The American Journal of Human Genetics, Volume *93*

Supplemental Data

Genetic Evidence for Recent

Population Mixture in India

Priya Moorjani, Kumarasamy Thangaraj, Nick Patterson, Mark Lipson, Po-Ru Loh, Periyasamy Govindaraj, Bonnie Berger, David Reich, and Lalji Singh

Table of Contents

Figure S1. Historical relationships assumed for *F_4 Ratio Estimation*

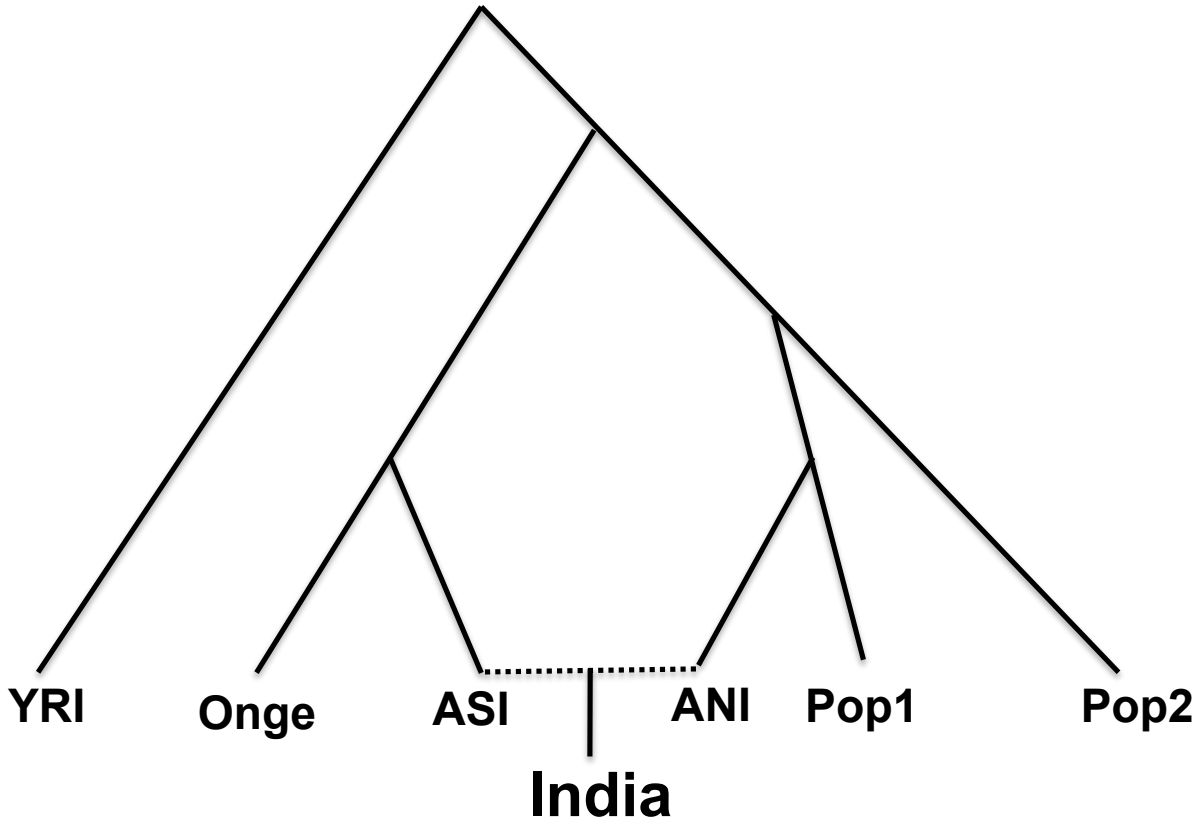
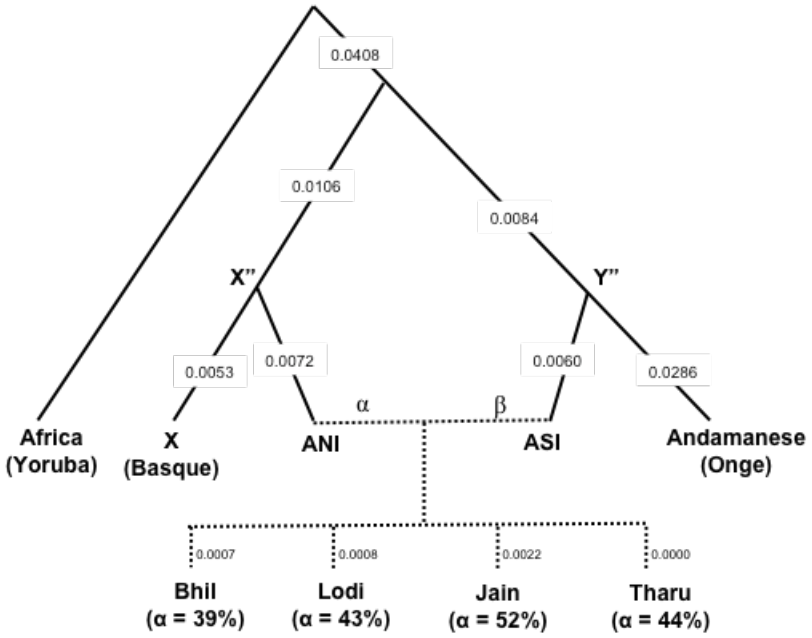


Figure S2. *admixture graph* fitted models of Indian history. We fit a model population relationships to the Indo-European and Dravidian rank 1 sets using *admixture graph*. We use the Affymetrix dataset (*n* = 210,482 SNPs). The drift lengths shown below were estimated using *X* = Basque.

(a) Indo-European rank 1 set



(b) Dravidian rank 1 set

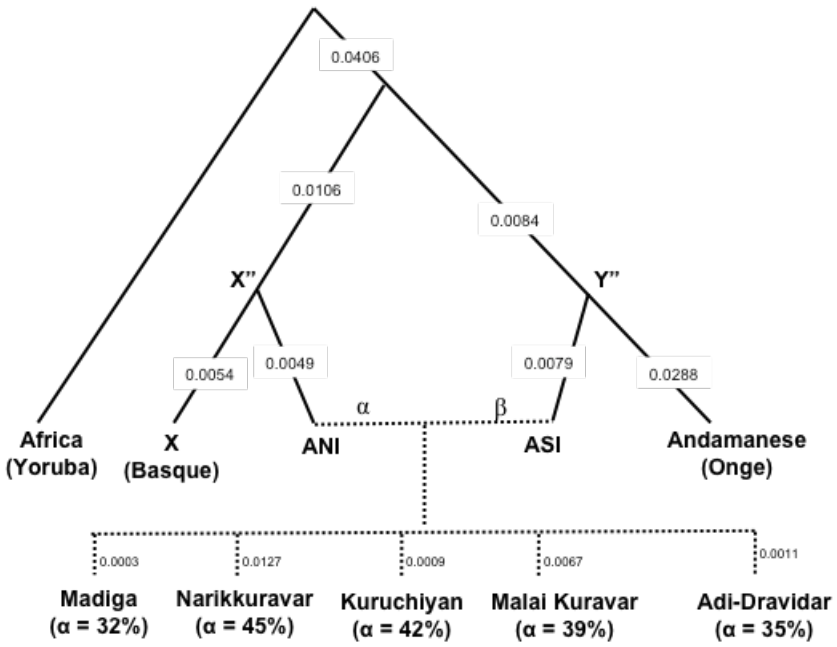
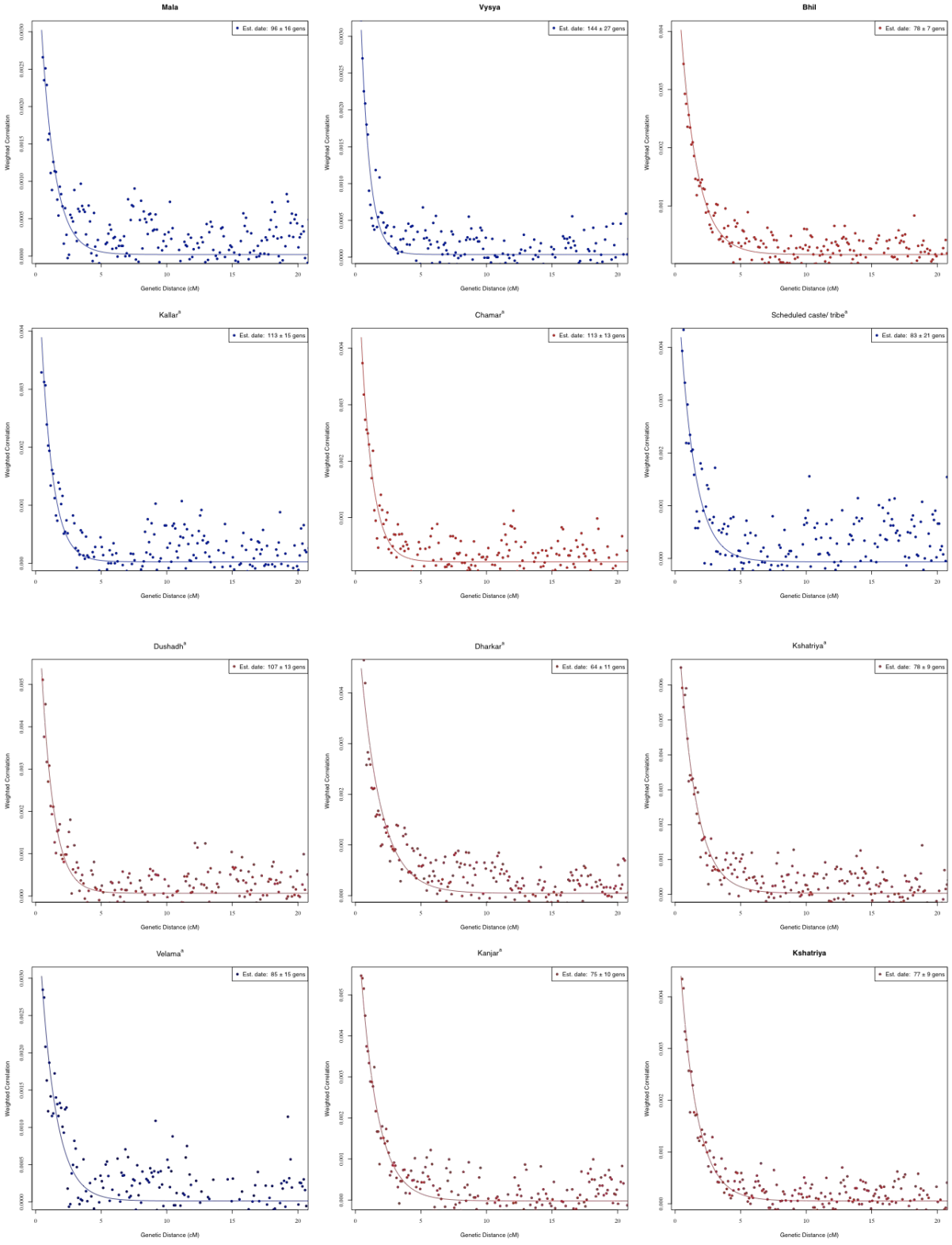
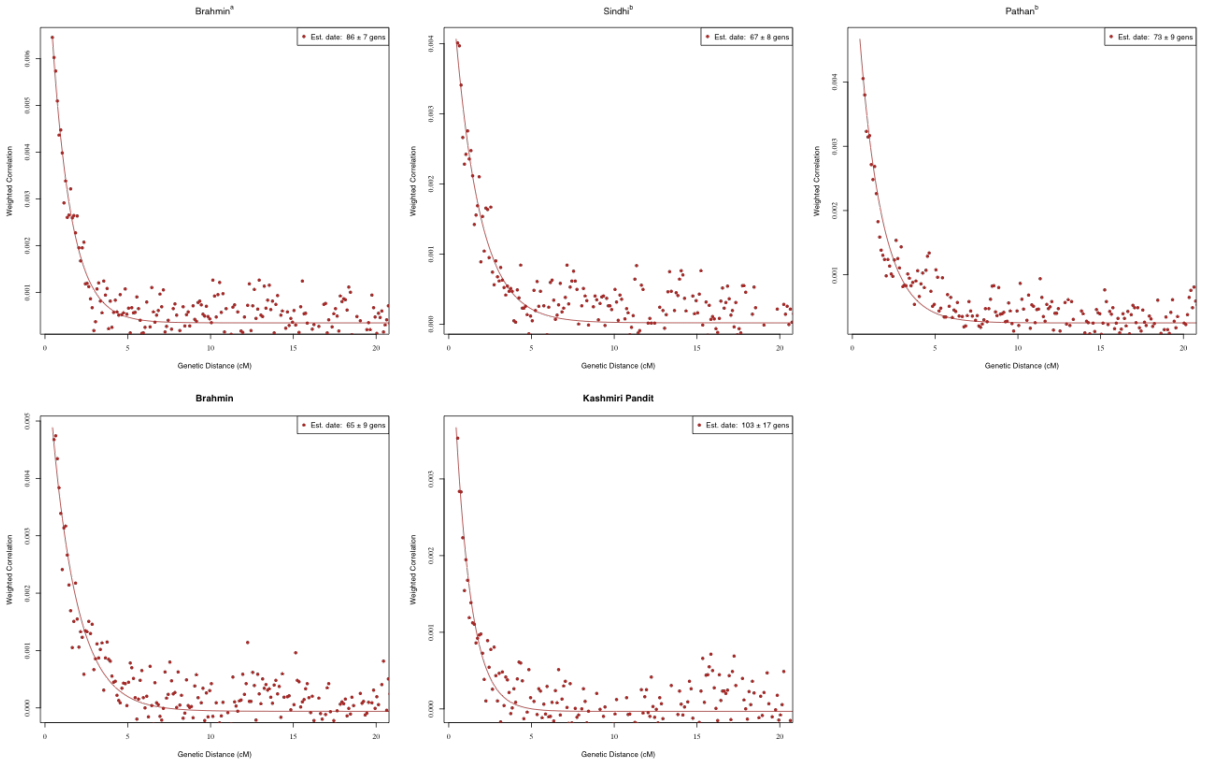


Figure S3. *rolloff* curves for each Indian group.
(a) Full dataset. We use the full Affymetrix (494,863 SNPs) or Illumina (500,703 SNPs) dataset to increase precision. We run *rolloff* with weights computed by performing Principal Component Analysis (PCA) on data from all populations on the Indian-cline and CEU (excluding the test population). We ignore inter-SNP distances less than 0.5 cM to avoid confounding by background LD. The output is colored based on the linguistic affiliation of the group. Standard errors were computed by weighted block jackknife (see Material and Methods).





***(b) Rank 1* groups with a simple history of ANI-ASI admixture.** We performed *rolloff* analysis for *rank 1* groups, computing PCA based SNP loadings for Basque and Indian cline groups (not including the target admixed groups for computing the weights) with data for 210,482 SNPs. We ignore inter-SNP distances less than 0.5 cM to avoid confounding by background LD.

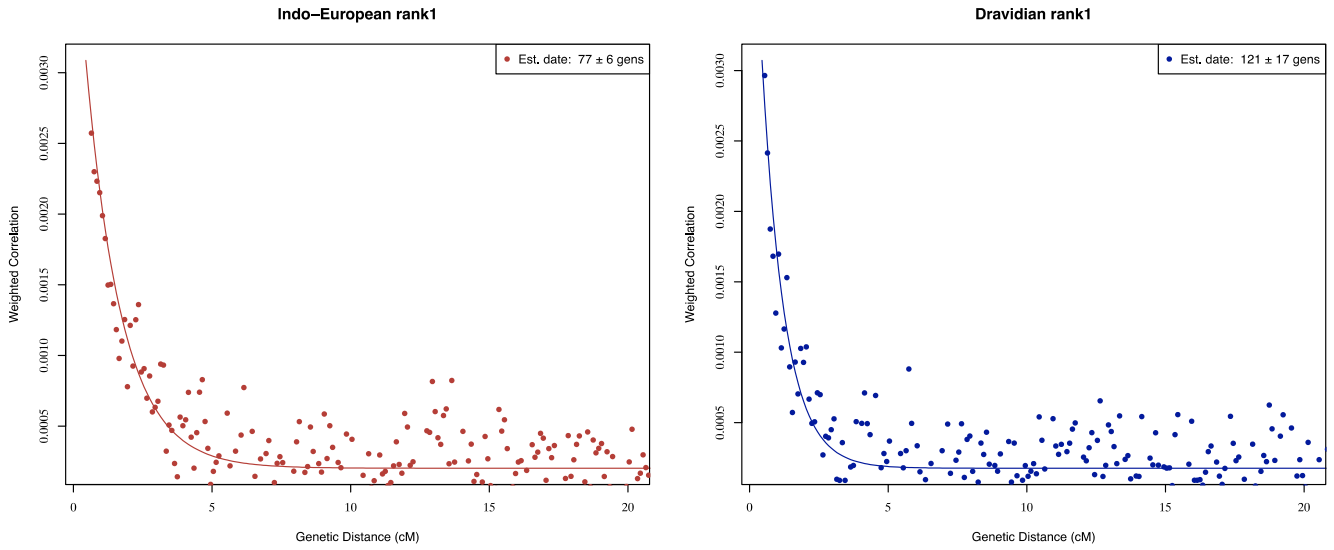


Figure S4. Distribution of nominal p-values in simulations of the likelihood ratio test. We performed 100,000 simulations based on the null model of a single pulse of mixture with noise. We use least squares to fit a null model of a single pulse of mixture ($y = Ae^{-nd} + c$) and an alternative model of two pulses of mixture ($y = Ae^{-n_1 d} + Be^{-n_2 d} + c$), where n, n_1, n_2 are parameters capturing the times since mixture, and d is the genetic distance. We performed a likelihood ratio test that is χ^2 distributed with 2 degrees of freedom and plotted the distribution of the nominal p-values. We computed the observed tail (y) as the proportion of observed p-values that are less than or equal to the theoretical p-values (x), normalized by the total number of simulations. The values below $-\log_{10}(0.05)$ are not shown. The dotted line indicates the regression line for the linear model between $\log_{10}(y)$ and $\log_{10}(x)$.

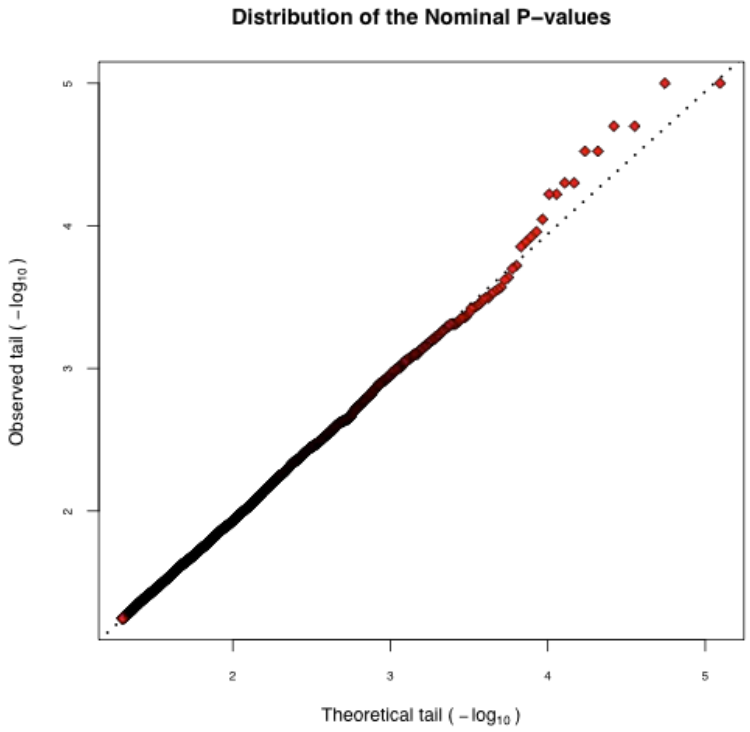


Figure S5. Phylogenetic relationships of simulated populations Pop1-15. Panel (a) shows the phylogenetic relationships of Pop1-15, and (b) shows PCA of Pop2-15. SNPs were ascertained in Pop1 and hence this population is not included in the PCA.

(a) ***(b)***

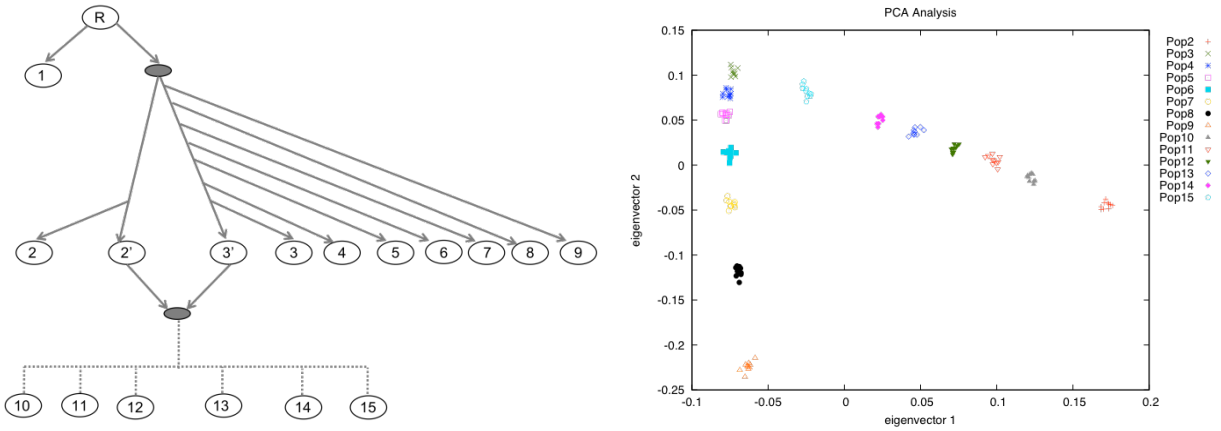


Figure S6. *Admixture graph* fitted to simulated data used in *ALDER* analysis. We simulated data for three groups (Sim-group1, Sim-group2, Sim-group3) using phased data from HGDP Han and HapMap CEU samples, where the admixed populations have a proportion α of ancestry from CEU. We fit a model of population relationships using *admixture graph.* The drift lengths shown below are based on simulated data where Sim-group1 ($\alpha = 30\%$), Sim-group2 ($\alpha = 50\%$) and Sim-group3 ($\alpha = 70\%$).

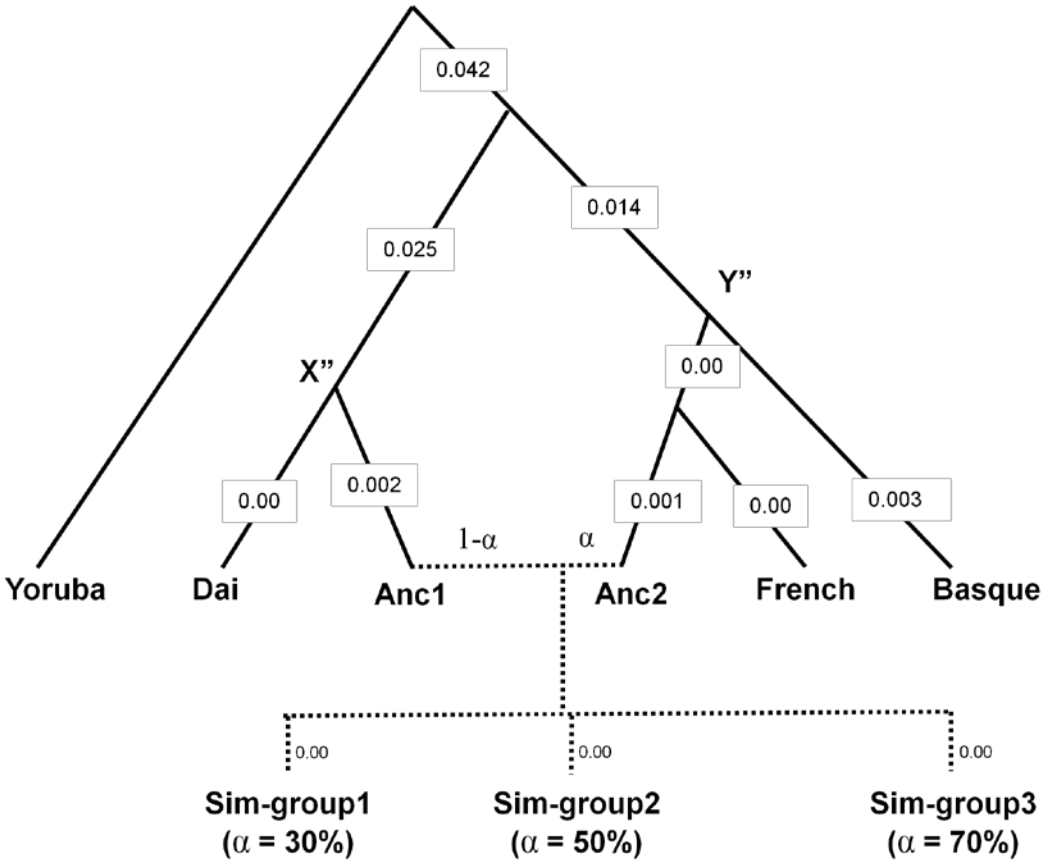


Table S1: Data curation

Pop	Dataset	Samples removed pre-PCA curation	Samples removed post-PCA curation	Total count (Post-curation)	Exclusion Criteria[c]
Adi-Dravidar	This study			5	
Bhil	Reich 2009 & this study			17	
Bhumij	This study		5	0	(1)
Birhor	This study	1	4	0	(a); (1)
Brahmin	This study	5		10	(b)
Changpa	This study		5	0	(1)
Gond	This study	1	14	0	(a); (1)
Ho	This study		5	0	(1)
Irula	This study		5	0	(2)
Jain	This study			5	
Jews	This study		5	0	(2)
Kallar	This study			5	
Kattunayakan	This study			5	
Korku	This study	1	4	0	(a); (1)
Kshatriya	This study	5		15	(b)
Kuruchiyan	This study			5	
Gounder	This study			5	
Madiga	Reich 2009 & this study	5	1	13	(b); (3)
Malai Kuravar	This study			5	
Mala	Reich 2009 & this study	5		13	(b)
Mali	This study			5	
Minicoy	This study		1	4	(3)
Munda	This study		5	0	(1)
Narikkuravar	This study			5	
Palliyar	This study			5	
Kashmiri Pandit	Reich 2009 & this study	5		15	(b)
Paniya	This study			5	
Sherpa	This study		5	0	(1)
Siddi	Reich 2009 & this study	2	14	0	(a); (1)[1; 2]
Subba	This study		5	0	(1)
Tibet-refugees	This study		5	0	(1)
Vedda	This study			4	
Vysya	Reich 2009 & this study	5	1	14	(b); (3)
Tharu	Reich 2009		4	5	(3)
Meghawal	Reich 2009			5	
Chenchu	Reich 2009			6	
Kurumba	Reich 2009		9	0	(2)
Hallaki	Reich 2009		7	0	(2)
Santhal	Reich 2009		7	0	(1)
Kharia	Reich 2009		6	0	(1)
Vaish	Reich 2009			4	
Srivastava	Reich 2009			2	
Naidu	Reich 2009			4	
Velama	Reich 2009			4	
Sahariya	Reich 2009		4	0	(1)
Lodi	Reich 2009			5	

Satnami	Reich 2009		1	3	(3)
Kamsali	Reich 2009			4	
Onge	Reich 2009			9	
Great_Andamanese	Reich 2009		7	0	(1)[2]
Nyshi	Reich 2009		4	0	(1)
Ao Naga	Reich 2009		4	0	(1)
Brahmin[a]	Metspalu 2011			8	
Kanjar[a]	Metspalu 2011	1		8	(b)
Chamar[a]	Metspalu 2011			10	
Dushadh[a]	Metspalu 2011	3		7	(b)
Kshatriya[a]	Metspalu 2011			7	
Kol[a]	Metspalu 2011		16	0	(1)
Dharkar[a]	Metspalu 2011	1		11	(b)
Muslim[a]	Metspalu 2011			5	
Scheduled caste[a]	Metspalu 2011		6	0	(2)
central_mix1_nihali	Metspalu 2011		5	0	(2)
Gond[a]	Metspalu 2011		4	0	(2)
Munda[a]	Metspalu 2011		1	0	(1)
Scheduled caste/ tribe[a]	Metspalu 2011			6	
Hakkipikki[a]	Metspalu 2011			4	
Ao Naga[a]	Metspalu 2011		4	0	(1)
Chenchu[a]	Metspalu 2011			4	
Kallar[a]	Metspalu 2011			8	
Velama[a]	Metspalu 2011	1		9	(a)
Palliyar[a]	Metspalu 2011			5	
Sindhi[b]	Li 2008		14	10	(1)[2]
Pathan[b]	Li 2008		7	15	(1)[2]

[c]Samples were removed based on the following exclusion criteria:
Pre-PCA curation:
(a) Remove all duplicate samples: for each pair for samples that have >90% genotype matching, remove one sample.
(b) Remove all related samples: In the case of trios, the child was excluded and in case of first-degree relatives, one sample from the pair was excluded.
(c) Remove all samples previously excluded in Metspalu et al (2011): 8 samples were excluded (not shown in table above)[3].

Post-PCA curation (Figure 1):
(1) Remove samples and groups that have evidence of recent ancestry from groups other than ANI and ASI based on PCA.
(2) Remove groups that are not homogenous in PCA.
(3) Remove samples that do not cluster with the majority of samples from their group.

Table S2: Summary of *D*-statistics

Population (X)	*n*	Language Group	Social/ caste group	Pop with highest *D*-statistic mean	Pop with 2nd highest *D*-statistic (Z-score)	Pop with 3rd highest *D*-statistic (Z-score)
Paniya	5	Dravidian	Tribal	Georgian	Armenian (0.1)	Cypriot (0.3)
Palliyar	5	Dravidian	Tribal	Georgian	Kurd (0.1)	Abhkasian (0.2)
Kattunayakan	5	Dravidian	Tribal	Georgian	Kurd (0.2)	Armenian (0.4)
Palliyar[a]	5	Dravidian	Lower caste	Cypriot	Abhkasian (0.4)	Georgian (0.5)
Madiga	13	Dravidian	Lower caste	Georgian	Lezgin (1.1)	Abhkasian (1.2)
Mala	13	Dravidian	Lower caste	Georgian	Abhkasian (0.5)	Armenian (1.2)
Adi-Dravidar	5	Dravidian	Lower caste	Georgian	Abhkasian (1.1)	Armenian (1.5)
Hakkipikki[a]	4	Dravidian	Tribal	Georgian	Armenian (1.2)	Abhkasian (1.3)
Vedda	4	Indo-European	Tribal	Georgian	Abhkasian (0.1)	Kurd (0.4)
Kamsali	4	Dravidian	Lower caste	Georgian	Armenian (0.8)	Lezgin (0.6)
Chenchu[a]	4	Dravidian	Tribal	Georgian	Abhkasian (1.4)	Armenian (2)
Chamar[a]	10	Indo-European	Tribal	Georgian	Abhkasian (0.6)	Lezgin (1.1)
Chenchu	6	Dravidian	Tribal	Georgian	Abhkasian (0.7)	Armenian (1.2)
Bhil	17	Indo-European	Tribal	Georgian	Abhkasian (0.4)	Armenian (1.0)
Kallar	5	Dravidian	Lower caste	Georgian	Armenian (0.9)	Cypriot (1.0)
Kallar[a]	8	Dravidian	Tribal	Georgian	Abhkasian (1.2)	Lezgin (2.0)
Vysya	14	Dravidian	Middle caste	Georgian	Abhkasian (1.1)	Armenian (1.7)
Malai Kuravar	5	Dravidian	Tribal	Georgian	Abhkasian (0.4)	Armenian (1.5)
Satnami	3	Indo-European	Lower caste	Georgian	Abhkasian (0.2)	Tuscan (0.8)
Kuruchiyan	5	Dravidian	Tribal	Georgian	Abhkasian (1.1)	Armenian (1.7)
Dushadh[a]	7	Indo-European	Lower caste	Georgian	Abhkasian (0.2)	Lezgin (1.1)
Scheduled caste/ tribe[a]	6	Dravidian	Lower caste	Georgian	Kurd (0.6)	Abhkasian (0.8)
Mali	5	Dravidian	Lower caste	Georgian	Armenian (1.3)	Abhkasian (1.4)
Minicoy	4	Indo-European	Lower caste	Georgian	Abhkasian (2.3)	Cypriot (2.3)
Gounder	5	Dravidian	Middle caste	Georgian	Kurd (0.6)	Abhkasian (0.9)
Lodi	5	Indo-European	Lower caste	Georgian	Armenian (2)	Abhkasian (2.0)
Naidu	4	Dravidian	Upper caste	Georgian	Armenian (1.3)	Abhkasian (1.1)
Velama	4	Dravidian	Upper caste	Georgian	Abhkasian (1.3)	Armenian (2.8)
Velama[a]	9	Dravidian	Upper caste	Georgian	Armenian (1)	Abhkasian (0.9)
Narikkuravar	5	Dravidian	Tribal	Georgian	Abhkasian (0.5)	Cypriot (1.3)
Tharu	5	Indo-European	Tribal	Georgian	Lezgin (0.8)	Abhkasian (1.1)
Dharkar[a]	11	Indo-European	Nomadic group	Georgian	Tuscan (1.0)	Armenian (2.0)
Kanjar[a]	8	Indo-European	Nomadic group	Georgian	Abhkasian (1.2)	Tuscan (0.8)
Muslim[a]	5	Indo-European	Religious group	Georgian	Abhkasian (1.2)	Armenian (1.9)
Srivastava	2	Indo-European	Upper caste	Georgian	Abhkasian (-0.3)	Cypriot (0.1)
Jain	5	Indo-European	Religious group	Georgian	Abhkasian (0.8)	Lezgin (1.2)
Meghawal	5	Indo-European	Lower caste	Georgian	Abhkasian (0.8)	Cypriot (1.3)
Kshatriya[a]	7	Indo-European	Upper caste	Georgian	Abhkasian (1.4)	Lezgin (1.9)
Vaish	4	Indo-European	Upper caste	Georgian	Lezgin (1.0)	Armenian (2.2)
Brahmin[a]	8	Indo-European	Upper caste	Tuscan	Lezgin (0.0)	Georgian (0.1)
Kshatriya	15	Indo-European	Upper caste	Georgian	Abhkasian (0.9)	Tuscan (0.8)
Brahmin	10	Indo-European	Upper caste	Georgian	Tuscan (1.1)	Lezgin (1.6)
Sindhi[b]	10	Indo-European	Urban group	Georgian	Armenian (2.6)	Abhkasian (2.4)
Kashmiri Pandit	15	Indo-European	Upper caste	Georgian	Abhkasian (2.0)	Armenian (2.6)
Pathan[b]	15	Indo-European	Urban group	Georgian	Armenian (2.1)	Abhkasian (1.6)

We compute *D*(Onge, *X*; YRI, *Y*) where *X* is an Indian group shown above and *Y* is a West Eurasian group chosen from a panel of 43 groups including Europeans, Central Asians, Middle Easterners and Caucasian populations. We display the results for the population *Y* with the highest *D*-statistic mean, the 2nd highest *D*-statistic mean (*Z*-score for the difference between highest and 2nd highest group), and the 3nd highest *D*-statistic mean (*Z*-score for the difference between the highest and 3nd highest). We consider $|Z| > 3$ to be statistically significant. [a] indicates samples from Metspalu et al (2011) and [b] indicates samples from HGDP.

Table S3 (A): *D*-statistic differences: Madiga

Population (Y)	*n*	Sampling Location	Geographic Group	Mean difference in *D*-statistics	Z-score of difference in *D*-statistics
Lezgin	18	Caucasus	Caucasus	0.0001	1.1
Abhkasian	20	Caucasus	Caucasus	0.0001	1.2
Armenian	35	Armenia	Caucasus	0.0002	2.0
Kurd	6	Kazakhstan	Central Asia	0.0002	1.3
Chechen	20	Caucasus	Caucasus	0.0002	2.1
Cypriot	12	Cyprus	Europe	0.0003	2.2
Iranian	20	Iran	Central Asia	0.0003	2.8
Druze	42	Israel	Near East	0.0004	3.4
Syrian	16	Syria	Near East	0.0004	3.3
Adygei	17	Caucasus	Caucasus	0.0004	3.4
Tuscan	7	Italy	Europe	0.0004	3.0
North Ossetian	15	Russia	Caucasus	0.0004	3.5
TSI	87	Italy	Europe	0.0005	4.6
Lebanese	7	Lebanon	Near East	0.0005	3.1
Turk	19	Turkey	Near East	0.0005	4.7
Basque	24	France	Europe	0.0005	3.9
Kumyk	14	Russia	Caucasus	0.0005	4.5
CEU	110	United States	Europe	0.0005	5.0
Orcadian	15	United Kingdom	Europe	0.0005	3.9
Italian	12	Italy	Europe	0.0006	4.1
Jordanian	19	Jordania	Near East	0.0006	4.9
Hungarian	20	Hungary	Europe	0.0006	4.5
French	28	France	Europe	0.0006	5.0
Lithuanian	10	Lithuania	Europe	0.0006	3.9
Spaniard	12	Spain	Europe	0.0006	4.6
Bulgarian	13	Bulgaria	Europe	0.0006	5.0
Balkar	19	Caucasus	Caucasus	0.0006	5.8
Ukranian	20	Ukraine	Europe	0.0006	5.3
Palestinian	46	Israel	Near East	0.0007	6.1
Romanian	16	Romania	Europe	0.0007	5.6
Sardinian	28	Italy	Europe	0.0007	5.4
Saudi	19	Saudi Arabia	Near East	0.0007	5.2
Bedouin	45	Israel	Near East	0.0008	6.2
Belorussian	9	Belorussia	Europe	0.0008	5.5
Mordovian	15	Russia	Europe	0.0010	7.7
Russian	27	Russia	Europe	0.0011	9.0
Tajik	15	Tajikstan	Central Asia	0.0011	8.5
Yemenese	10	Yemen	Near East	0.0013	8.3
Turkmen	15	Turkmenistan	Central Asia	0.0016	11.3
Nogai	16	Russia	Caucasus	0.0017	13.3
Chuvash	17	Russia	Europe	0.0020	13.1
Uzbek	15	Uzbekstan	Central Asia	0.0032	19.7

We compare *D*(Onge, Madiga; YRI, Georgian) = 0.0335 (*Z* = 16.7) with *D*(Onge, Madiga; YRI, *Y*) where *Y* is any West Eurasian group chosen from a panel of 42 groups including Europeans, Central Asians, Middle Easterners and Caucasian populations. For each *Y*, we display the difference in mean and Z-score of *D*-statistics (i.e. *D*(Onge, Madiga; YRI, Georgian) - *D*(Onge, Madiga; YRI, Georgian) and the corresponding *Z*-score for the difference).

Table S3 (B): *D*-statistic differences: Kashmiri Pandit

Population (X)	*n*	Sampling Location	Geographic Group	Mean difference in *D*-statistics	Z-score of difference in *D*-statistics
Abhkasian	20	Caucasus	Caucasus	0.0002	2.0
Lezgin	18	Caucasus	Caucasus	0.0003	2.2
Armenian	35	Armenia	Caucasus	0.0003	2.6
Cypriot	12	Cyprus	Europe	0.0004	2.8
Tuscan	7	Italy	Europe	0.0004	2.4
Chechen	20	Caucasus	Caucasus	0.0004	3.6
TSI	87	Italy	Europe	0.0005	4.9
Kurd	6	Kazakhstan	Central Asia	0.0005	3.0
Orcadian	15	United Kingdom	Europe	0.0005	3.6
CEU	110	United States	Europe	0.0005	4.8
Basque	24	France	Europe	0.0005	4.0
Italian	12	Italy	Europe	0.0005	3.8
Lithuanian	10	Lithuania	Europe	0.0005	3.5
French	28	France	Europe	0.0006	4.8
Druze	42	Israel	Near East	0.0006	5.4
Hungarian	20	Hungary	Europe	0.0006	5.0
Spaniard	12	Spain	Europe	0.0007	5.2
Bulgarian	13	Bulgaria	Europe	0.0008	5.8
Sardinian	28	Italy	Europe	0.0008	5.8
Adygei	17	Caucasus	Caucasus	0.0008	6.4
Ukranian	20	Ukraine	Europe	0.0008	6.3
Belorussian	9	Belorussia	Europe	0.0009	5.4
North Ossetian	15	Russia	Caucasus	0.0009	7.0
Syrian	16	Syria	Near East	0.0009	7.5
Romanian	16	Romania	Europe	0.0010	7.4
Turk	19	Turkey	Near East	0.0010	8.6
Iranian	20	Iran	Central Asia	0.0010	8.4
Balkar	19	Caucasus	Caucasus	0.0010	9.2
Lebanese	7	Lebanon	Near East	0.0010	6.4
Kumyk	14	Russia	Caucasus	0.0011	8.8
Jordanian	19	Jordania	Near East	0.0012	10.1
Palestinian	46	Israel	Near East	0.0013	11.1
Bedouin	45	Israel	Near East	0.0014	11.2
Mordovian	15	Russia	Europe	0.0014	10.6
Saudi	19	Saudi Arabia	Near East	0.0014	10.6
Russian	27	Russia	Europe	0.0015	12.2
Tajik	15	Tajikstan	Central Asia	0.0026	18.5
Yemenese	10	Yemen	Near East	0.0030	17.3
Nogai	16	Russia	Caucasus	0.0031	22.1
Turkmen	15	Turkmenistan	Central Asia	0.0031	20.2
Chuvash	17	Russia	Europe	0.0032	20.7
Uzbek	15	Uzbekstan	Central Asia	0.0057	31.6

We compare *D*(Onge, Kashmiri Pandit; YRI, Georgian) = 0.0627 (Z = 29.7) with *D*(Onge, Kashmiri Pandit; YRI, *Y*) where *Y* is any West Eurasian group chosen from a panel of 42 groups including Europeans, Central Asians, Middle Easterners and Caucasian groups. For each *Y*, we display the difference in mean and Z-score of *D*-statistics (*D*(Onge, Kashmiri Pandit; YRI, Georgian) - *D*(Onge, Kashmiri Pandit; YRI, Georgian) and the corresponding *Z*-score for the difference).

Table S4: Ancestry estimates from *F_4 Ratio Estimation*

Population (X)	*n*	Language Group	Social/ caste group	§ANI Ancestry (Pop2 = Basque)	§ANI Ancestry (Pop2=Abhkasian)	§ANI ancestry (Reich 09)
Paniya	5	Dravidian	Tribal	16.7 ± 2.4	16.8 ± 2.1	22.5 ± 1.6
Palliyar	5	Dravidian	Tribal	21.2 ± 2.3	22.8 ± 2	29.1 ± 1.4
Kattunayakan	5	Dravidian	Tribal	24.6 ± 2.1	25.1 ± 1.9	30.8 ± 1.5
Palliyar[a]	5	Dravidian	Lower caste	24.2 ± 2.4	25.6 ± 2.1	31.4 ± 1.5
Madiga	13	Dravidian	Lower caste	32 ± 1.7	33.1 ± 1.5	40.6 ± 1.1
Mala	13	Dravidian	Lower caste	34.3 ± 1.7	35.8 ± 1.5	39.9 ± 1.1
Adi-Dravidar	5	Dravidian	Lower caste	34.7 ± 2	35.7 ± 1.7	40.9 ± 1.3
Hakkipikki[a]	4	Dravidian	Tribal	36.2 ± 2	35.4 ± 1.8	40.8 ± 1.4
Vedda	4	Indo-European	Tribal	36 ± 2.5	38 ± 2.2	41.3 ± 1.6
Kamsali	4	Dravidian	Lower caste	36.5 ± 2.1	38 ± 1.8	43.1 ± 1.4
Chenchu[a]	4	Dravidian	Tribal	37.2 ± 2.1	38.1 ± 1.9	43.4 ± 1.4
Chamar[a]	10	Indo-European	Tribal	38.7 ± 1.7	38.5 ± 1.5	43.1 ± 1.1
Chenchu	6	Dravidian	Tribal	39 ± 2.2	38.4 ± 2	41.7 ± 1.4
Bhil	17	Indo-European	Tribal	38.9 ± 1.6	39.3 ± 1.4	45.8 ± 1
Kallar	5	Dravidian	Lower caste	37.3 ± 2.1	39.4 ± 1.8	44.5 ± 1.3
Kallar[a]	8	Dravidian	Tribal	37.7 ± 1.8	40.4 ± 1.5	47.1 ± 1.1
Vysya	14	Dravidian	Middle caste	37.9 ± 1.8	41.2 ± 1.5	47.2 ± 1.1
Malai Kuravar	5	Dravidian	Tribal	38.8 ± 2.1	41.2 ± 1.9	46.8 ± 1.3
Satnami	3	Indo-European	Lower caste	40.7 ± 2.1	40.8 ± 1.9	43.4 ± 1.4
Kuruchiyan	5	Dravidian	Tribal	41.9 ± 1.9	43.2 ± 1.7	48.6 ± 1.2
Dushadh[a]	7	Indo-European	Lower caste	41 ± 1.8	42.8 ± 1.6	48.2 ± 1.2
Scheduled caste/ tribe[a]	6	Dravidian	Lower caste	40.5 ± 1.9	43.5 ± 1.6	48.8 ± 1.2
Mali	5	Dravidian	Lower caste	44 ± 2	43.3 ± 1.8	53.1 ± 1.2
Minicoy	4	Indo-European	Lower caste	42.9 ± 2	43.1 ± 1.7	48.9 ± 1.3
Gounder	5	Dravidian	Middle caste	42.9 ± 1.9	45.8 ± 1.7	51.8 ± 1.2
Lodi	5	Indo-European	Lower caste	43.1 ± 1.9	43.4 ± 1.7	49.4 ± 1.2
Naidu	4	Dravidian	Upper caste	43.2 ± 2	44.3 ± 1.8	48.5 ± 1.3
Velama	4	Dravidian	Upper caste	42.7 ± 2	46.3 ± 1.7	53.9 ± 1.3
Velama[a]	9	Dravidian	Upper caste	43.4 ± 1.7	45.3 ± 1.5	51.1 ± 1.1
Narikkuravar	5	Dravidian	Tribal	45 ± 2.2	46.1 ± 1.9	50.5 ± 1.5
Tharu	5	Indo-European	Tribal	43.6 ± 1.9	43.3 ± 1.7	50.5 ± 1.2
Dharkar[a]	11	Indo-European	Nomadic group	47.8 ± 1.5	47.3 ± 1.3	54.6 ± 1
Kanjar[a]	8	Indo-European	Nomadic group	48.2 ± 1.7	47.1 ± 1.5	53.5 ± 1.1
Muslim[a]	5	Indo-European	Religious group	49.4 ± 1.8	49.4 ± 1.5	55.1 ± 1.2
Srivastava	2	Indo-European	Upper caste	52.3 ± 2.5	51.6 ± 2.2	56.4 ± 1.5
Jain	5	Indo-European	Religious group	51.6 ± 1.9	52.1 ± 1.7	58 ± 1.2
Meghawal	5	Indo-European	Lower caste	53.6 ± 1.8	53.2 ± 1.6	58.2 ± 1.1
Kshatriya[a]	7	Indo-European	Upper caste	54.6 ± 1.6	53 ± 1.4	60.7 ± 0.9
Vaish	4	Indo-European	Upper caste	56.5 ± 1.7	54.5 ± 1.5	60.1 ± 1.2
Brahmin[a]	8	Indo-European	Upper caste	61.2 ± 1.4	57.8 ± 1.3	63.9 ± 0.9
Kshatriya	15	Indo-European	Upper caste	60.9 ± 1.3	58.4 ± 1.2	63.6 ± 0.8
Brahmin	10	Indo-European	Upper caste	62.8 ± 1.4	59.2 ± 1.3	64.5 ± 0.9
Sindhi[b]	10	Indo-European	Urban group	64.3 ± 1.3	62.7 ± 1.2	71.8 ± 0.8
Kashmiri Pandit	15	Indo-European	Upper caste	65.2 ± 1.3	63.8 ± 1.1	68.6 ± 0.8
Pathan[b]	15	Indo-European	Urban group	70.4 ± 1.2	67.9 ± 1	74.8 ± 0.7

We performed *F_4 Ratio Estimation* to estimate the proportion of ANI ancestry in Indians. Specifically, we use the following statistics: **§ANI ancestry (Pop2 = Basque)** = f_4(YRI, Basque; X, Onge)/ f_4(YRI, Basque; Georgian, Onge); **§ANI ancestry (Pop2 = Abhkasian)** = f_4(YRI, Abhkasian; X, Onge)/ f_4(YRI, Abhkasian; Georgian, Onge); **§ANI ancestry (Reich et al., 09)** = f_4(Adygei, Papuan; X, Onge)/ f_4(Adygei, Papuan; CEU, Onge). We computed standard errors using a Block Jackknife with a block size of 5cM. [a] indicates samples from Metspalu et al (2011) and [b] indicates samples from HGDP.

Table S5. Comparison of expected and observed weighted LD amplitudes for simulated data.

		Single wave				***Two waves***			
Europe %	**Ref. in *ALDER***	**Expected Amplitude x 10000**	**Observed Amplitude x 10000**	***Z***	α_{old}	**Expected Amplitude x 10000**	**Observed Amplitude x 10000**	***Z***	α_{old}
Simulation Set 1:									
30%	Basque	3.18 ± 0.13	3.30 ± 0.27	0.4	-0.8 ± 1.8	3.16 ± 0.15	1.76 ± 0.37	-3.7	11.3 ± 3.6
50%	Basque	1.83 ± 0.07	1.79 ± 0.23	-0.2	0.5 ± 3.1	1.81 ± 0.09	1.28 ± 0.12	-4.2	8.7 ± 2.3
70%	Basque	0.54 ± 0.03	0.56 ± 0.13	0.1	-0.6 ± 4.3	0.52 ± 0.04	0.11 ± 0.03	-14.2	36.6 ± 4.2
30%	Dai	0.60 ± 0.04	0.64 ± 0.14	0.3	-1.5 ± 4.5	0.58 ± 0.05	0.43 ± 0.12	-1.1	6.1 ± 5.6
50%	Dai	1.88 ± 0.09	1.70 ± 0.24	-0.8	2.6 ± 3.5	1.88 ± 0.10	1.16 ± 0.11	-4.1	11.9 ± 3.0
70%	Dai	3.04 ± 0.11	3.31 ± 0.31	0.8	-1.7 ± 2.0	3.07 ± 0.14	1.78 ± 0.23	-6.8	12.4 ± 2.7
Simulation Set 2:									
20%	Basque	3.36 ± 0.20	3.28 ± 0.23	-0.3	0.4 ± 1.5	3.45 ± 0.24	1.81 ± 0.23	-6.4	9.4 ± 1.5
30%	Basque	3.13 ± 0.18	3.37 ± 0.31	0.7	-1.7 ± 2.6	3.32 ± 0.25	1.88 ± 0.28	-3.6	10.9 ± 3.1
40%	Basque	2.55 ± 0.16	2.85 ± 0.31	0.9	-2.8 ± 3.2	2.68 ± 0.25	1.50 ± 0.24	-4.0	13.2 ± 3.4
20%	Dai	0.22 ± 0.03	0.10 ± 0.07	-1.7	10.8 ± 6.8	0.18 ± 0.04	0.22 ± 0.44	0.1	-3.3 ± 31.7
30%	Dai	0.63 ± 0.05	0.59 ± 0.09	-0.4	1.4 ± 3.5	0.51 ± 0.08	0.29 ± 0.09	-2.1	11.1 ± 5.2
40%	Dai	1.21 ± 0.09	1.29 ± 0.19	0.4	-1.5 ± 3.4	1.11 ± 0.16	0.92 ± 0.12	-1.2	4.5 ± 3.7

We simulated 14 diploid individuals with under two admixture models: (a) a single CEU-Han admixture event 100 generations ago, and (b) Two waves of CEU admixture into Han, 300 and 75 generations ago with the European admixture proportion (Europe %) shown above. We performed *ALDER* analysis with the reference population shown above. Standard errors shown are based on jackknife estimates from a single simulation (not standard errors from averaging over multiple simulations). To infer the statistical uncertainty of (Observed - Expected) amplitude, we use a weighted block jackknife dropping each chromosome in turn and repeating the entire procedure. This produces a standard error that allows us to test whether the difference is consistent with zero ($|Z| < 3$).

Table S6. Dates of admixture using PCA loadings and one reference group

Pop (X)	***n***	**PCA based weights (*rolloff*)**	**PCA based weights (*ALDER*)**	**One reference (*ALDER*)**
Indo-European rank 1 set	32	77 ± 6	70 ± 7	68 ± 12
Dravidian rank 1 set	33	121 ± 17	101 ± 17	105 ± 14

We performed *rolloff* and *ALDER* analysis using SNP loadings computed based on a PCA of Basque and all Indian cline groups (except the target (X)). We also performed *ALDER* analysis using Basque as one reference group. To remove the effects of LD in the ancestral populations, we ignore bins corresponding to genetic distance separation less than 0.5 cM: this threshold is set by *ALDER* after comparison of shared LD between Basque and the admixed groups.

Table S7. Simulations to test bias in estimated dates of admixture for demographic parameters relevant to Indian groups

Group (X)	Sam-ples	ANI ancestry%	Date of mixture estimated using *rolloff* from real data (gens)	Mean date estimate over 100 simulations
Brahmin	10	62.8 ± 1.4	65 ± 9	66
Mala	13	34.3 ± 1.7	96 ± 16	99
Pathan[b]	15	70.4 ± 1.2	73 ± 9	76
Dravidian rank 1 set	33	36.9 ± 1.5	121 ± 17	123
Indo-European rank 1 set	32	42.3 ± 1.4	77 ± 6	79

We simulated individuals of mixed European (CEU) and East Asian (CHB) ancestry where we set sample size, ANI ancestry proportion, and the date of admixture to match the parameters observed in the real data for each Indian group (X). We performed *rolloff* using French and Han as the reference groups and computed the average admixture date (in generations) for 100 simulations.

Table S8. Record of testing for consistency with simple ANI-ASI mixture

Set size	Sets tested	Sets that are rank 1	Sets also passing *admixture graph*
3	7,770	3,692	1,152
4	25,425	5,152	860
5	19,239	1,293	90
6	1,852	30	1

Table S9. Number of times each of 37 Indian groups is included in a rank 1 set

Population (X)	Linguistic affiliation	Traditional social status	Sampling Location	Set size = 3	Set size = 4	Set size = 5	Set size = 6	Set size = 7
Vysya	Dravidian	Middle caste	Andhra Pradesh	0	0	0	0	0
Sindhi[b]	Indo-European	Urban groups	Pakistan	2	0	0	0	0
Brahmin[a]	Indo-European	Upper caste	Uttar Pradesh	4	0	0	0	0
Chenchu	Dravidian	Tribal	Andhra Pradesh	8	0	0	0	0
Brahmin	Indo-European	Upper caste	Uttar Pradesh	15	0	0	0	0
Pathan[b]	Indo-European	Urban groups	Pakistan	19	2	0	0	0
Kshatriya	Indo-European	Upper caste	Uttar Pradesh	22	1	0	0	0
Mali	Dravidian	Lower caste	Lakshadweep	24	9	1	0	0
Kashmiri Pandit	Indo-European	Upper caste	Kashmir	33	10	0	0	0
Kshatriya[a]	Indo-European	Upper caste	Uttar Pradesh	40	13	0	0	0
Kanjar[a]	Indo-European	Nomadic group	Uttar Pradesh	48	23	0	0	0
Kallar[a]	Dravidian	Tribal	Tamil Nadu	50	81	20	1	0
Scheduled caste/ tribe[a]	Dravidian	Lower caste	Tamil Nadu	52	47	11	1	0
Gounder	Dravidian	Middle caste	Tamil Nadu	54	59	12	0	0
Velama[a]	Dravidian	Upper caste	Andhra Pradesh	62	71	11	0	0
Chamar[a]	Indo-European	Tribal	Uttar Pradesh	80	83	14	0	0
Dharkar[a]	Indo-European	Nomadic group	Uttar Pradesh	85	53	5	0	0
Kallar	Dravidian	Lower caste	Tamil Nadu	93	84	14	1	0
Narikkuravar	Dravidian	Nomadic group	Tamil Nadu	106	53	4	1	0
Paniya	Dravidian	Tribal	Kerela	107	81	3	0	0
Tharu	Indo-European	Tribal	Uttarkhand	118	91	10	0	0
Meghawal	Indo-European	Lower caste	Rajasthan	118	107	11	0	0
Bhil	Indo-European	Tribal	Gujarat	119	89	5	0	0
Madiga	Dravidian	Lower caste	Andhra Pradesh	123	112	8	0	0
Kattunayakan	Dravidian	Tribal	Kerela	123	82	10	0	0
Mala	Dravidian	Lower caste	Andhra Pradesh	123	167	24	0	0
Muslim[a]	Indo-European	Religious Group	Uttar Pradesh	124	87	5	0	0
Palliyar	Dravidian	Tribal	Tamil Nadu	124	99	15	0	0
Palliyar[a]	Dravidian	Lower caste	Tamil Nadu	128	81	8	0	0
Adi-Dravidar	Dravidian	Lower caste	Tamil Nadu	143	175	19	0	0
Dushadh*	Indo-European	Lower caste	Uttar Pradesh	158	169	15	0	0
Lodi	Indo-European	Lower caste	Uttar Pradesh	162	212	18	0	0
Malai Kuravar	Dravidian	Tribal	Tamil Nadu	164	195	29	1	0
Jain	Indo-European	Religious Group	Gujarat	169	193	32	0	0
Kuruchiyan	Dravidian	Tribal	Kerela	213	322	49	1	0
Onge	Jarawa-Onge	Hunter-gatherer	Andaman & Nicobar	215	311	53	0	0
Vedda	Indo-European	Tribal	Sri Lanka	227	278	44	0	0

[a] indicates samples from Metspalu et al (2011) and [b] indicates groups from HGDP.

References

1. Shah, A.M., Tamang, R., Moorjani, P., Rani, D.S., Govindaraj, P., Kulkarni, G., Bhattacharya, T., Mustak, M.S., Bhaskar, L., and Reddy, A.G. (2011). Indian siddis: African descendants with Indian admixture. The American Journal of Human Genetics 89, 154-161.
2. Reich, D., Thangaraj, K., Patterson, N., Price, A., and Singh, L. (2009). Reconstructing Indian population history. Nature 461, 489-494.
3. Metspalu, M., Romero, I.G., Yunusbayev, B., Chaubey, G., Mallick, C.B., Hudjashov, G., Nelis, M., Mägi, R., Metspalu, E., and Remm, M. (2011). Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. The American Journal of Human Genetics 89, 731-744.