Supplementary Information Guide

Reconstructing Indian Population History

Reich D, Thangaraj K, Patterson N, Price AL and Singh L

doi: 10.1038/nature08365

Table S1. Autosomal F_{ST} among 25 Indian groups (no inbreeding correction)

			Kashmiri Pandit	Vaish	Srivastava	Sahariya	Lodi	Satnami	Bhil	Tharu	Meghawal	Vysya	Naidu	Velama	Madiga	Mala	Kamsali	Chenchu	Kurumba	Hallaki	Santhal	Kharia	Nyshi	Ao Naga	Siddi	Onge	Great Andamanese
Kashmiri Pandit	5	Kashmir		0.0005	0.0058	0.0228	0.0085	0.0134	0.0113	0.0049	0.0039	0.0190	0.0112	0.0116	0.0147	0.0150	0.0141	0.0599	0.0122	0.0136	0.0236	0.0324	0.0824	0.0860	0.0772	0.1335	0.0522
Vaish	4	Uttar Pradesh	0.0007		0.0035	0.0166	0.0050	0.0083	0.0061	0.0009	0.0034	0.0141	0.0091	0.0078	0.0103	0.0108	0.0096	0.0563	0.0064	0.0101	0.0163	0.0251	0.0801	0.0836	0.0746	0.1277	0.0484
Srivastava	2	Uttar Pradesh	0.0012	0.0013		0.0142	0.0054	0.0062	0.0059	0.0029	0.0058	0.0120	0.0076	0.0088	0.0068	0.0103	0.0091	0.0554	0.0069	0.0094	0.0168	0.0233	0.0786	0.0835	0.0774	0.1292	0.0477
Sahariya	4	Uttar Pradesh	0.0010	0.0010	0.0015		0.0130	0.0089	0.0113	0.0095	0.0187	0.0203	0.0146	0.0202	0.0119	0.0113	0.0141	0.0636	0.0113	0.0176	0.0111	0.0138	0.0667	0.0692	0.0838	0.1205	0.0443
Lodi	5	Uttar Pradesh	0.0007	0.0008	0.0012	0.0009		0.0069	0.0058	0.0029	0.0071	0.0116	0.0065	0.0092	0.0065	0.0062	0.0079	0.0563	0.0064	0.0081	0.0146	0.0213	0.0794	0.0824	0.0788	0.1281	0.0488
Satnami	4	Madhya Pradesh	0.0009	0.0010	0.0013	0.0010	0.0009		0.0057	0.0038	0.0099	0.0140	0.0087	0.0118	0.0053	0.0062	0.0069	0.0566	0.0059	0.0104	0.0078	0.0125	0.0665	0.0695	0.0789	0.1204	0.0419
Bhil	7	Gujarat	0.0007	0.0008	0.0011	0.0007	0.0006	0.0007		0.0022	0.0082	0.0129	0.0081	0.0092	0.0053	0.0052	0.0077	0.0560	0.0036	0.0102	0.0094	0.0170	0.0765	0.0796	0.0805	0.1235	0.0447
Tharu	9	Uttarkhand	0.0006	0.0006	0.0011	0.0007	0.0005	0.0007	0.0004		0.0049	0.0108	0.0052	0.0078	0.0047	0.0049	0.0055	0.0524	0.0021	0.0072	0.0080	0.0150	0.0701	0.0740	0.0753	0.1204	0.0409
Meghawal	5	Rajasthan	0.0007	0.0008	0.0012	0.0009	0.0007	0.0008	0.0006	0.0005		0.0158	0.0090	0.0107	0.0103	0.0110	0.0108	0.0592	0.0096	0.0117	0.0192	0.0279	0.0818	0.0858	0.0777	0.1300	0.0509
Vysya	5	Andhra Pradesh	0.0008	0.0009	0.0013	0.0010	0.0008	0.0009	0.0007	0.0006	0.0007		0.0147	0.0155	0.0135	0.0121	0.0119	0.0646	0.0128	0.0164	0.0211	0.0280	0.0858	0.0893	0.0877	0.1354	0.0563
Naidu	4	Andhra Pradesh	0.0009	0.0010	0.0014	0.0010	0.0009	0.0010	0.0008	0.0007	0.0008	0.0009		0.0083	0.0077	0.0073	0.0082	0.0599	0.0082	0.0109	0.0166	0.0230	0.0809	0.0843	0.0812	0.1297	0.0515
Velama	4	Andhra Pradesh	0.0008	0.0009	0.0014	0.0010	0.0009	0.0010	0.0007	0.0007	0.0008	0.0009	0.0009		0.0097	0.0103	0.0105	0.0626	0.0100	0.0114	0.0200	0.0288	0.0834	0.0871	0.0802	0.1318	0.0537
Madiga	4	Andhra Pradesh	0.0009	0.0009	0.0013	0.0009	0.0008	0.0009	0.0007	0.0006	0.0008	0.0009	0.0009	0.0009		0.0038	0.0063	0.0576	0.0046	0.0094	0.0115	0.0185	0.0775	0.0798	0.0803	0.1239	0.0469
Mala	3	Andhra Pradesh	0.0010	0.0011	0.0015	0.0012	0.0010	0.0011	0.0009	0.0008	0.0010	0.0010	0.0011	0.0011	0.0010		0.0060	0.0582	0.0040	0.0089	0.0105	0.0169	0.0780	0.0805	0.0830	0.1250	0.0475
Kamsali	4	Andhra Pradesh	0.0009	0.0009	0.0014	0.0010	0.0008	0.0009	0.0007	0.0006	0.0008	0.0009	0.0009	0.0009	0.0009	0.0011		0.0090	0.0068	0.0101	0.0139	0.0210	0.0790	0.0824	0.0824	0.1269	0.0492
Chenchu	6	Andhra Pradesh	0.0013	0.0014	0.0017	0.0014	0.0013	0.0014	0.0013	0.0012	0.0013	0.0014	0.0014	0.0014	0.0014	0.0015	0.0015		0.0543	0.0612	0.0577	0.0655	0.1205	0.1233	0.1245	0.1711	0.0918
Kurumba	9	Kerala	0.0006	0.0007	0.0011	0.0007	0.0006	0.0007	0.0004	0.0004	0.0006	0.0006	0.0007	0.0007	0.0006	0.0009	0.0007	0.0012		0.0095	0.0070	0.0136	0.0747	0.0767	0.0795	0.1185	0.0409
Hallaki	7	Karnataka	0.0007	0.0007	0.0012	0.0008	0.0007	0.0008	0.0005	0.0005	0.0007	0.0007	0.0008	0.0008	0.0007	0.0009	0.0007	0.0013	0.0005		0.0170	0.0243	0.0816	0.0848	0.0828	0.1295	0.0508
Santhal	7	Jharkhand	0.0008	0.0008	0.0012	0.0007	0.0007	0.0007	0.0006	0.0005	0.0007	0.0008	0.0008	0.0009	0.0007	0.0009	0.0008	0.0013	0.0005	0.0006		0.0045	0.0638	0.0674	0.0865	0.1152	0.0386
Kharia	6	Madhya Pradesh	0.0008	0.0009	0.0013	0.0008	0.0008	0.0008	0.0007	0.0006	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010	0.0008	0.0014	0.0006	0.0007	0.0006		0.0616	0.0633	0.0907	0.1197	0.0411
Nyshi	4	Arunachal Pradesh	0.0013	0.0013	0.0017	0.0013	0.0013	0.0013	0.0011	0.0011	0.0013	0.0013	0.0014	0.0014	0.0013	0.0015	0.0013	0.0018	0.0011	0.0012	0.0011	0.0011		0.0215	0.1315	0.1559	0.0729
Ao Naga	4	Nagaland	0.0014	0.0014	0.0018	0.0013	0.0013	0.0013	0.0012	0.0012	0.0013	0.0014	0.0015	0.0014	0.0014	0.0016	0.0014	0.0018	0.0012	0.0013	0.0012	0.0012	0.0011		0.1338	0.1584	0.0752
Siddi	4	Karnataka	0.0017	0.0016	0.0020	0.0018	0.0017	0.0017	0.0017	0.0016	0.0016	0.0018	0.0017	0.0017	0.0018	0.0018	0.0018	0.0019	0.0016	0.0016	0.0016	0.0016	0.0018	0.0018		0.1748	0.1079
Onge	9	Andaman & Nicobar	0.0013	0.0014	0.0017	0.0014	0.0013	0.0013	0.0012	0.0012	0.0013	0.0013	0.0014	0.0015	0.0013	0.0016	0.0014	0.0018	0.0011	0.0012	0.0012	0.0013	0.0016	0.0016	0.0017		0.0905
Great Andamanese	7	Andaman & Nicobar	0.0010	0.0010	0.0014	0.0010	0.0009	0.0010	0.0008	0.0008	0.0010	0.0010	0.0010	0.0010	0.0010	0.0012	0.0010	0.0014	0.0008	0.0008	0.0008	0.0008	0.0012	0.0013	0.0017	0.0012	

Note: F_{ST} values are presented in the top right of the matrix, and standard errors are presented in the bottom left.

Table S2. Autosomal F_{ST} among 25 Indian groups (inbreeding correction)

			Kashmiri Pandit	Vaish	Srivastava	Sahariya	Lodi	Satnami	Bhil	Tharu	Meghawal	Vysya	Naidu	Velama	Madiga	Mala	Kamsali	Chenchu	Kurumba	Hallaki	Santhal	Kharia	Nyshi	Ao Naga	Siddi	Onge	Great Andamanese
Kashmiri Pandit	5	Kashmir		0.0023	0.0059	0.0234	0.0091	0.0142	0.0125	0.0059	0.0052	0.0194	0.0089	0.0111	0.0150	0.0158	0.0122	0.0618	0.0124	0.0119	0.0252	0.0337	0.0826	0.0860	0.0783	0.1349	0.0550
Vaish	4	Uttar Pradesh	0.0008		0.0036	0.0173	0.0057	0.0092	0.0074	0.0020	0.0048	0.0147	0.0070	0.0074	0.0107	0.0117	0.0078	0.0583	0.0067	0.0084	0.0180	0.0265	0.0803	0.0837	0.0758	0.1293	0.0513
Srivastava	2	Uttar Pradesh	0.0014	0.0015		0.0133	0.0045	0.0054	0.0055	0.0023	0.0056	0.0108	0.0037	0.0068	0.0054	0.0096	0.0055	0.0558	0.0055	0.0060	0.0168	0.0230	0.0772	0.0819	0.0771	0.1291	0.0490
Sahariya	4	Uttar Pradesh	0.0010	0.0010	0.0016		0.0125	0.0087	0.0114	0.0094	0.0189	0.0197	0.0112	0.0187	0.0111	0.0110	0.0110	0.0644	0.0104	0.0147	0.0117	0.0140	0.0658	0.0682	0.0839	0.1209	0.0461
Lodi	5	Uttar Pradesh	0.0007	0.0008	0.0015	0.0009		0.0066	0.0059	0.0028	0.0073	0.0110	0.0030	0.0076	0.0057	0.0058	0.0048	0.0571	0.0055	0.0051	0.0151	0.0215	0.0784	0.0814	0.0788	0.1284	0.0506
Satnami	4	Madhya Pradesh	0.0009	0.0010	0.0015	0.0010	0.0009		0.0061	0.0039	0.0103	0.0135	0.0055	0.0104	0.0046	0.0061	0.0040	0.0577	0.0053	0.0077	0.0085	0.0129	0.0657	0.0686	0.0792	0.1210	0.0438
Bhil	7	Gujarat	0.0007	0.0008	0.0013	0.0008	0.0006	0.0007		0.0027	0.0091	0.0129	0.0054	0.0083	0.0051	0.0055	0.0053	0.0574	0.0033	0.0079	0.0105	0.0178	0.0762	0.0791	0.0811	0.1245	0.0471
Tharu	9	Uttarkhand	0.0006	0.0006	0.0013	0.0008	0.0005	0.0007	0.0004		0.0056	0.0106	0.0022	0.0066	0.0043	0.0050	0.0028	0.0536	0.0017	0.0047	0.0089	0.0156	0.0695	0.0733	0.0757	0.1212	0.0430
Meghawal	5	Rajasthan	0.0007	0.0008	0.0014	0.0010	0.0007	0.0008	0.0006	0.0006		0.0159	0.0064	0.0099	0.0102	0.0114	0.0084	0.0608	0.0094	0.0096	0.0204	0.0288	0.0816	0.0854	0.0784	0.1311	0.0534
Vysya	5	Andhra Pradesh	0.0008	0.0009	0.0016	0.0010	0.0008	0.0009	0.0007	0.0006	0.0008		0.0111	0.0138	0.0125	0.0116	0.0087	0.0653	0.0118	0.0133	0.0215	0.0281	0.0847	0.0881	0.0877	0.1357	0.0579
Naidu	4	Andhra Pradesh	0.0009	0.0010	0.0016	0.0010	0.0009	0.0010	0.0008	0.0007	0.0008	0.0009		0.0038	0.0040	0.0041	0.0022	0.0579	0.0044	0.0051	0.0142	0.0204	0.0772	0.0804	0.0787	0.1272	0.0504
Velama	4	Andhra Pradesh	0.0009	0.0009	0.0015	0.0010	0.0009	0.0010	0.0008	0.0007	0.0008	0.0009	0.0009		0.0078	0.0090	0.0063	0.0624	0.0081	0.0074	0.0194	0.0280	0.0814	0.0850	0.0793	0.1312	0.0544
Madiga	4	Andhra Pradesh	0.0009	0.0009	0.0014	0.0010	0.0008	0.0009	0.0007	0.0006	0.0008	0.0009	0.0009	0.0009		0.0031	0.0028	0.0581	0.0034	0.0062	0.0117	0.0184	0.0762	0.0784	0.0801	0.1239	0.0483
Mala	3	Andhra Pradesh	0.0011	0.0011	0.0017	0.0012	0.0010	0.0011	0.0009	0.0009	0.0010	0.0010	0.0012	0.0012	0.0011		0.0030	0.0592	0.0033	0.0061	0.0112	0.0173	0.0772	0.0795	0.0833	0.1256	0.0494
Kamsali	4	Andhra Pradesh	0.0009	0.0009	0.0016	0.0010	0.0008	0.0009	0.0007	0.0006	0.0008	0.0009	0.0009	0.0009	0.0008	0.0011		0.0090	0.0033	0.0045	0.0118	0.0186	0.0755	0.0788	0.0801	0.1247	0.0484
Chenchu	6	Andhra Pradesh	0.0013	0.0014	0.0019	0.0014	0.0013	0.0014	0.0013	0.0012	0.0013	0.0014	0.0014	0.0014	0.0014	0.0015	0.0015		0.0547	0.0596	0.0595	0.0670	0.1209	0.1235	0.1257	0.1727	0.0948
Kurumba	9	Kerala	0.0006	0.0007	0.0013	0.0007	0.0006	0.0007	0.0004	0.0004	0.0006	0.0006	0.0007	0.0007	0.0006	0.0009	0.0006	0.0012		0.0062	0.0071	0.0134	0.0734	0.0753	0.0792	0.1185	0.0423
Hallaki	7	Karnataka	0.0007	0.0007	0.0014	0.0008	0.0007	0.0008	0.0005	0.0005	0.0007	0.0007	0.0008	0.0008	0.0007	0.0010	0.0007	0.0013	0.0005		0.0151	0.0220	0.0783	0.0814	0.0807	0.1275	0.0501
Santhal	7	Jharkhand	0.0008	0.0008	0.0014	0.0008	0.0007	0.0007	0.0006	0.0005	0.0007	0.0008	0.0008	0.0009	0.0008	0.0010	0.0008	0.0013	0.0005	0.0006		0.0057	0.0638	0.0673	0.0874	0.1166	0.0414
Kharia	6	Madhya Pradesh	0.0009	0.0009	0.0015	0.0009	0.0008	0.0008	0.0007	0.0006	0.0008	0.0009	0.0009	0.0009	0.0009	0.0011	0.0009	0.0014	0.0006	0.0007	0.0006		0.0613	0.0629	0.0914	0.1208	0.0436
Nyshi	4	Arunachal Pradesh	0.0013	0.0013	0.0019	0.0013	0.0013	0.0013	0.0012	0.0011	0.0013	0.0013	0.0014	0.0014	0.0013	0.0015	0.0013	0.0018	0.0011	0.0012	0.0011	0.0011		0.0198	0.1311	0.1557	0.0742
Ao Naga	4	Nagaland	0.0014	0.0014	0.0020	0.0014	0.0013	0.0013	0.0013	0.0012	0.0013	0.0014	0.0015	0.0015	0.0014	0.0016	0.0014	0.0018	0.0012	0.0013	0.0012	0.0013	0.0012		0.1334	0.1581	0.0764
Siddi	4	Karnataka	0.0017	0.0017	0.0021	0.0019	0.0017	0.0017	0.0017	0.0016	0.0017	0.0018	0.0018	0.0017	0.0018	0.0018	0.0018	0.0020	0.0016	0.0017	0.0016	0.0017	0.0019	0.0019		0.1756	0.1099
Onge	9	Andaman & Nicobar	0.0013	0.0014	0.0018	0.0014	0.0013	0.0013	0.0012	0.0012	0.0013	0.0013	0.0014	0.0015	0.0013	0.0016	0.0014	0.0018	0.0011	0.0012	0.0012	0.0013	0.0016	0.0017	0.0018		0.0934
Great Andamanese	7	Andaman & Nicobar	0.0010	0.0010	0.0016	0.0010	0.0009	0.0010	0.0008	0.0008	0.0010	0.0010	0.0010	0.0010	0.0010	0.0012	0.0010	0.0015	0.0008	0.0008	0.0008	0.0008	0.0012	0.0013	0.0017	0.0012	

Note: F_{ST} values are presented in the top right of the matrix, and standard errors are presented in the bottom left.

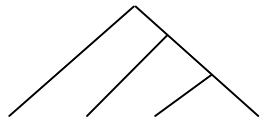
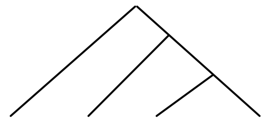
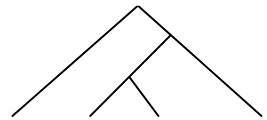
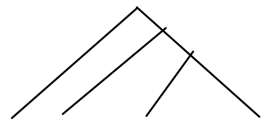
Table S3. Pairwise F_{ST} for combinations of Indian groups

Category of comparison	Details of comparison	No. of groups	Average F_{ST}	Average F_{ST} correcting for inbreeding
All India[†]	All pairs	19	0.0109	0.0100
	Comparing matched groups (both Uttar Pradesh or both Andhra Pradesh and both traditionally upper caste or both traditionally lower or middle caste)	9 pairs	0.0087	0.0069
Restricting to language	Indo-European speaking pairs	9	0.0076	0.0080
	Dravidian speaking pairs	8	0.0096	0.0067
Restricting to caste level	Traditionally upper caste pairs	5	0.0074	0.0061
	Traditionally lower and middle caste pairs	6	0.0010	0.0093
Restricting to a state	Uttar Pradesh pairs	4	0.0096	0.0095
	Andhra Pradesh pairs	6	0.0097	0.0069

* We exclude 6 outlier groups: the Onge, Great Andamanese, Ao Naga, Nyshi, Siddi and Chenchu. Individual pairwise F_{ST} values for all possible pairs of 25 groups are presented in Tables S1 and S2.

[†] The inbreeding corrected average F_{ST} between all pairs of 19 Indian groups (0.0100) is higher than the average F_{ST} between all pairs of 23 European groups in ref. 1 (0.0033). This phenomenon persists when we restrict to pairs of Indian groups of the same traditional caste level that are matched by geographic region (0.0069), and compare this to pairs of European groups that are matched by geographic region (0.0018). For performing a regional analysis of the European data in ref. 1, we defined five European “regions”: Scandinavia (Helsinki, Førde, and Uppsala), Northern Europe (Kopenhagen, Rotterdam, Dublin, London and Kiel), Central Europe (Budapest, Lausanne, Augsburg, Innsbruck and Lyon), Eastern Europe (Prague, Belgrade, Bucharest and Warsaw), and Southern Europe (Rome, Lisbon, Madrid, Greece, Ancona and Barcelona).

Table S4. Formal tests for mixture on the Indian Cline (expansion of Table 2 in the main text)

Group (ordered from most ASI-related to most ANI)	**No. samples after pruning**	**Z-score for *3 Population Test*** $(P_X-P_{CEU})(P_X-P_{Santhal})$ (negative values indicate violation)	**Z-score for *4 Pop Test*** $(P_{YRI}-P_{CEU})(P_{Onge}-P_X)$  YRI CEU Onge X	**Z-score for *4 Pop Test*** $(P_{YRI}-P_{Onge})(P_{CEU}-P_X)$  YRI Onge CEU X	**Z-score for *4 Pop Test*** $(P_{YRI}-P_X)(P_{CEU}-P_{Onge})$  YRI Onge CEU X	**Z-score for *4 Pop Test*** $(P_{YRI}-P_{Papuan})(P_{Dai}-P_X)$††  YRI Papuan Dai X
Onge	9	77.3 (not significant)	n/a	n/a	n/a	1.7 (not significant)
Mala	3	-2.5	13.8	20.4	7.1	-9.7
Madiga	4	-2.7	15.1	21.6	6.8	-11.2
Chenchu	6	31.3 (not significant)	16.9	21.2	5.6	-9.7
Kurumba	6	-12.6	17.1	24.5	6.0	-11.8
Bhil	7	-10.6	18.1	23.9	5.0	-13.0
Kamsali	3	-6.5	17.1	20.7	3.5	-10.9
Satnami	3	-5.6	16.7	19.0	3.4	-10.5
Vysya	5	5.4 (not significant)	18.1	21.1	1.8 (not significant†)	-11.5
Naidu	4	-3.3	18.4	20.2	-0.3 (not significant†)	-12.8
Lodi	5	-8.9	21.9	20.8	-1.1 (not significant†)	-12.9
Tharu	5	-20.6	20.6	21.5	-1.4 (not significant†)	-14.3
Velama	4	-3.2	19.4	17.2	-2.7	-14.4
Srivastava	2	-7.5	19.8	14.1	-5.5	-11.9
Meghawal	5	-13.3	24.8	18.0	-8.1	-15.6
Vaish	4	-22.0	25.7	18.0	-10.1	-15.6
Kashmiri Pandit	5	-20.6	30.7	17.0	-15.7	-17.1
Sindhi *	10	-26.3	27.8	13.0	-18.3	-20.7
Pathan *	15	-34.3	30.8	14.3	-21.2	-20.0

*Tests using HGDP samples use the reduced set of 119,744 autosomal SNPs, while all other tests use 560,123 autosomal SNPs.

† Four groups in the middle of the Indian Cline (from the Vysya to the Tharu) give non-significant Z-scores for the *4 Population Test* for the third tree topology ((YRI,X),(CEU,Onge)), which we hypothesize reflects the fact that two other topologies are both present (due to ancient mixture) and balance in their contribution to the *4 Population Test* statistic. However, we can show by another argument that this topology is not consistent with the data in the absence of mixture. Fitting this topology to the data and using a Weighted Block Jackknife to obtain a standard error, we estimate that the internal branches have negative length with high statistical significance (normally distributed Z-scores of -34 (Vysya), -34 (Naidu), -39 (Lodi) and -38 (Tharu) (Note S3)). Since the internal branch length is proportional to genetic drift under the null hypothesis of a correct topology, the topology cannot be correct.

†† The Onge are the only ASI-related group with no evidence at all of ANI-related mixture, as assessed by a *4 Population Test* of the topology ((YRI,Papuan),(Dai,X)) in the last column. The f_4 statistic is extremely significantly different from 0 (Z-score << -9 standard deviations) for all Indian Cline groups, but is consistent with 0 (Z = 1.7) for the Onge. Thus, all the Indian Cline group have a component of mixture that the Onge do not.

Table S5. ANI ancestry estimates based on three alternative methods

Group	f_3 *Ancestry Estimation*					f_4 *Ancestry Estimation* *					*Regression Ancestry Estimation* †
	Auto-somes	**Stand. error**	**X chrom.**	**Stand. error**	**P-value for X-autosome difference**	**Auto-somes**	**Stand. error**	**X chrom.**	**Stand. error**	**P-value for X-autosome difference**	
Mala	38.8%	1.2%	38%	9%	0.46	38.2%	1.7%	40%	13%	0.54	41%
Madiga	40.6%	1.2%	35%	14%	0.33	40.6%	1.7%	49%	13%	0.73	41%
Chenchu	40.7%	1.3%	31%	11%	0.18	42.1%	1.7%	23%	9%	0.021	42%
Bhil	42.9%	1.1%	42%	10%	0.45	42.5%	1.4%	37%	10%	0.30	44%
Satnami	43.0%	1.3%	33%	15%	0.26	43.6%	1.8%	39%	11%	0.35	46%
Kurumba	43.2%	1.1%	28%	10%	0.06	42.3%	1.5%	36%	10%	0.25	43%
Kamsali	44.5%	1.3%	44%	10%	0.50	43.8%	1.7%	49%	18%	0.62	45%
Vysya	46.2%	1.2%	40%	11%	0.29	44.7%	1.7%	44%	10%	0.48	49%
Lodi	49.9%	1.1%	43%	10%	0.25	47.7%	1.6%	47%	8%	0.48	52%
Naidu	50.1%	1.2%	54%	12%	0.62	48.6%	1.6%	54%	11%	0.69	52%
Tharu	51.0%	1.2%	34%	9%	0.03	50.9%	1.5%	35%	9%	0.04	53%
Velama	54.7%	1.3%	53%	11%	0.43	52.4%	1.7%	44%	13%	0.26	57%
Srivastava	56.4%	1.5%	43%	11%	0.11	55.0%	1.9%	47%	15%	0.30	60%
Meghawal	60.3%	1.2%	67%	13%	0.69	57.1%	1.4%	58%	11%	0.53	61%
Vaish	62.6%	1.2%	55%	13%	0.26	60.3%	1.5%	51%	12%	0.23	64%
Kashmiri Pandit	70.6%	1.2%	64%	11%	0.28	69.3%	1.3%	52%	7%	0.004	72%
Sindhi	73.7%	1.1%	81%	12%	0.71	70.7%	1.0%	65%	6%	0.17	78%
Pathan	76.9%	1.1%	83%	11%	0.71	74.2%	0.9%	73%	6%	0.40	81%

* For f_4 *Ancestry Estimation*, we use the statistic f_4(Adygei,Papuan; India,Onge)/f_4(Adygei,Papuan; CEU,Onge) to estimate ANI ancestry proportion, and obtain a standard error for each group by a Block Jackknife. This calculation only analyzes one Indian Cline group at a time, and hence the estimates are not expected to be biased by the outlier-removal procedure we used to eliminate specific groups from the Indian Cline (i.e. Kharia, Santhal, Sahariya and Hallaki).

† For *Regression Ancestry Estimation*, we plot f_4(YRI,Adygei; Onge,$India_k$), a number proportional to ANI ancestry, against f_4(YRI,Onge; Adygei,$India_k$), a number proportional to ASI ancestry. We then use regression analysis over all 18 groups to extrapolate the x-intercept and y-intercept, and interpolate the ANI ancestry proportion for each group (Note S5).

doi: 10.1038/nature08365

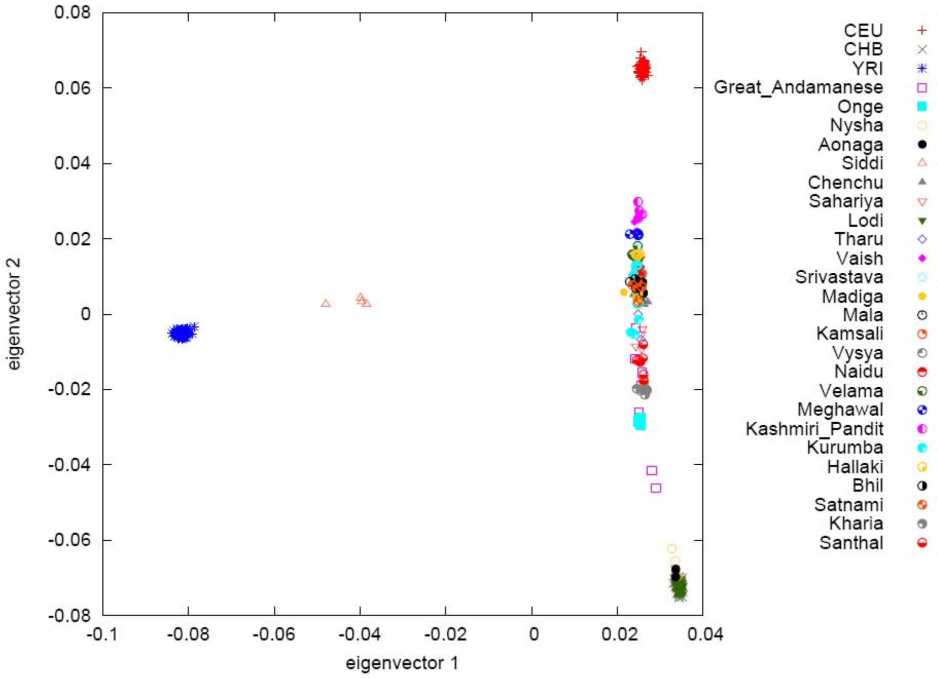
Table S6. mtDNA and Y chromosome data

		Mala	Madiga	Chenchu	Kurumba	Bhil	Kamsali	Satnami	Vysya	Naidu	Lodhi	Tharu	Velama	Srivastava	Meghawal	Vaish	Kash. Pandit	Ao Naga	Kharia	Santhal	Sahariya	Nyshi	Siddi	Hallaki	Onge	Gr. Andaman
mtDNA																										
Samples		37	20	75	68	147	43	28	58	60	49	104	16	35	30	9	61	60	101	110	23	45	94	29	33	9
ASI %		38%	90%	48%	56%	52%	67%	39%	31%	20%	51%	8%	25%	46%	50%	89%	31%	12%	47%	50%	57%	18%	4%	45%	0%	0%
ANI %		0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	13%	0%	0%	0%	73%	89%
M18	ASI	1	4	0	0	0	0	1	0	4	0	0	0	2	3	0	0	0	0	6	0	0	0	0	0	0
M2	ASI	0	3	30	15	10	1	3	1	3	0	0	0	0	0	0	0	1	17	11	2	3	0	0	0	0
M25	ASI	5	0	0	0	10	1	0	2	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
M2a	ASI	0	1	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	2	0	0	0	9	0	0
M2b	ASI	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	3	0	4	0	1	0	0
M3	ASI	0	1	0	4	9	3	0	5	3	9	0	2	2	3	7	3	1	9	1	3	1	0	1	0	0
M3a	ASI	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M4	ASI	2	1	5	5	30	0	1	3	1	4	2	0	0	1	1	5	0	6	5	2	0	0	0	0	0
M4a	ASI	0	0	0	0	0	3	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
M5	ASI	2	2	1	10	5	9	2	4	0	5	1	0	0	1	0	6	5	13	19	0	0	0	0	0	0
M5?	ASI	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M5a	ASI	0	2	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	2	0	0	0	0	0
M6	ASI	0	2	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
R5	ASI	4	0	0	4	10	3	0	3	0	0	2	0	0	0	0	2	0	0	0	1	0	0	2	0	0
R6	ASI	0	0	0	0	3	0	4	0	0	0	0	1	0	2	0	3	0	0	5	1	0	0	0	0	0
R7	ASI	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
U2	ASI	0	1	0	0	0	0	0	0	0	7	0	0	4	0	0	0	0	0	0	0	0	4	0	0	0
U2c	ASI	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M40	ASI	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
M31	ASI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
M31a	ASI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	8
M32	ASI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	1
M35	ASI	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
I	ASI	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
M*	ASI	11	0	37	23	34	12	6	22	18	21	23	5	13	8	0	20	31	30	29	6	27	52	11	0	0
U3	ANI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
U8	ANI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
U7	ANI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0
M30	ANI	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
M39	ANI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
B4	ANI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
B5a	ANI	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	ANI	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F1	ANI	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	5	0	0	0	6	0	0	0	0
F1a	ANI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
F1c	ANI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	2	0	0	0	0
R	ANI	7	0	1	5	25	2	5	10	24	3	66	2	5	3	0	13	11	17	20	0	1	10	3	0	0
R1	ANI	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
T	ANI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
U	ANI	2	0	1	2	11	0	6	5	5	0	1	0	0	4	0	4	0	6	1	0	0	0	1	0	0
U10	ANI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
U11	ANI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
U9	ANI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Ua	ANI	2	0	0	0	0	0	0	1	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ub	ANI	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0
W	ANI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L0	unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
L2	unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0
L3	unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
L3?	unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
Y chromosome																										
Samples		17	21	41	21	27	42	30	9	17	74	68	15	38	10	19	19	57	30	249	23	45	65	27	10	10
ASI %		47%	67%	59%	38%	89%	43%	70%	67%	65%	26%	65%	80%	21%	10%	74%	32%	61%	80%	71%	65%	82%	26%	89%	0%	40%
ANI %		53%	33%	41%	62%	11%	57%	30%	33%	35%	74%	35%	20%	79%	90%	26%	68%	39%	20%	29%	35%	18%	11%	11%	0%	60%
H	ASI	1	6	15	7	10	15	9	0	2	3	22	1	8	0	2	0	12	2	56	0	2	0	13	0	0
H1	ASI	1	0	0	0	0	0	0	0	3	0	0	0	0	0	3	1	0	9	0	1	0	9	0	0	0
H2	ASI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
L	ASI	4	1	6	0	0	3	0	0	0	4	0	9	0	0	0	1	5	0	0	0	0	0	0	0	0
O	ASI	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	10	0	0	1	0	1
O2	ASI	0	0	0	0	14	0	2	0	0	12	14	0	0	0	0	0	0	0	117	0	0	0	9	0	1
O3	ASI	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	35	0	0	0	2
R2	ASI	0	4	3	1	0	0	10	6	2	0	0	2	0	0	6	4	18	0	3	4	0	0	1	0	0
F	ASI	2	2	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	5	0	0	0
R	ASI	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	ANI	1	1	3	0	0	5	0	0	0	40	0	1	8	6	3	0	0	0	0	1	0	2	0	0	0
R1	ANI	8	6	12	3	3	19	9	2	5	7	13	2	22	1	2	11	21	1	18	7	8	2	1	0	0
C	ANI	0	0	2	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0
K	ANI	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K*	ANI	0	0	0	4	0	0	0	1	0	6	11	0	0	0	0	0	1	3	47	0	0	3	1	0	2
P	ANI	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	1	0	2	0	0	0	0	1	0	4
G	ANI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
B	unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
E2	unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
E3a	unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0
B2	unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0
D*	unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0

Note: Haplogroups were designated as typical of Ancient South Indians (ASI) or Ancient North Indians (ANI) based on the judgement of an expert on mtDNA and Y chromosome variation (KT) who was blinded to ancestry estimates from the autosomes.

Figure S1

(a)



(b)

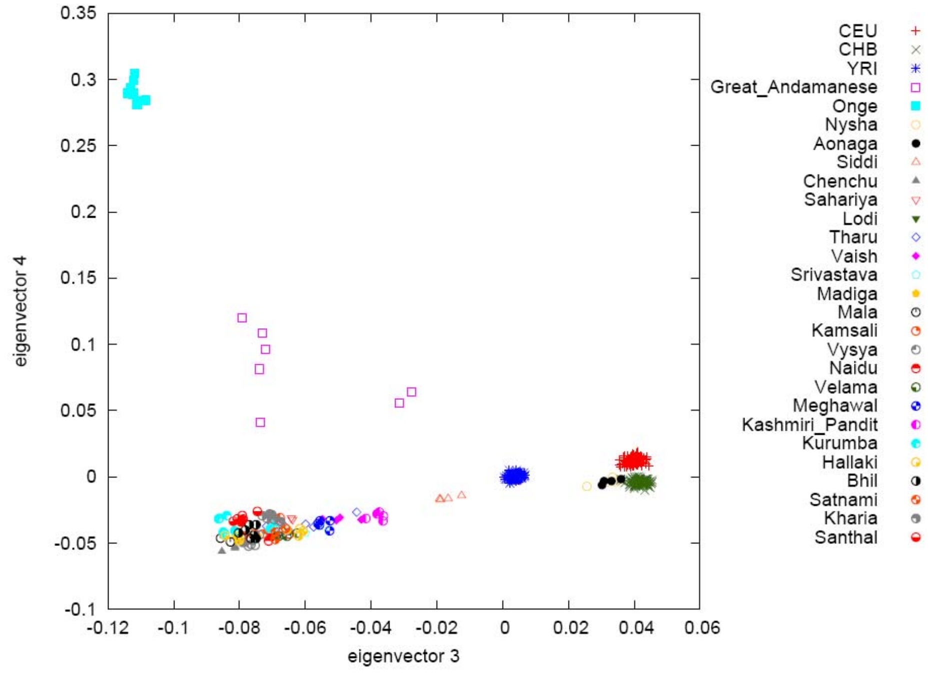


Figure S1 Legend: Principal components analysis of the 25 groups, together with CEU, CHB and YRI from HapMap. **(a)** The top two PCs show that the Siddi are an outlying group with ancestry that is related to West Africans (YRI), consistent with the known origin of this group in the Arab slave trade. They also show that the Nyshi and Ao Naga are closely related to East Asians (CHB), as expected from the fact that these groups speak a Tibeto-Burman language. **(b)** The third and fourth PCs distinguish the Andaman Island groups, and show that the Great Andamanese do not cluster in the plot. This is a signature of recent gene flow from the mainland in the last handful of generations (Note S1).

Figure S2

[see next page for the figure]

Figure S2 Legend: Decay of allele sharing provides evidence for ancient founder effects, which in many Indian Cline groups appear to have occurred at least 30 generations ago. For each of the groups that we genotyped (except for the Srivastava with just two individuals), we examined all pairs of samples, and recorded whether 0, 1 or 2 alleles were shared at each SNP (we scored SNPs that were heterozygous in both individuals as sharing 1 allele to account for phase ambiguity). Founder events are expected to cause segments of the genome to be identical by state (IBS) for at least one allele over a stretch of sequence due to their descent from a shared founder, with the extent of the shared segment providing information about the age of the event. To correct for allele sharing inherited from the ancestral population, we subtracted the curve obtained by comparisons across different Indian Cline groups, picking the closest match among the groups with 65% ± 5% ANI ancestry (Meghawal, Vaish and Kashmiri Pandit), 58% ± 5% ANI ancestry (Velama, Srivastava, Meghawal and Vaish), 53% ± 5% ANI ancestry (Lodi, Naidu, Tharu, Velama and Srivastava), 47% ± 5% ANI ancestry (Bhil, Satnami, Kurumba, Kamsali, Vysya, Lodi, Naidu and Tharu), and 42% ± 5% ANI ancestry (Mala, Madiga, Chenchu, Bhil, Satnami, Kurumba, Kamsali and Vysya). We performed a least-squares fit of $y = a + be^{-2Dt}$ to the data from each group where a, b and *t* are constants, D is the distance in Morgans between SNPs, and the factor of 2 corresponds to the fact that a recombination can occur on either haplotype that is being compared. Computer simulations reported in Figure S3 show that this procedure can infer the age *t* of founder events with reasonable accuracy under the assumption of a single founder event. As an example, in the Vysya, allele sharing decreases with an exponential decay of 0.461 cM, suggesting a founder event roughly 100/(2*0.461) = 108 generations ago (see also Figure 2). There are 6 Indo-European and Dravidian speaking groups with estimated founder events of >30 generations ago: Bhil (40), Hallaki (32), Meghawal (59), Sahariya (108), Vysya (108) and Velama (88).