

The American Journal of Human Genetics, Volume 89

Supplemental Data

Indian Siddis: African Descendants with Indian admixture

Anish M. Shah, Rakesh Tamang, Priya Moorjani, Deepa Selvi Rani, Periyasamy Govindaraj, Gururaj Kulkarni, Tanmoy Bhattacharya, Mohammed S. Mustak, LVKS Bhaskar, Alla G Reddy, Dharmendra Gadhvi, Pramod B. Gai, Gyaneshwer Chaubey, Nick Patterson, David Reich, Chris Tyler-Smith, Lalji Singh and Kumarasamy Thangaraj

Table S1. Fst distribution in the present study

	Portugal	YRI	LWK	Siddi Karnataka1	Siddi Karnataka2	Siddi Gujarat	Mala	Madiga	Kurumba	Bhil	Kamsali	Satnami	Vysya	Naidu	Lodi	Tharu	Velama	Srivastava	Meghawal	Vaish	Kashmiri Pandit	Hallaki
Portugal	0.00	0.149	0.137	0.09	0.101	0.091	0.05	0.05	0.051	0.045	0.05	0.051	0.056	0.045	0.043	0.039	0.042	0.035	0.034	0.027	0.027	0.048
YRI	0.149	0.00	0.008	0.042	0.027	0.031	0.147	0.147	0.145	0.145	0.149	0.146	0.155	0.147	0.146	0.141	0.146	0.143	0.143	0.141	0.145	0.148
LWK	0.137	0.008	0.00	0.035	0.023	0.026	0.136	0.135	0.134	0.134	0.137	0.135	0.144	0.135	0.134	0.13	0.134	0.132	0.132	0.129	0.133	0.137
Siddi Karnataka1	0.09	0.042	0.035	0.00	0.026	0.002	0.081	0.081	0.079	0.079	0.082	0.08	0.089	0.082	0.078	0.075	0.08	0.078	0.077	0.074	0.079	0.083
Siddi Karnataka2	0.101	0.027	0.023	0.026	0.00	0.021	0.09	0.09	0.089	0.088	0.092	0.09	0.098	0.09	0.089	0.084	0.089	0.086	0.087	0.085	0.09	0.091
Siddi Gujarat	0.091	0.031	0.026	0.002	0.021	0.00	0.083	0.083	0.081	0.081	0.084	0.083	0.091	0.083	0.081	0.077	0.083	0.08	0.079	0.077	0.081	0.086
Mala	0.05	0.147	0.136	0.081	0.09	0.083	0.00	0.002	0.003	0.004	0.005	0.005	0.014	0.006	0.005	0.003	0.008	0.006	0.009	0.008	0.015	0.008
Madiga	0.05	0.147	0.135	0.081	0.09	0.083	0.002	0.00	0.003	0.004	0.006	0.006	0.014	0.006	0.005	0.003	0.009	0.006	0.009	0.008	0.015	0.008
Kurumba	0.051	0.145	0.134	0.079	0.089	0.081	0.003	0.003	0.00	0.004	0.006	0.006	0.014	0.007	0.006	0.002	0.009	0.006	0.009	0.006	0.015	0.009
Bhil	0.045	0.145	0.134	0.079	0.088	0.081	0.004	0.004	0.004	0.00	0.008	0.007	0.014	0.007	0.005	0.003	0.009	0.005	0.007	0.006	0.012	0.01
Kamsali	0.05	0.149	0.137	0.082	0.092	0.084	0.005	0.006	0.006	0.008	0.00	0.007	0.014	0.008	0.009	0.005	0.01	0.009	0.011	0.009	0.016	0.01
Satnami	0.051	0.146	0.135	0.08	0.09	0.083	0.005	0.006	0.006	0.007	0.007	0.00	0.016	0.009	0.007	0.004	0.012	0.007	0.01	0.009	0.016	0.011
Vysya	0.056	0.155	0.144	0.089	0.098	0.091	0.014	0.014	0.014	0.014	0.014	0.016	0.00	0.016	0.013	0.013	0.017	0.013	0.017	0.015	0.021	0.018
Naidu	0.045	0.147	0.135	0.082	0.09	0.083	0.006	0.006	0.007	0.007	0.008	0.009	0.016	0.00	0.006	0.005	0.008	0.007	0.008	0.009	0.013	0.011
Lodi	0.043	0.146	0.134	0.078	0.089	0.081	0.005	0.005	0.006	0.005	0.009	0.007	0.013	0.006	0.00	0.003	0.009	0.005	0.007	0.005	0.01	0.008
Tharu	0.039	0.141	0.13	0.075	0.084	0.077	0.003	0.003	0.002	0.003	0.005	0.004	0.013	0.005	0.003	0.00	0.007	0.002	0.005	0.001	0.008	0.007
Velama	0.042	0.146	0.134	0.08	0.089	0.083	0.008	0.009	0.009	0.009	0.01	0.012	0.017	0.008	0.009	0.007	0.00	0.009	0.011	0.008	0.013	0.012
Srivastava	0.035	0.143	0.132	0.078	0.086	0.08	0.006	0.006	0.006	0.005	0.009	0.007	0.013	0.007	0.005	0.002	0.009	0.00	0.006	0.002	0.007	0.009
Meghawal	0.034	0.143	0.132	0.077	0.087	0.079	0.009	0.009	0.009	0.007	0.011	0.01	0.017	0.008	0.007	0.005	0.011	0.006	0.00	0.003	0.007	0.012
Vaish	0.027	0.141	0.129	0.074	0.085	0.077	0.008	0.008	0.006	0.006	0.009	0.009	0.015	0.009	0.005	0.001	0.008	0.002	0.003	0.00	0.004	0.01
Kashmiri Pandit	0.027	0.145	0.133	0.079	0.09	0.081	0.015	0.015	0.015	0.012	0.016	0.016	0.021	0.013	0.01	0.008	0.013	0.007	0.007	0.004	0.00	0.016
Hallaki	0.048	0.148	0.137	0.083	0.091	0.086	0.008	0.008	0.009	0.01	0.01	0.011	0.018	0.011	0.008	0.007	0.012	0.009	0.012	0.01	0.016	0.00

NOTE: Fst values were computed using EIGENSOFT.

Table S2a. Formal test to confirm if Siddis have ancestry from Africans, Europeans and Indians

a. Is Ancestral pop2 necessary?				b. Is Ancestral pop2 sufficient?			
Admixed pop	Ancestral pop1	Ancestral pop2	Z-score	Admixed	Ancestral pop1	Ancestral pop2	Z-score
Siddi_Gujarat	Vaish	YRI	25.87	Siddi_Gujarat	YRI	Vaish	19.53
Siddi_Gujarat	Tharu	YRI	30.55	Siddi_Gujarat	YRI	Tharu	19.66
Siddi_Gujarat	Lodi	YRI	29.56	Siddi_Gujarat	YRI	Lodi	18.80
Siddi_Gujarat	Kashmiri Pandit	YRI	36.00	Siddi_Gujarat	YRI	Kashmiri Pandit	19.34
Siddi_Gujarat	Bhil	YRI	33.10	Siddi_Gujarat	YRI	Bhil	19.83
Siddi_Gujarat	Meghawal	YRI	30.52	Siddi_Gujarat	YRI	Meghawal	19.10
Siddi_Gujarat	Kurumba	YRI	32.34	Siddi_Gujarat	YRI	Kurumba	19.99
Siddi_Gujarat	Madiga	YRI	32.88	Siddi_Gujarat	YRI	Madiga	19.58
Siddi_Gujarat	Mala	YRI	31.91	Siddi_Gujarat	YRI	Mala	19.59
Siddi_Gujarat	Kamsali	YRI	29.58	Siddi_Gujarat	YRI	Kamsali	19.84
Siddi_Gujarat	Velama	YRI	28.81	Siddi_Gujarat	YRI	Velama	19.65
Siddi_Gujarat	Naidu	YRI	29.43	Siddi_Gujarat	YRI	Naidu	19.48
Siddi_Gujarat	Satnami	YRI	29.05	Siddi_Gujarat	YRI	Satnami	19.40
Siddi_Gujarat	Hallaki	YRI	30.80	Siddi_Gujarat	YRI	Hallaki	18.75
Siddi_Gujarat	Vysya	YRI	33.77	Siddi_Gujarat	YRI	Vysya	19.02
Siddi_Gujarat	Portugal	YRI	39.26	Siddi_Gujarat	YRI	Portugal	19.22
Siddi_Gujarat	ICP	YRI	35.25	Siddi_Gujarat	YRI	ICP	18.87

NOTE: The Z-scores evaluates the significance of incorporating Ancestral pop2 in model - that is compares the model with ancestral pop2 against a null model with Ancestral pop2 excluded. Z-scores > |3| are statistically significant and implies that ancestral pop1 alone does not provide a good fit to the data. ICP - combined data from 16 Indian groups and Portuguese- represents the ancestral non-African population.

Table S2B. Identify the models that provide a good fit to the Siddi data

Admixed pop	pop1	pop2	pop3	Z-score
Siddi_Gujarat	YRI	Vaish	Portugal	-1.45
Siddi_Gujarat	YRI	Tharu	Portugal	0.97
Siddi_Gujarat	YRI	Kurumba	Portugal	4.22
Siddi_Gujarat	YRI	Lodi	Portugal	3.13
Siddi_Gujarat	YRI	Bhil	Portugal	3.59
Siddi_Gujarat	YRI	Madiga	Portugal	3.98
Siddi_Gujarat	YRI	Mala	Portugal	5.20
Siddi_Gujarat	YRI	Kamsali	Portugal	3.43
Siddi_Gujarat	YRI	Satnami	Portugal	3.37
Siddi_Gujarat	YRI	Meghawal	Portugal	0.95
Siddi_Gujarat	YRI	Pandit	Portugal	0.29
Siddi_Gujarat	YRI	Naidu	Portugal	2.99
Siddi_Gujarat	YRI	Velama	Portugal	2.57
Siddi_Gujarat	YRI	Vysya	Portugal	6.42
Siddi_Gujarat	YRI	Hallaki	Portugal	5.70
Siddi_Gujarat	YRI	cline	Portugal	2.68
Siddi_Gujarat	YRI	Portugal	Vaish	8.01
Siddi_Gujarat	YRI	Portugal	Tharu	8.61
Siddi_Gujarat	YRI	Portugal	Kurumba	8.98
Siddi_Gujarat	YRI	Portugal	Lodi	7.78
Siddi_Gujarat	YRI	Portugal	Bhil	9.40
Siddi_Gujarat	YRI	Portugal	Madiga	9.02
Siddi_Gujarat	YRI	Portugal	Mala	8.49
Siddi_Gujarat	YRI	Portugal	Kamsali	8.38
Siddi_Gujarat	YRI	Portugal	Satnami	7.92
Siddi_Gujarat	YRI	Portugal	Meghawal	6.99
Siddi_Gujarat	YRI	Portugal	Pandit	8.53
Siddi_Gujarat	YRI	Portugal	Naidu	8.12
Siddi_Gujarat	YRI	Portugal	Velama	7.48
Siddi_Gujarat	YRI	Portugal	Vysya	8.54
Siddi_Gujarat	YRI	Portugal	Hallaki	6.59
Siddi_Gujarat	YRI	Portugal	cline	9.90
Siddi_Gujarat	YRI	ICP	MEX	0.87

NOTE: The Z-scores evaluates the significance of incorporating Ancestral pop3 in model - that is compares the model with ancestral pop3 against a null model with Ancestral pop3 excluded. Highlighted in bold implies that the model of Ancestral Pop1 + Ancestral Pop2 provides a good fit to the data. “Indian cline” - combined data from 16 Indian groups- represents the ancestral Indian group. ICP - combined data from 16 Indian groups and Portuguese- represents the ancestral non-African population. HapMap3 Mexican ancestry in Los Angeles, California (MEX) are used to test if the YRI + ICP provides a good fit to the Siddi data as the variation in the Mexicans is unrelated to the genetics of Siddis.

Table S2C. Testing the robustness of the models that emerges from Table S2b

Admixed pop	Ancestral pop1	Ancestral pop2	Ancestral pop3	Z-score
Siddi_Karnataka-1				
Siddi_Karnataka-1	YRI	Vaish	Portugal	0.52
Siddi_Karnataka-1	YRI	Tharu	Portugal	-0.75
Siddi_Karnataka-1	YRI	Meghawal	Portugal	1.68
Siddi_Karnataka-1	YRI	Kashmiri Pandit	Portugal	0.62
Siddi_Karnataka-1	YRI	Naidu	Portugal	1.43
Siddi_Karnataka-1	YRI	Velama	Portugal	1.21
Siddi_Karnataka-1	YRI	Indian cline	Portugal	0.25
Siddi_Karnataka-1	YRI	ICP	MEX	-0.89
Siddi_Karnataka-2				
Siddi_Karnataka-2	YRI	Vaish	Portugal	-0.10
Siddi_Karnataka-2	YRI	Tharu	Portugal	-0.72
Siddi_Karnataka-2	YRI	Meghawal	Portugal	1.36
Siddi_Karnataka-2	YRI	Kashmiri Pandit	Portugal	-0.12
Siddi_Karnataka-2	YRI	Naidu	Portugal	0.51
Siddi_Karnataka-2	YRI	Velama	Portugal	1.07
Siddi_Karnataka-2	YRI	Indian cline	Portugal	1.72
Siddi_Karnataka-2	YRI	ICP	MEX	1.80

NOTE: ICP - combined data from 16 Indian groups and Portuguese- represents the ancestral non-African population and Indian cline - represents combined data for 16 Indian groups.

Table S2D. Testing the robustness of the models to the African population chosen for the analysis

Admixed	Ancestral pop1	Ancestral pop2	Ancestral pop3	Z-score
Siddi_Gujarat				
Siddi_Gujarat	LWK	Vaish	Portugal	1.07
Siddi_Gujarat	LWK	Tharu	Portugal	0.87
Siddi_Gujarat	LWK	Meghawal	Portugal	1.22
Siddi_Gujarat	LWK	Kashmiri Pandit	Portugal	1.08
Siddi_Gujarat	LWK	Naidu	Portugal	1.16
Siddi_Gujarat	LWK	Velama	Portugal	2.07
Siddi_Gujarat	LWK	Indian cline	Portugal	2.07
Siddi_Gujarat	LWK	ICP	MEX	-1.12

Table S3. Result observed with the Batwing Analysis for determining the effective male population

MCMC cycles	N posterior				
	2.5%	Median	97.5%	Plus error	Minus error
Full dataset					
10 ⁶	1,675	2,794	5,964	3,169	1,120
10 ⁷	689	1,729	3,683	1,955	1,039
10 ⁸	699	1,666	3,721	2,055	967
10 ⁹					
Subset of 40 random chromosomes					
10 ⁶	966	1,881	4,164	2,283	915
10 ⁷	697	1,704	4,360	2,656	1,007
10 ⁸	588	1,463	3,996	2,533	876
10 ⁹					
Biased subset lacking haplogroup B					
10 ⁶					
10 ⁷	604	1,392	2,385		
10 ⁸					
10 ⁹					

NOTE-A random subset of 40 samples was analyzed using 10⁶ to 10⁸ MCMC cycles and we obtained the same posterior probability for effective population size (N) as that obtained for 10⁷ cycles. We estimated the effective male population size of the African ancestors of Siddis brought to India as ~1,400 individuals

Table S4. *G6PD* variants observed in the Siddis and other Indian populations

Location (State)	Population	Linguistic Family	N	<i>G6PD</i> variant	
				A - variant rs1050828 rs1050829	Med variant rs5030868
Andhra Pradesh	Dudekula	Dravidian	18	-	-
	Erukala	Dravidian	39	-	-
Chhattisgarh	Patkar	Dravidian	11	-	-
	Thalari	Dravidian	12	-	-
	Vysya	Dravidian	25	-	-
Gujarat	Sindhi	Indo-European	91	-	5
	Bhil	Indo-European	24	-	-
Kerala	Jat	Indo-European	8	-	-
	Siddi	Indo-European	60	6	1
	Kurumba	Dravidian	9	-	-
Madhya Pradesh	Muduga	Dravidian	25	-	-
	Muduvar	Dravidian	62	-	-
Nagaland	Rajgond	Indo-European	24	-	-
	Sonr	Indo-European	72	-	-
Uttar Pradesh	Ao Naga	Tibeto-Burman	35	-	-
	Nagasema	Tibeto-Burman	38	-	-
	Chakhesang Naga	Tibeto-Burman	16	-	-
Uttaranchal	Yadav	Indo-European	22	-	-
	Lodhi	Indo-European	47	-	-
Karnataka	Bisth	Indo-European	36	-	-
	Uniyal	Indo-European	34	-	-
	Siddi	Dravidian	67	7	-
Tamil Nadu	Kare Vokkal	Dravidian	30	-	-
	GramVokkal	Dravidian	54	-	-
	Korava	Dravidian	38	-	-
	Medar	Dravidian	56	-	-
Tamil Nadu	Khani	Dravidian	46	-	-
	Badaga	Dravidian	57	-	-
Total			1056		

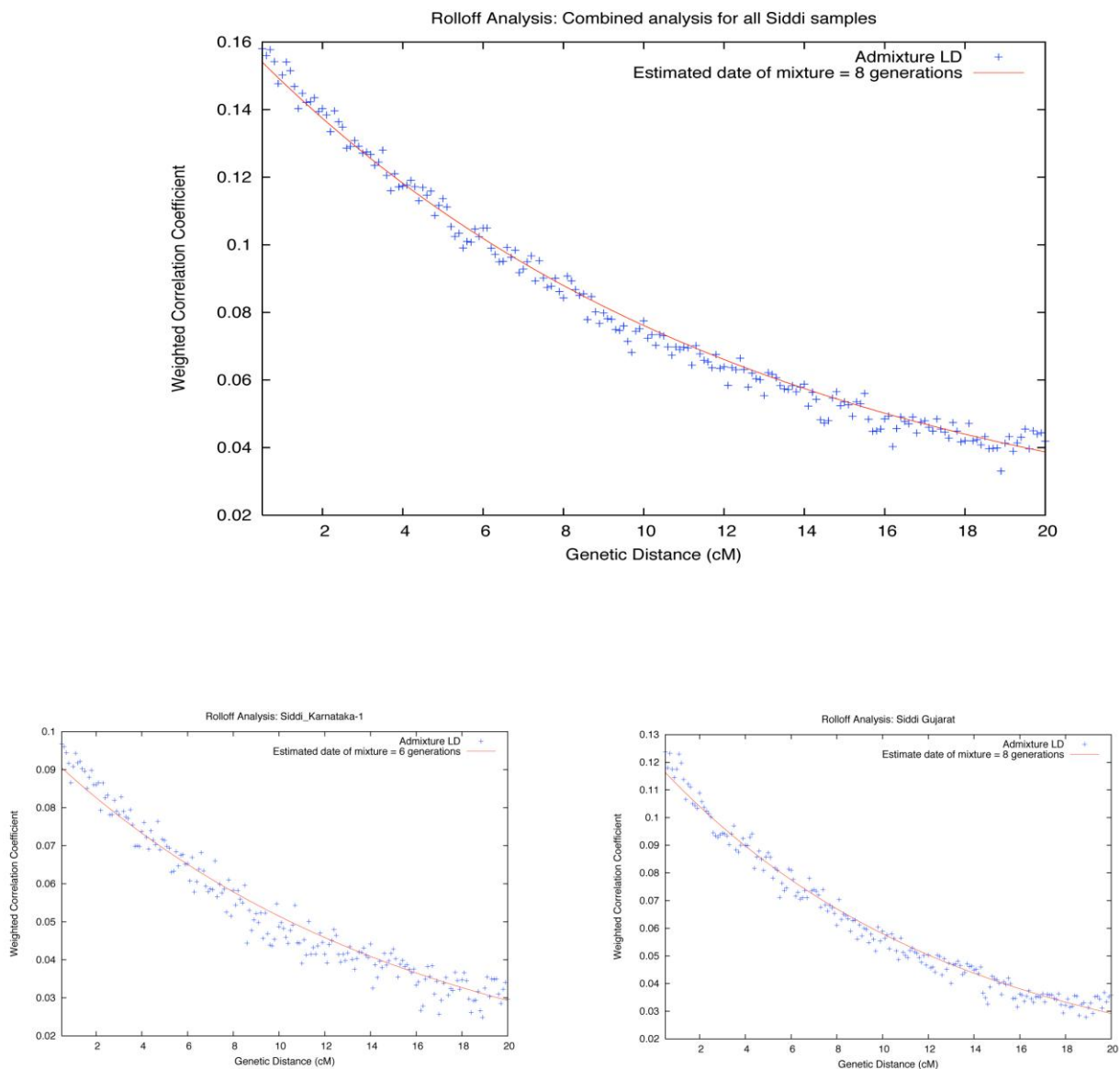


Figure S1 – ROLLOFF Analysis using East Africans as the ancestral population

We performed the ROLLOFF analysis to estimate the date of admixture in Siddis, using East Africans (LWK) and ICP samples as the reference populations. The estimated dates of mixture are as following: Siddi-Karnataka-1 = 6 ± 1 generation, Siddi-Gujarat = 8 ± 1 generation and the result for the combined data set ($n = 12$) is 8 ± 1 generation or ~ 200 years. Due to limited number of samples, we were not able to perform separate analysis of the Siddi_Karnataka-2 group.

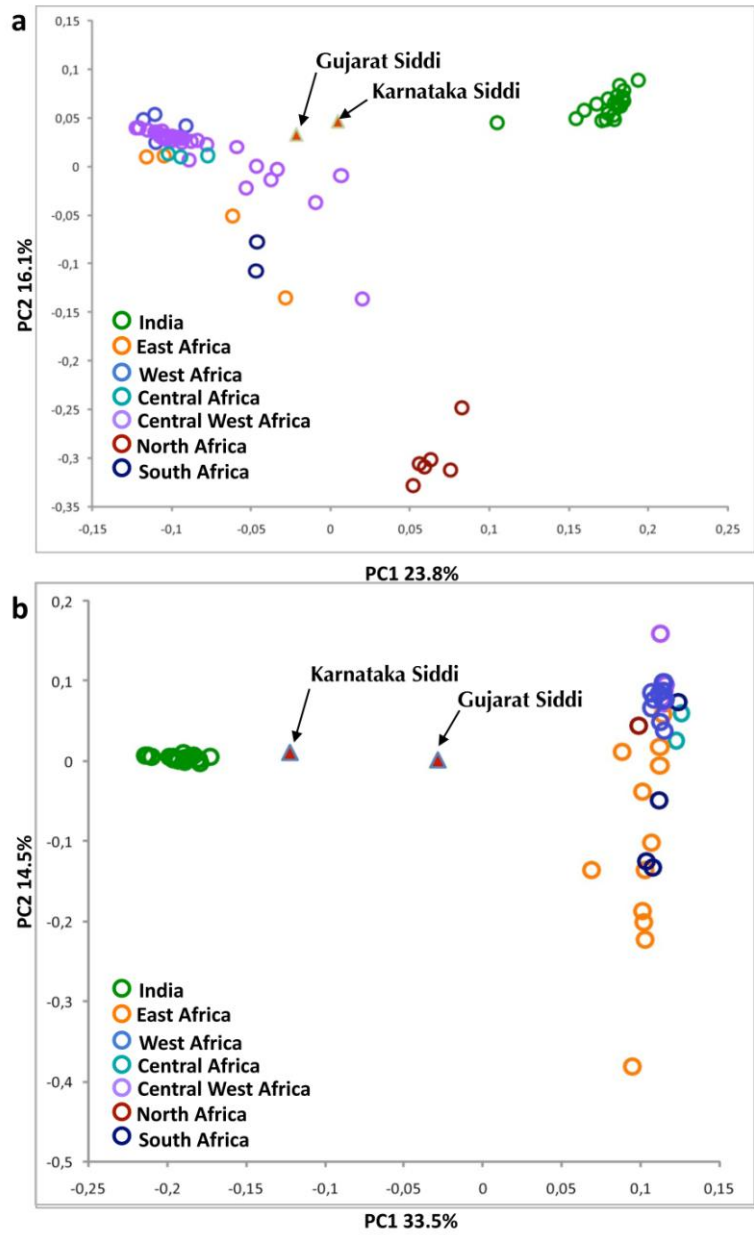


Figure S2- Principal Component Analysis (PCA) using data for uniparentally transmitted markers

PCA was performed using Y-chromosome and mitochondrial DNA haplogroups frequencies among Siddis, Indian and African populations. (a) PCA based on Y-chromosome haplogroup frequencies (Supporting Dataset S1) shows that the Siddis from Gujarat and Karnataka are related to African populations and nearby Indian populations. The Siddis also appear to be most closely related to the Bantu speaking populations from West and Central West Africa. (b) PCA based on mitochondrial DNA (mtDNA) haplogroup frequencies (Supporting Dataset S2) shows the Siddi falling between the two major clusters on PC1 of Indians and Africans. The Siddis from Karnataka state appear to be closer to the Indians than Africans, likely because of high level of admixture with the nearby Indian groups.

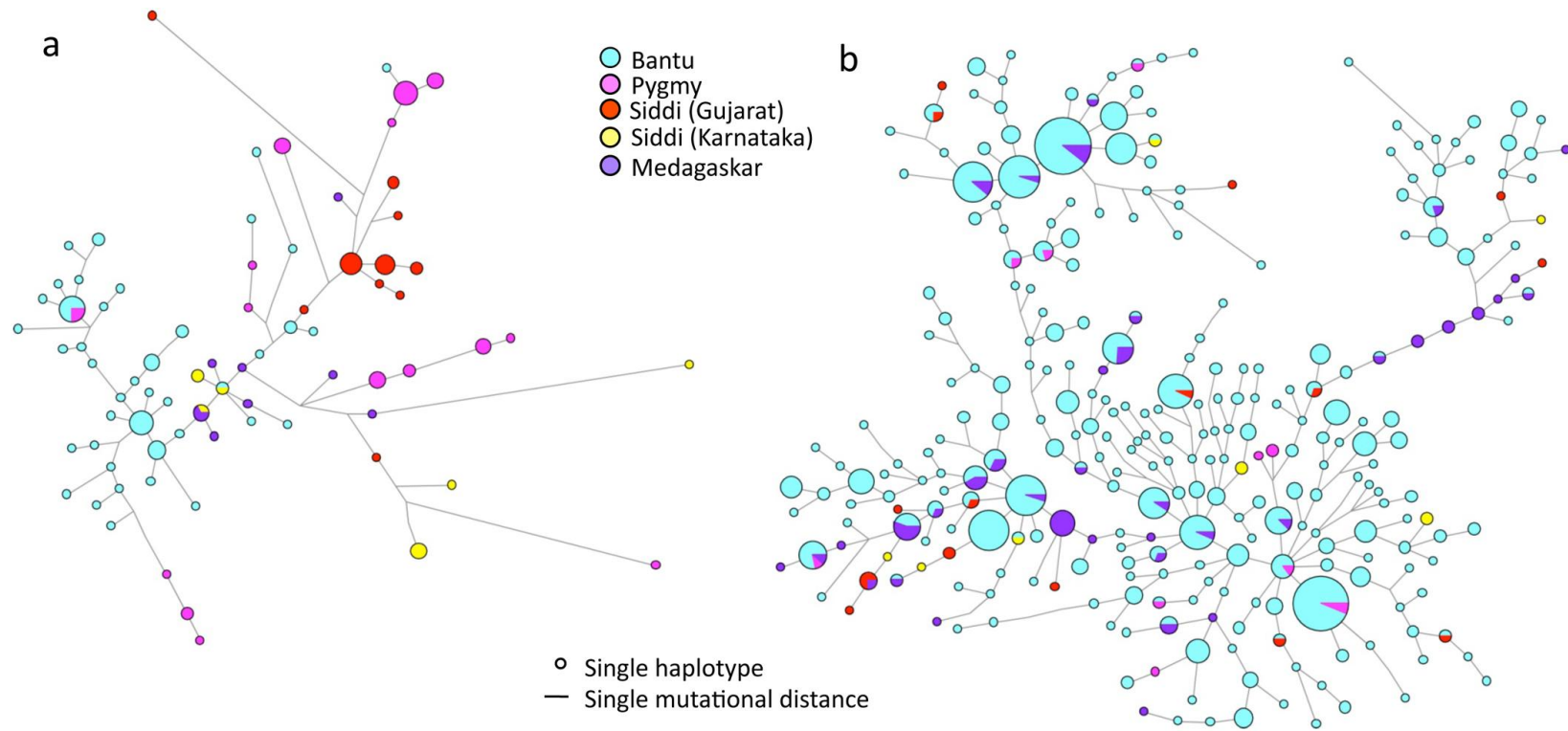


Figure S3- Phylogenetic Network analysis

Phylogenetic networks were constructed based on ten common Y-chromosomal short tandem repeats (Y-STRs). (a) Haplotype networks were generated for haplogroups B2 and (b) E1b1 by using the median-joining algorithm of Network 4.6 (<http://www.fluxus-engineering.com>). Because of the high level of reticulation in the E1b1a-M2 samples, data were post-processed using the MP calculation option in Network 4.6. The size of the circles is proportional to the number of samples. Published data from Berniell-Lee et al. 2009¹⁸ and Tofanelli et al. 2009¹⁹ was used for Bantu, Pygmy and Madagascar populations.

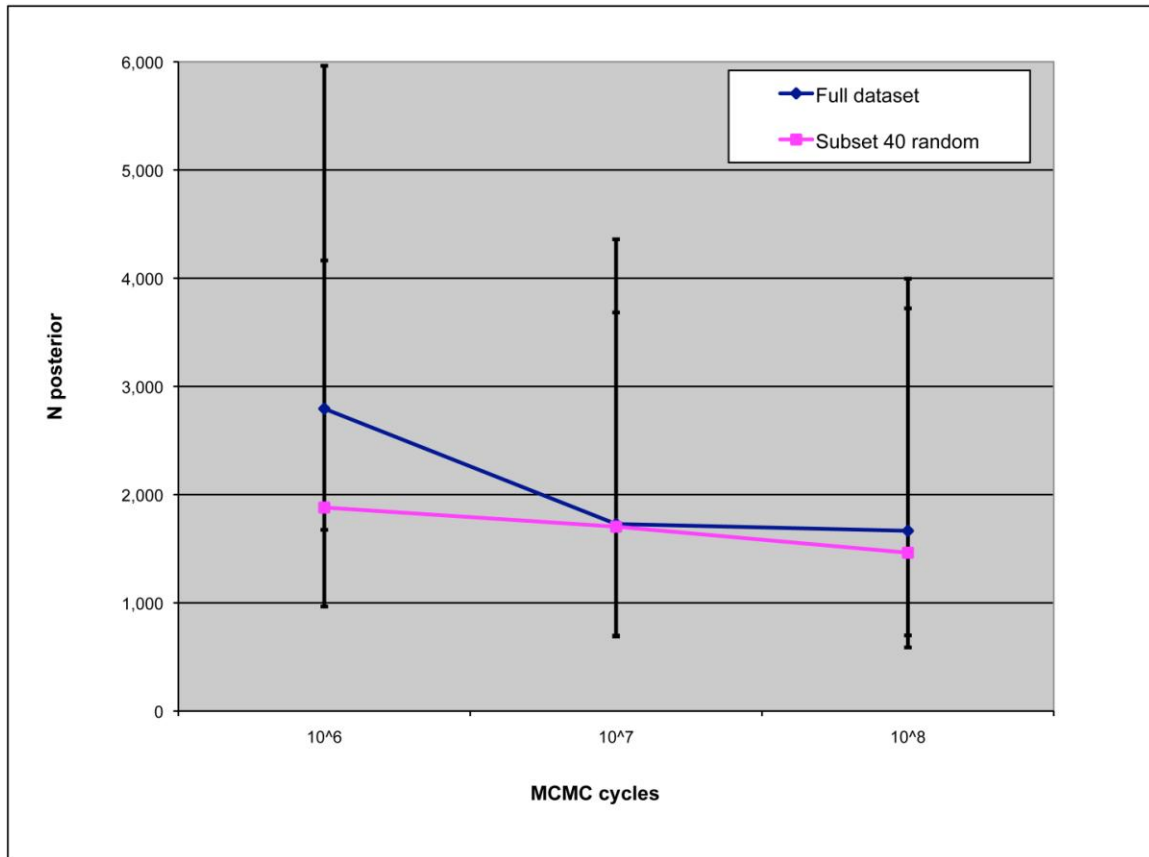


Figure S4- Optimization of parameters for BATWING analysis

We performed BATWING analysis with a demographic model that assumes a period of constant size followed by exponential growth. To find the optimal number of MCMC cycles at which the method converges, we varied the MCMC cycles between 10^6 - 10^8 cycles and compared the results based on full dataset and a random subset of 40 samples. The number of MCMC cycles performed is shown on the X-axis and the posterior value of the effective population size (N) is shown on the Y-axis.

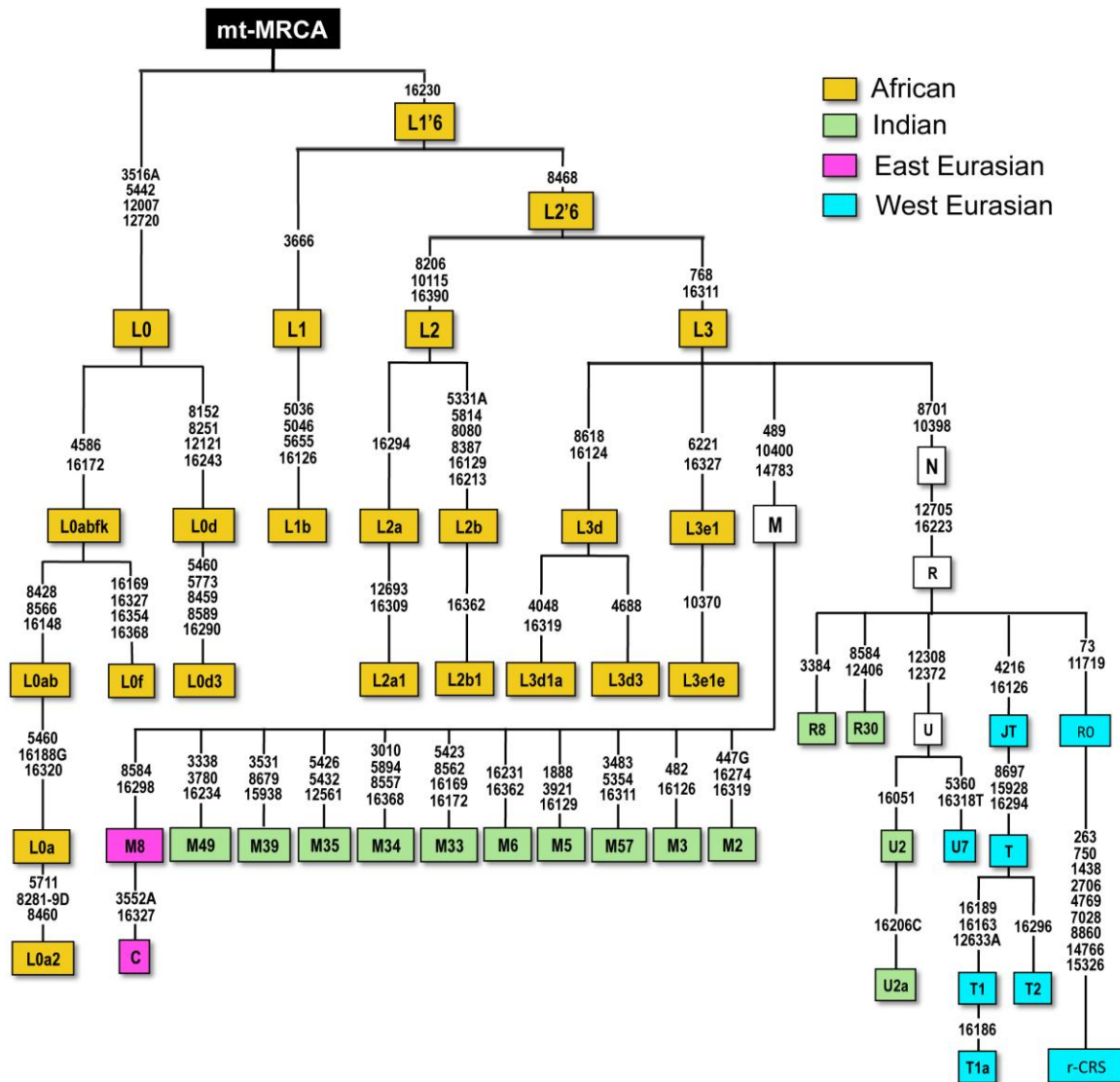


Figure S5- Phylogenetic tree based on mitochondrial DNA markers

We performed mtDNA markers analysis for 153 Siddis and 269 individuals from the nearby Indian populations (Supporting Dataset S2). The following maternal lineages – African-specific L haplogroups, Indian-specific M, R and U lineages, East Asian branch of M and the Eurasian specific branches of R- were seen in the Siddi samples.