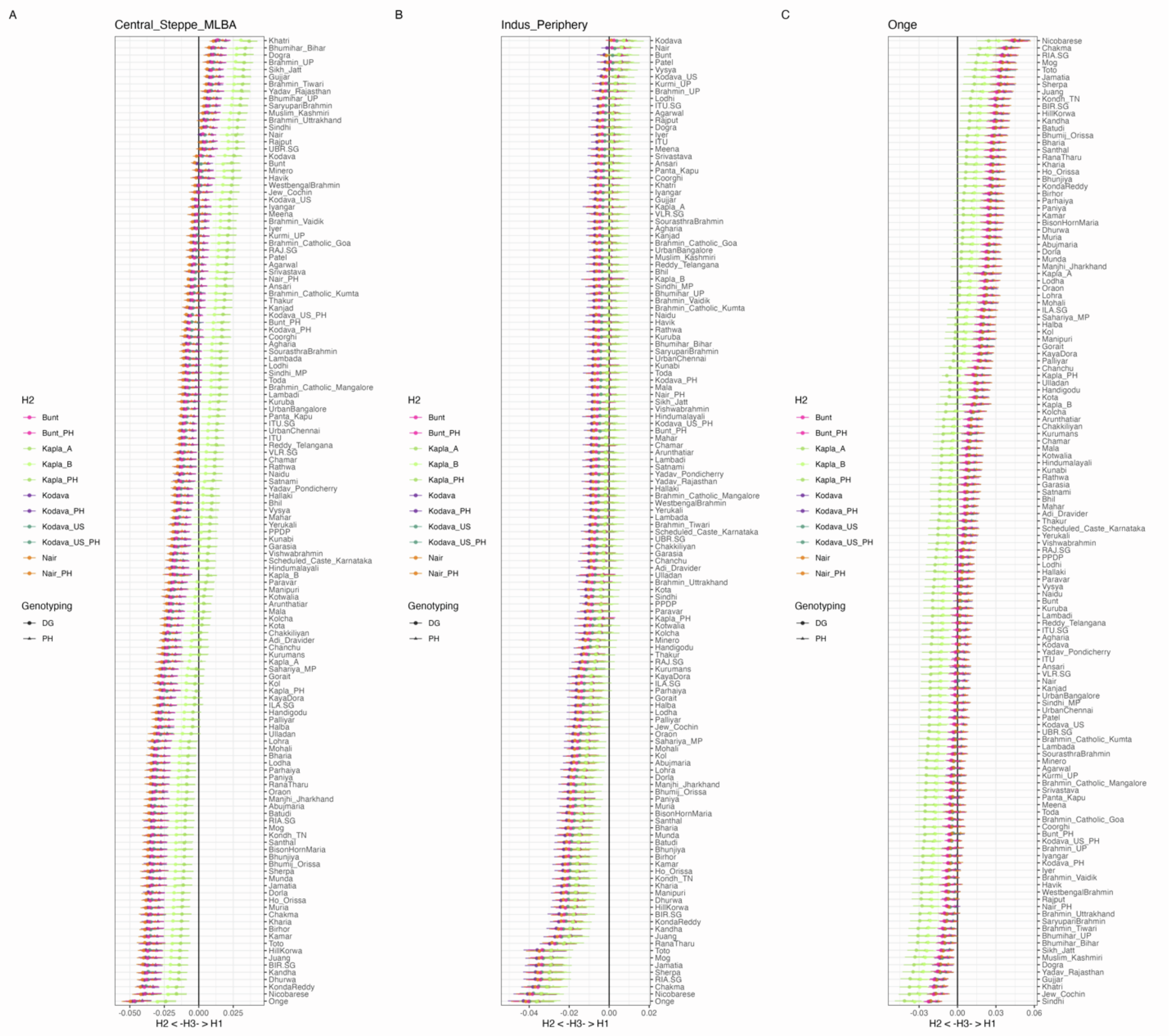


Figure S14: Summary of co-ancestry matrix as average length count per population from ChromoPainter.

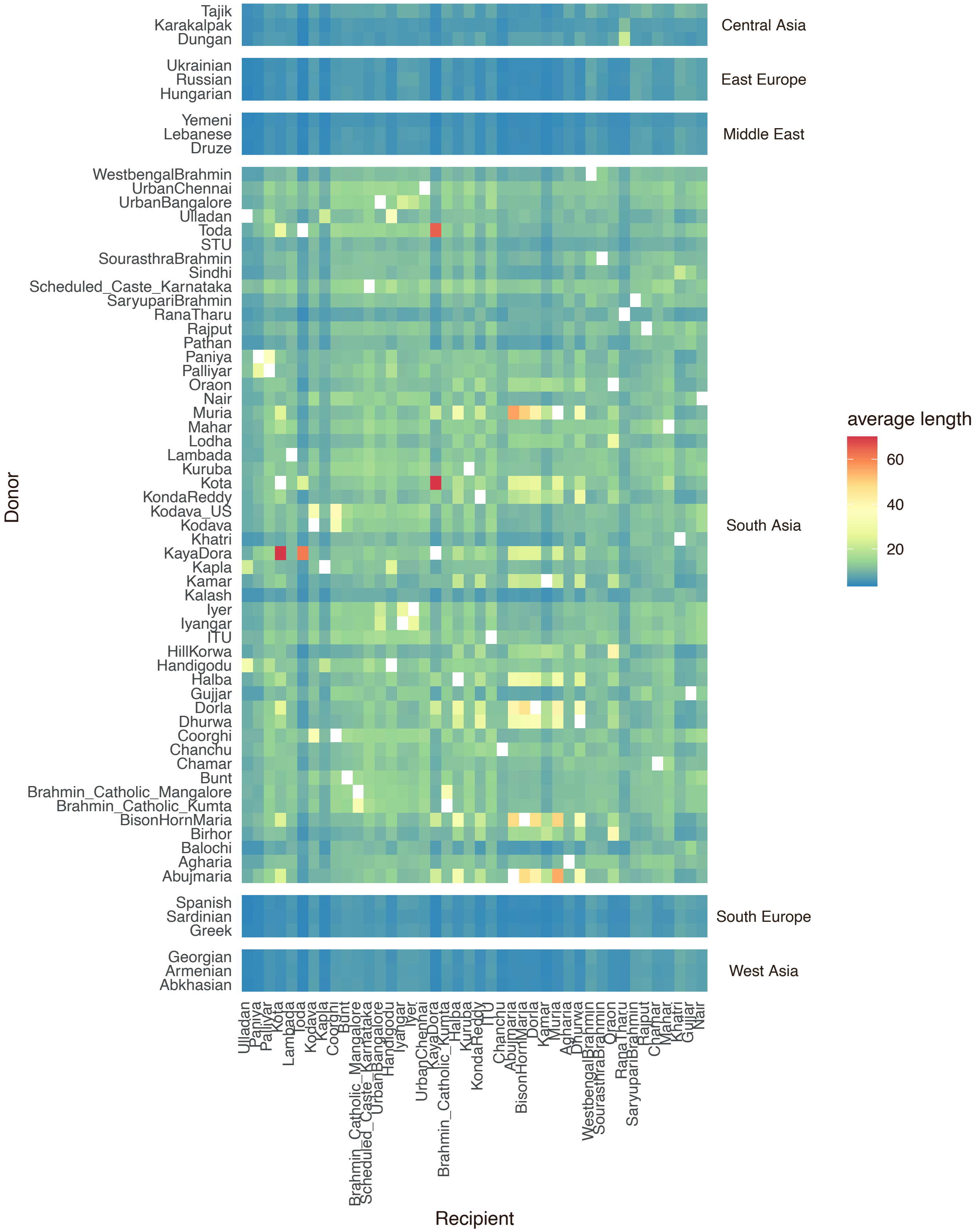


Figure S15: fineSTRUCTURE clustering dendrogram. This figure shows individual-level relatedness based on haplotypic similarity.

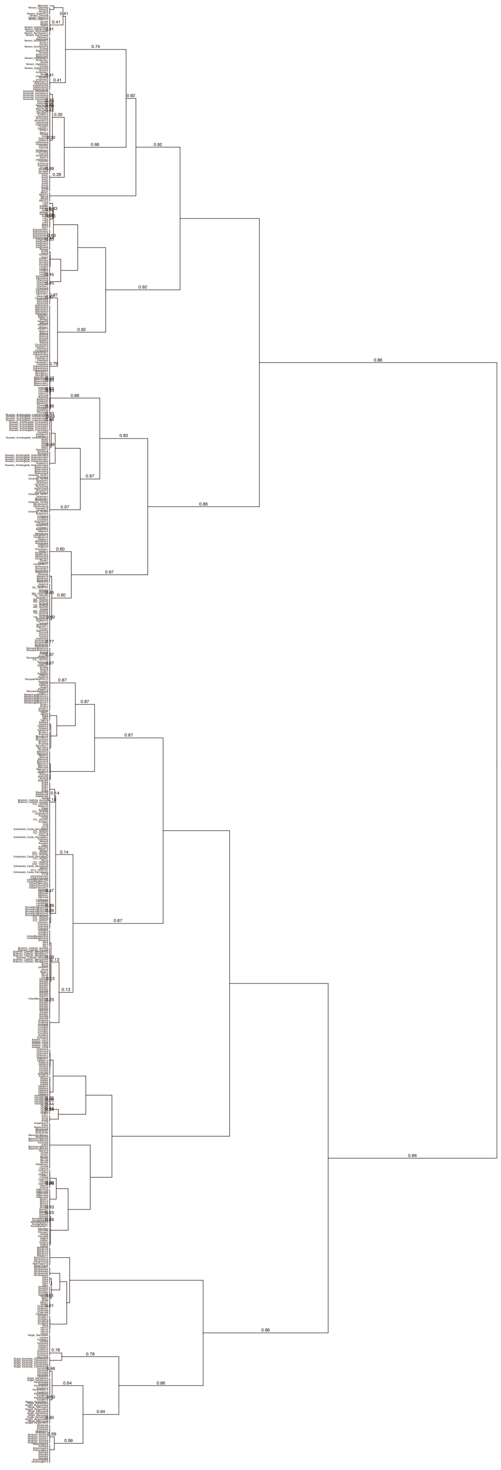


Figure S16: Ancestry-specific haplotype copying across sampled populations from Southwest India. The median per-locus copying probability from ChromoPainter represents the haplotype-level similarity between individuals from the focal populations from India – Rajput and Paniya added to reflect high ANI & high ASI populations, respectively – and potential sources of Western Eurasian ancestry.

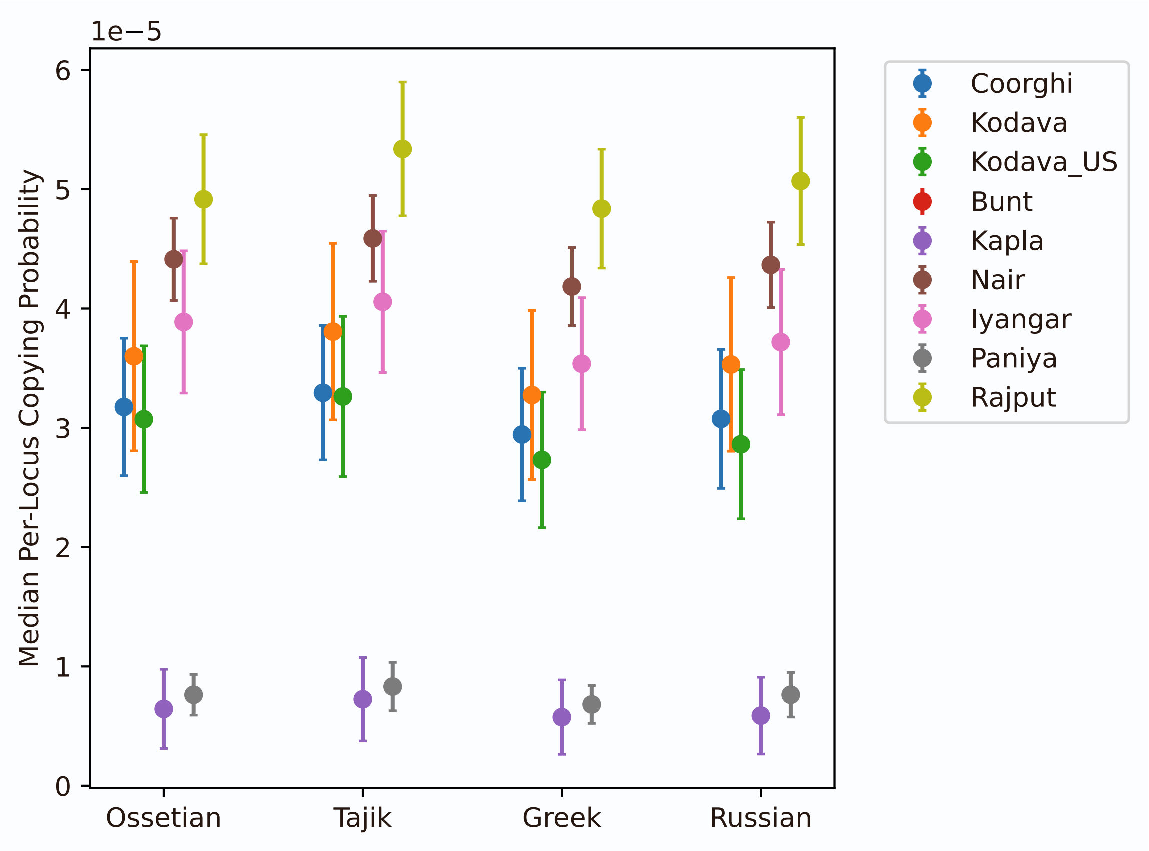


Figure S17: Total length of ROH (SROH) versus the total number of ROH (NROH). Averaged within populations.

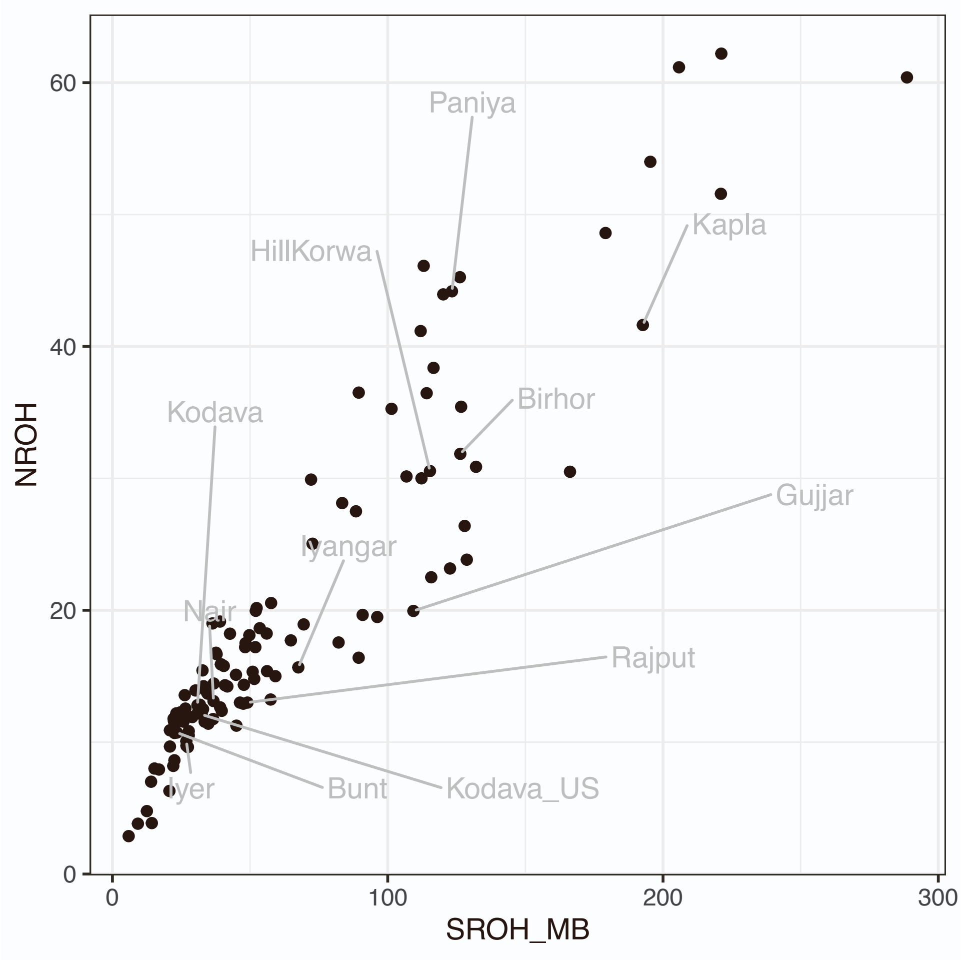


Figure S18: Short versus long ROH for select populations. A) Sum of total length of short (2-5Mb) ROH; B) Sum of total length of long (>10Mb) ROH.

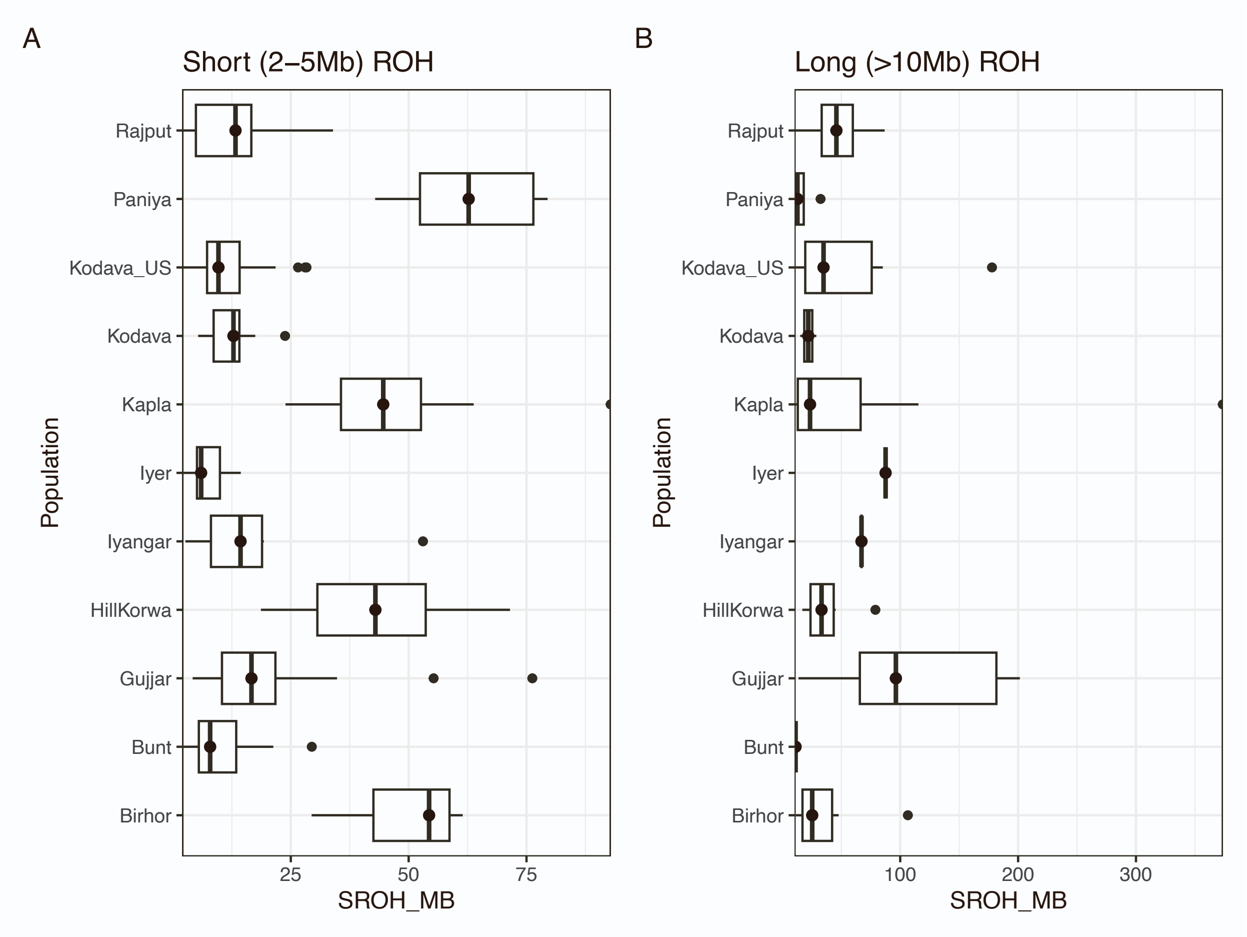


Figure S19: Decay curves for the Southwest India populations (Kapla, Bunt, Kodava, Kodava US, and Nair) and other South Asian populations. X-axis represents the genetic distance in cM and the Y-axis the allele sharing correlation. The legend for each figure shows the mean and CI for the founder event (Tf) in generations before the present (gBP, 1 generation = 28 years), the intensity of the founder event (If), and the normalized root-mean-square deviation (NRMSD). The algorithm was run using Mbuti as the outgroup population since South Asian populations do not share a recent bottleneck/founder event with this African population. Populations names highlighted in yellow lacked evidence for a significant founder event (see Methods).

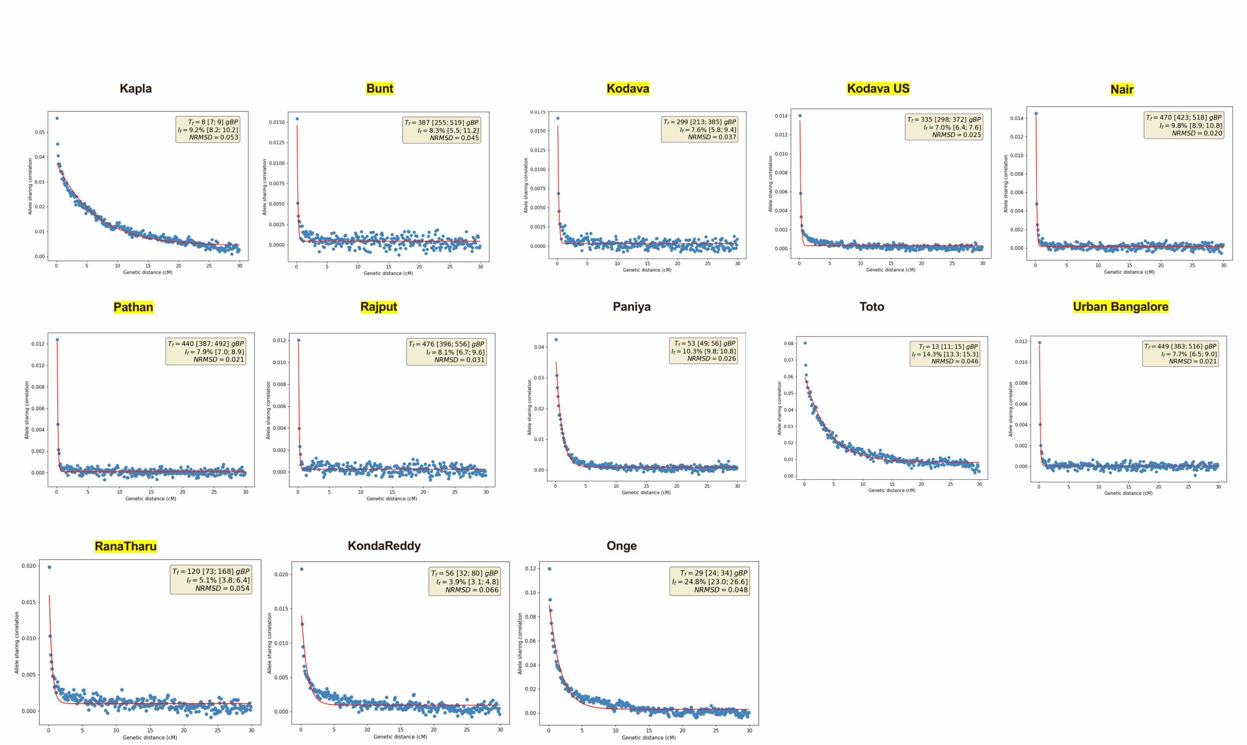


Figure S20: Functional and novel variation in high-coverage WGS samples from Southwest India. (A) Estimation of burden of functional mutations across all four study populations using 5000 binomial resampling iterations from carrier frequencies at functional alleles; (B) Estimating fraction of novel mutations per study population using binomial resampling of frequencies.

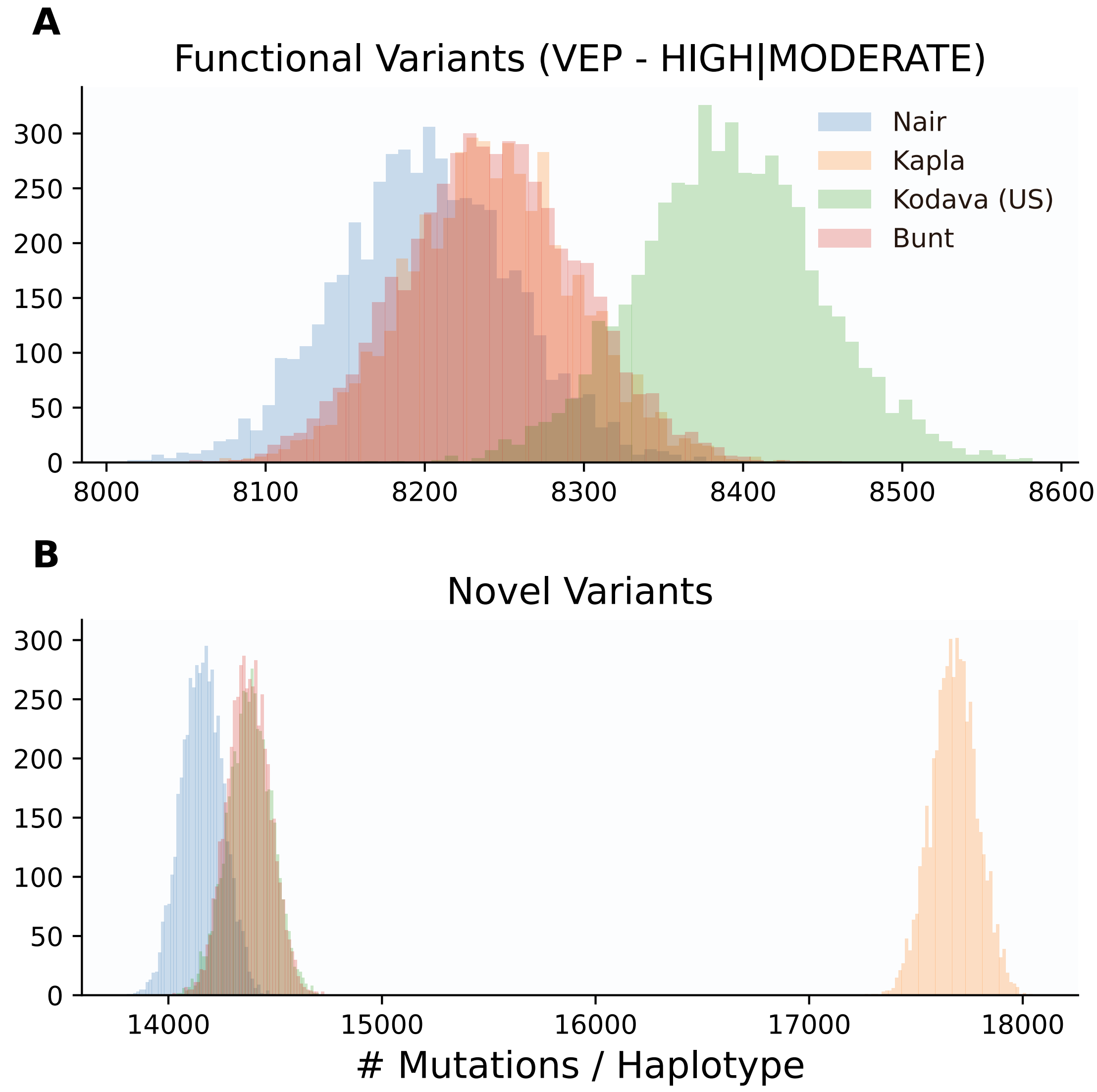


Table S2: Populations included in Treemix analysis

Population Label	Sample size
Assyrian	11
Balochi	21
Bunt	11
Coorghi	5
Dai	10
Druze	39
French	27
Georgian	13
Greek	20
Gujjar	23
Iranian	30
Iyangar	6
Iyer	13
Japanese	29
Jew_Cochin	5
Kapla_A	5
Kapla_B	3
Khatri	11
Kodava	15
Kodava_US	78
Mbuti	10
Nair	44
Onge	5
Palliyar	11
Paniya	11
Rajput	14
Central_Steppe_MLBA	35
Sardinian	27
Sikh_Jatt	44

Table S4: Populations used for calculating PCA-based SNP-loadings

Population Label	**Sample size**	**Language Family**
Handigodu	6	Dravidian
Toda	12	Dravidian
Paniya	11	Dravidian
Yadav_Pondicherry	12	Dravidian
Iyangar	6	Dravidian
Vysya	39	Dravidian
Iyer	13	Dravidian
Yerukali	6	Dravidian
Palliyar	11	Dravidian
UrbanChennai	34	Dravidian
STU	8	Dravidian
Arunthatiar	18	Dravidian
Nair	44	Dravidian
Kodava_US	78	Dravidian
Ulladan	6	Dravidian
UrbanBangalore	34	Dravidian
Kodava	15	Dravidian
ITU	25	Dravidian
Bunt	11	Dravidian
Chakkiliyan	9	Dravidian
Paravar	7	Dravidian
Panta_Kapu	15	Dravidian
Kapla_A	5	Dravidian
Kapla_B	3	Dravidian
Adi_Dravider	7	Dravidian
SourasthraBrahmin	9	Indo-European
Agarwal	36	Indo-European
SaryupariBrahmin	13	Indo-European
Brahmin_Catholic_Goa	14	Indo-European
Brahmin_Vaidik	25	Indo-European

Patel	7	Indo-European
Mahar	19	Indo-European
Brahmin_Catholic_Mangalore	6	Indo-European
Brahmin_Catholic_Kumta	10	Indo-European
Sindhi	30	Indo-European
Lambada	11	Indo-European
Brahmin_Tiwari	15	Indo-European
WestbengalBrahmin	10	Indo-European
Pathan	33	Indo-European
Khatri	14	Indo-European
Gujjar	23	Indo-European
Bhumihar_Bihar	7	Indo-European
Rajput	17	Indo-European
Sikh_Jatt	44	Indo-European
Bhil	8	Indo-European
Basque	33	Indo-European

Table S10: Allele sharing of Paniya and Kapla_A/Kapla_B/Kapla with ancient and present-day African groups using *D*-statistics. *D*-statistics results for both GATK and pseudo-haploid (PH) versions of Kapla_A, Kapla_B, and Kapla have been reported here.These analyses were performed to evaluate patterns of allele sharing between Paniya and Kapla_A/Kapla_B/Kapla with groups from Africa. Kapla refers to a pool of Kapla_A and Kapla_B individuals.

GATK version									
H1	**H2**	**H3**	**Outgroup**	***D***	**SE**	**Z-score**	**BABA**	**ABBA**	**Total no snps**
Paniya	Kapla_A	Luhya	Mbuti	-0.003	0.001	-2.002	20439	20547	375257
Paniya	Kapla_A	Yoruba	Mbuti	-0.002	0.001	-1.489	20406	20479	375397
Paniya	Kapla_A	Ethiopia_4500BP_published.SG	Mbuti	-0.002	0.002	-0.830	20814	20898	375184
Paniya	Kapla_A	BantuKenya	Mbuti	-0.002	0.001	-1.852	20513	20611	375397
Paniya	Kapla_B	Luhya	Mbuti	-0.003	0.001	-1.713	20582	20688	375257
Paniya	Kapla_B	Yoruba	Mbuti	-0.002	0.001	-1.628	20542	20635	375397
Paniya	Kapla_B	Ethiopia_4500BP_published.SG	Mbuti	-0.002	0.003	-0.679	20962	21041	375184
Paniya	Kapla_B	BantuKenya	Mbuti	-0.002	0.001	-1.480	20659	20750	375397
Paniya	Kapla	Luhya	Mbuti	-0.003	0.001	-2.085	20488	20593	375257
Paniya	Kapla	Yoruba	Mbuti	-0.002	0.001	-1.712	20452	20532	375397
Paniya	Kapla	Ethiopia_4500BP_published.SG	Mbuti	-0.002	0.002	-0.822	20867	20944	375184
Paniya	Kapla	BantuKenya	Mbuti	-0.002	0.001	-1.885	20563	20658	375397

PH version									
H1	**H2**	**H3**	**Outgroup**	***D***	**SE**	**Z-score**	**BABA**	**ABBA**	**Total no snps**
Paniya	Kapla_A_PH	Luhya	Mbuti	-0.002	0.001	-1.237	20715	20786	425476
Paniya	Kapla_A_PH	Yoruba	Mbuti	-0.001	0.001	-0.893	20669	20714	425476
Paniya	Kapla_A_PH	Ethiopia_4500BP_published.SG	Mbuti	0.001	0.002	0.262	21134	21107	425394
Paniya	Kapla_A_PH	BantuKenya	Mbuti	-0.002	0.001	-1.093	20781	20842	425476
Paniya	Kapla_B_PH	Luhya	Mbuti	-0.002	0.002	-1.002	20830	20895	425476
Paniya	Kapla_B_PH	Yoruba	Mbuti	-0.002	0.001	-1.091	20777	20844	425476
Paniya	Kapla_B_PH	Ethiopia_4500BP_published.SG	Mbuti	-0.001	0.003	-0.365	21219	21263	425394
Paniya	Kapla_B_PH	BantuKenya	Mbuti	-0.002	0.002	-1.118	20889	20961	425476
Paniya	Kapla_PH	Luhya	Mbuti	-0.002	0.001	-1.288	20759	20828	425476

Paniya	Kapla_PH	Yoruba	Mbuti	-0.001	0.001	-1.087	20712	20764	425476
Paniya	Kapla_PH	Ethiopia_4500BP_published.SG	Mbuti	0.000	0.002	-0.037	21166	21170	425394
Paniya	Kapla_PH	BantuKenya	Mbuti	-0.002	0.001	-1.241	20823	20888	425476

Table S11: ANI-ASI admixture timings in select populations from India inferred using ALDER (fit started at d > 0.30 cM). Study populations that were sequenced in this study are highlighted in bold. Admixture times inferred by ALDER are shown both in generations as well as in years (assuming a generation time of 28 years per generation (Narasimhan et al 2019)). Language-family assignments are taken from (Nakatsuka et al. 2017; GenomeAsia100K Consortium 2019). Populations are arranged by decreasing admixture time (generations) within the language family. Asterisks indicate admixture times whose decay rates have a Z-score >= 2 (corresponding to p-value < 0.05).

Population Label	Sample size	Language family	Admixture time (generations)	Admixture time SE (generations)	Admixture time (years)	Admixture time SE (years)
Paniya	11	Dravidian	195.93*	48.75	5486.04	1365.00
Ulladan	6	Dravidian	153.93*	36.30	4310.04	1016.40
Toda	12	Dravidian	151.72*	15.89	4248.16	444.92
Vysya	39	Dravidian	150.67*	12.06	4218.76	337.68
Arunthatiar	18	Dravidian	137.50*	16.17	3850.00	452.76
Palliyar	11	Dravidian	132.11*	27.54	3699.08	771.12
Iyangar	6	Dravidian	130.93*	22.45	3666.04	628.60
Adi_Dravider	7	Dravidian	128.96*	18.90	3610.88	529.20
Chakkiliyan	9	Dravidian	119.54*	27.04	3347.12	757.12
Yadav_Pondicherry	12	Dravidian	114.18*	20.32	3197.04	568.96
Kodava_US	78	Dravidian	110.93*	9.23	3106.04	258.44
Iyer	13	Dravidian	110.25*	9.28	3087.00	259.84
Panta_Kapu	15	Dravidian	109.28*	20.01	3059.84	560.28
ITU	25	Dravidian	108.32*	9.61	3032.96	269.08
Bunt	11	Dravidian	106.29*	8.65	2976.12	242.20
Paravar	7	Dravidian	103.74*	15.08	2904.72	422.24
Nair	44	Dravidian	101.15*	9.99	2832.20	279.72
STU	8	Dravidian	98.39*	22.93	2754.92	642.04
Yerukali	6	Dravidian	98.01*	21.14	2744.28	591.92
Handigodu	6	Dravidian	95.41*	16.41	2671.48	459.48
UrbanBangalore	34	Dravidian	94.44*	6.05	2644.32	169.40
Kodava	15	Dravidian	94.28*	6.27	2639.84	175.56
UrbanChennai	34	Dravidian	91.90*	6.74	2573.20	188.72
Kapla_A	5	Dravidian	21.46	269.40	600.88	7543.20
Kapla_B	3	Dravidian	5.03*	1.99	140.84	55.72

Mahar	19	Indo-European	117.27*	13.70	3283.56	383.60
Khatri	14	Indo-European	114.22*	17.60	3198.16	492.80
SaryupariBrahmin	13	Indo-European	111.94*	9.73	3134.32	272.44
SourasthraBrahmin	9	Indo-European	111.10*	14.13	3110.80	395.64
Agarwal	36	Indo-European	110.25*	6.59	3087.00	184.52
Brahmin_Tiwari	15	Indo-European	107.94*	11.05	3022.32	309.40
WestbengalBrahmin	10	Indo-European	106.25*	9.71	2975.00	271.88
Brahmin_Vaidik	25	Indo-European	106.20*	7.32	2973.60	204.96
Brahmin_Catholic_Mangalore	6	Indo-European	102.37*	15.31	2866.36	428.68
Brahmin_Catholic_Kumta	10	Indo-European	101.19*	14.38	2833.32	402.64
Patel	7	Indo-European	96.22*	18.89	2694.16	528.92
Brahmin_Catholic_Goa	14	Indo-European	94.55*	9.79	2647.40	274.12
Gujjar	23	Indo-European	93.56*	7.49	2619.68	209.72
Sindhi	30	Indo-European	92.01*	6.29	2576.28	176.12
Rajput	17	Indo-European	89.79*	12.70	2514.12	355.60
Lambada	11	Indo-European	86.65*	16.62	2426.20	465.36
Bhumihar_Bihar	7	Indo-European	82.90*	13.07	2321.20	365.96
Pathan	33	Indo-European	66.99*	9.16	1875.72	256.48
Bhil	8	Indo-European	65.30*	10.13	1828.40	283.64
Sikh_Jatt	44	Indo-European	59.86*	5.78	1676.08	161.84

Table S15: Haplotype and nucleotide diversity on mtDNA across matrilocal and patrilocal groups in Southwest India. For each study population with available mitogenome data, we computed the number of unique haplotypes (h) and haplotype diversity and its standard error based on 100 bootstrap resampling iterations (HD, HD_SD) and the overall nucleotide diversity as a per-basepair rate (pi) using the software pixy (Korunes & Samuk 2021). See Methods for further details on calculations.

Population	N	h	HD	HD_SD	pi
Bunt	11	6	0.836	0.045	0.006
Kapla	6	3	0.733	0.080	0.003
Kodava	15	8	0.924	0.049	0.003
Kodava_US	105	62	0.987	0.003	0.001
Nair	44	38	0.978	0.012	0.003