

## Indian Siddis: African Descendants with Indian Admixture

Anish M. Shah,<sup>1,9</sup> Rakesh Tamang,<sup>1,9</sup> Priya Moorjani,<sup>2,3,9</sup> Deepa Selvi Rani,<sup>1</sup> Periyasamy Govindaraj,<sup>1</sup> Gururaj Kulkarni,<sup>4</sup> Tanmoy Bhattacharya,<sup>1</sup> Mohammed S. Mustak,<sup>5</sup> L.V.K.S. Bhaskar,<sup>1</sup> Alla G. Reddy,<sup>1</sup> Dharmendra Gadhvi,<sup>1</sup> Pramod B. Gai,<sup>4</sup> Gyaneshwer Chaubey,<sup>6</sup> Nick Patterson,<sup>3</sup> David Reich,<sup>2,3</sup> Chris Tyler-Smith,<sup>7</sup> Lalji Singh,<sup>1,8</sup> and Kumarasamy Thangaraj<sup>1,\*</sup>

The Siddis (Afro-Indians) are a tribal population whose members live in coastal Karnataka, Gujarat, and in some parts of Andhra Pradesh. Historical records indicate that the Portuguese brought the Siddis to India from Africa about 300–500 years ago; however, there is little information about their more precise ancestral origins. Here, we perform a genome-wide survey to understand the population history of the Siddis. Using hundreds of thousands of autosomal markers, we show that they have inherited ancestry from Africans, Indians, and possibly Europeans (Portuguese). Additionally, analyses of the uniparental (Y-chromosomal and mitochondrial DNA) markers indicate that the Siddis trace their ancestry to Bantu speakers from sub-Saharan Africa. We estimate that the admixture between the African ancestors of the Siddis and neighboring South Asian groups probably occurred in the past eight generations (~200 years ago), consistent with historical records.

Siddis, or Habshis, are a unique tribe that has African ancestry and lives in South Asia. They are mainly found in three Indian states—Gujarat, Karnataka, and Andhra Pradesh—and according to the latest census, their total population size is about 0.25 million.<sup>1</sup> The first documented record of Siddis in India dates to 1100 AD, when the Siddis settled in Western India.<sup>2,3</sup> By the thirteenth century, substantial numbers of Siddis were being imported by the Nawabs and the Sultans of India to serve as soldiers and slaves. The major influx of Siddis occurred during the 17<sup>th</sup>–19<sup>th</sup> centuries, when the Portuguese brought them as slaves to India.<sup>2</sup> Previous genetic studies have shown that the Siddis have ancestry from up to three continental groups: Africans, Europeans, and South Asians.<sup>2,4,5</sup> Some genetic studies have suggested that they are most closely related to Africans.<sup>3,6</sup> However, the specific African group to which the Siddis trace their ancestry remains unknown. To obtain a high-resolution genome-wide perspective of ancestry, we analyzed data from three Siddi groups (from Karnataka and Gujarat) by genotyping them with ~850,000 autosomal and sex-linked markers. Applying statistical methods, we have estimated the contributions of various continental ancestries to the Siddis genome and investigated the likely source of the ancestral populations and the timing of the admixture events.

Blood samples (about 10 ml from each individual) were collected from Gujarat and Karnataka in India. Specifically, we collected samples from 60 Siddis (unrelated and healthy males) and 90 individuals belonging to the nearby tribal populations (Charan and Bharwad) of the Junagarh district of Gujarat and from 94 Siddis (65 males and 29

females) and 178 individuals belonging to neighboring tribal populations (Medar, Gram Vokkal, Kare Vokkal, and Korova) from the Uttara Kannad district of Karnataka. Informed written consent was obtained from all the donors. This project was approved by the Institutional Ethical Committee of the Centre for Cellular and Molecular Biology, Hyderabad, India. We genotyped 16 Siddi samples on Affymetrix (SNP 6.0) arrays by using standard protocols. We removed four duplicate samples and restricted the analysis to SNPs that had <5% missing data (846,418 SNPs). Two merged datasets were created for further analysis. Dataset I contained Siddi data that were merged with data from the International Haplotype Map (HapMap) phase 3<sup>7</sup> (n = 1,115 samples from 11 populations genotyped on an Illumina 1M array) and the India Project (n = 132 individuals from 25 groups genotyped on an Affymetrix 6.0 array). The merged dataset contained 574,197 SNPs. Dataset I was used for the principal-components analysis reported in Figure 1. Dataset II contained the Siddi data merged with three other datasets: the Population Reference Sample (POPRES)<sup>8</sup> (n = 3,845 samples from 37 populations genotyped on an Affymetrix 500K array), the International Haplotype Map (HapMap) phase 3<sup>7</sup> (n = 1,115 samples from 11 populations genotyped on an Illumina 1M array), and the India Project<sup>9</sup> (n = 132 individuals from 25 groups genotyped on an Affymetrix 6.0 array). The merged dataset contained 257,840 SNPs. Dataset II was used for estimating the admixture proportions and dates of admixture.

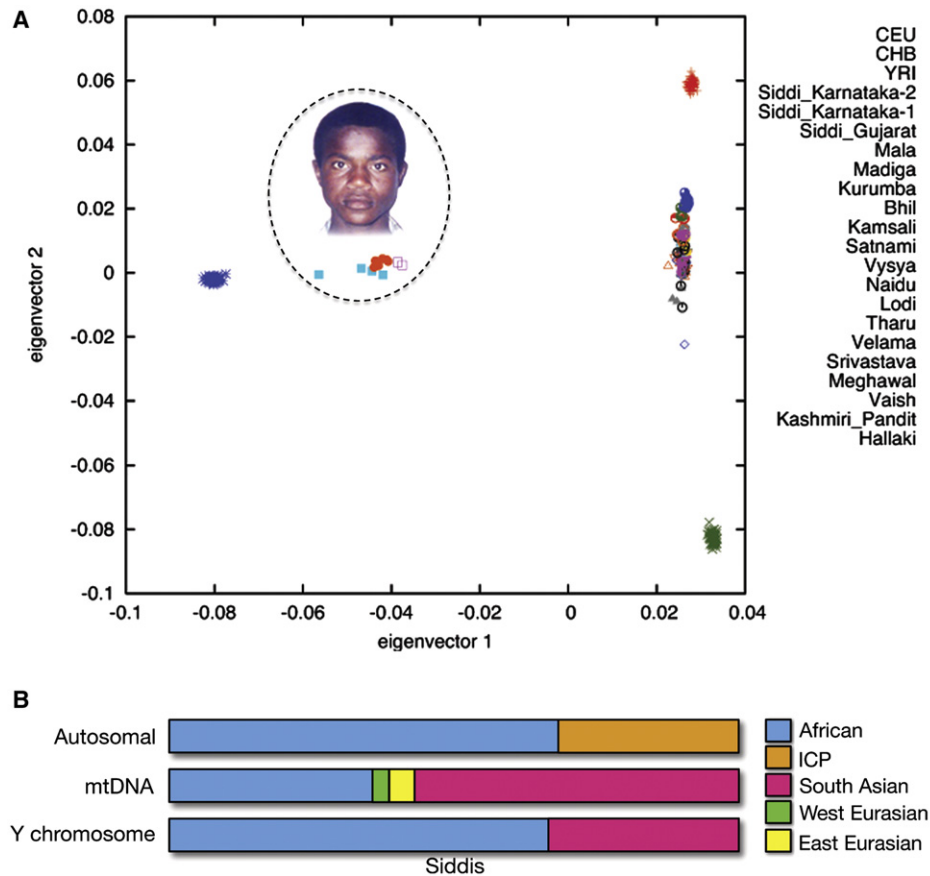
To explore patterns of population structure in the Siddis and to test their genetic affinity to other groups worldwide, we analyzed autosomal data from 12 individuals from

<sup>1</sup>Centre for Cellular and Molecular Biology, Council of Scientific and Industrial Research, Hyderabad 500007, India; <sup>2</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115, USA; <sup>3</sup>Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; <sup>4</sup>Department of Applied Genetics, Karnataka University, Dharwad 580003, Karnataka, India; <sup>5</sup>Department of Zoology, Mangalore University, Mangalore 574 119, India; <sup>6</sup>Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu and Estonian Biocentre, Tartu 51010, Estonia; <sup>7</sup>The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridgeshire CB10 1SA, UK; <sup>8</sup>Genome Foundation, Hyderabad 500007, India

<sup>9</sup>These authors contributed equally to this work

\*Correspondence: [thangs@ccmb.res.in](mailto:thangs@ccmb.res.in)

DOI 10.1016/j.ajhg.2011.05.030. ©2011 by The American Society of Human Genetics. All rights reserved.



**Figure 1. African Ancestries in Siddis**

(A) Principal-components analysis of three Siddi groups, HapMap Phase 3 populations (CEU, YRI, and CHB), and 16 Indian groups. (B) Schematic representation of the proportions of African and ICP ancestry in Siddis. ICP includes the 16 Indian groups and Portuguese and represents the ancestral non-African population.

three Siddi groups (six individuals from Karnataka and six individuals from Gujarat), 128 individuals from 16 Indian groups (Mala, Madiga, Kurumba, Bhil, Kamsali, Satnami, Vysya, Naidu, Lodi, Tharu, Velama, Srivastava, Meghawal, Vaish, Kashmiri Pandit, and Hallaki), and 300 individuals from three HapMap populations (Yoruba from Ibadan, Nigeria [YRI], Utah residents with Northern and Western European ancestry [CEU], and Han Chinese from Beijing, China [CHB]).<sup>7,9</sup> The 16 Indian groups were chosen because they spanned a high degree of diversity within India. It had been previously shown that most Indian populations have ancestry from two highly divergent groups: an Ancestral North Indian (ANI) population that is closely related to West Eurasians and an Ancestral South Indian (ASI) population that is not related to any population outside India. The ANI ancestry proportion lies within the range of 39%–71% across the 16 groups chosen.<sup>9</sup> The ANI and ASI have been inferred to be highly differentiated at the time that they mixed, and Reich et al. (2009)<sup>9</sup> estimated that the average allele frequency differentiation,  $F_{ST}$  (ANI, ASI), is  $\sim 0.09$ .

We performed principal-components analysis (PCA) on the autosomal SNP data with the EIGENSOFT software.<sup>10</sup>

A plot of the first and second principal components (PCs) suggests that the Siddis have ancestry from Africans as well as Eurasians (Figure 1A). Like other Indian populations, Siddis have both ANI and ASI ancestry, but they lie off the main cline of ANI-ASI admixture and are closely related to African individuals (Figure 1A). The average allele frequency differentiation between the two Siddi groups (Karnataka and Gujarat) is relatively high;  $F_{ST}$  (Siddi\_Karnataka-1, Siddi\_Gujarat) = 0.02 (Table S1), suggesting that the populations differ substantially, possibly as a result of endogamy, different ancestral origins, or admixture with different local South Asian groups. However, the diversity in the Siddis is not correlated with geography in our small sample; the individuals from the Karnataka-2 group are genetically closer to the Gujarat Siddis ( $F_{ST}$  [Siddi\_Karnataka-2, Siddi\_Gujarat] = 0.002) than to the other group from Karnataka ( $F_{ST}$  [Siddi\_Karnataka-1, Siddi\_Karnataka-2] = 0.026) (Table S1, available online). This suggests that the members of Karnataka2 might be recent migrants from Gujarat or that the ancestors of one of the Karnataka samples might have experienced a very strong recent founder event.

Previous genetic studies with traditional biochemical and autosomal markers have suggested that the Siddis have ancestry from up to three distinct ancestral groups: Africans, Indians, and Europeans.<sup>3,6</sup> To formally test whether the Siddis have ancestry from each of these three ancestral populations, we used a regression method proposed by Patterson et al. (2010).<sup>11</sup> This method allowed us to model the allele frequency of the admixed Siddis as a linear combination of the allele frequencies in the ancestral populations and to build optimal models with and without each ancestral population and then compute the error between our model and the data. For example, to test whether the Siddis contain genetic admixture from Africans, we built two models with the data; one included Africans as the ancestral population, and another excluded Africans from the model. Applying this method to the Siddi Gujarat samples, we observed that there is strong evidence that the Siddis have African ancestry (Z score  $\gg 25$ ), but the genetic variation in Africans does not fully explain the underlying genetic data in the Siddis (Table S2A). Next, we assessed whether a two-way model or three-way mixture model provides a better fit to the data. Table S2B shows that a two-way model of African + Portuguese or African + Mala (or any other group that has high ASI ancestry) provides a poor fit to the data. However, the model of African + Vaish (or any other group that has high ANI ancestry) provides just as good a fit to the data as a three-way model of African + any Indian population + Portuguese (Table S2B). This suggests that the Siddis have some West-Eurasian-related (ANI or Portuguese) ancestry, in addition to their African and ASI ancestry. However, the size of our dataset prevents our methods from being sensitive enough to differentiate between ANI and Portuguese ancestry. To represent the ancestral non-African population of the Siddis, we combined the data from 16 Indian groups and the Portuguese (“ICP”). To test the robustness of our models, we analyzed Siddi Karnataka samples with the models built from the Siddi Gujarat samples and showed that the models provided a good fit to the data (Table S2C).

Applying the regression-style method to all three Siddi groups with YRI and ICP as the ancestral populations, we estimated that the Siddis have on average ~67% African ancestry (Table 1). We obtained qualitatively similar results when we used East Africans (HapMap Luhya [LWK]) in place of YRI (Table 1 and Table S2D).

To characterize the temporal impact of admixture and to develop a historical interpretation of the results, we needed not only to qualitatively demonstrate a history of admixture but also to quantitatively estimate a date for the admixture event. We applied the *ROLLOFF* method,<sup>12</sup> which utilizes information related to admixture linkage disequilibrium (LD) to estimate the time since admixture. This method capitalizes on the fact that the genome of an admixed population contains chromosomal segments from ancestral populations, whose length is inversely proportional to the date of admixture. By modeling the

**Table 1. Estimation of Ancestry Proportions in the Siddis**

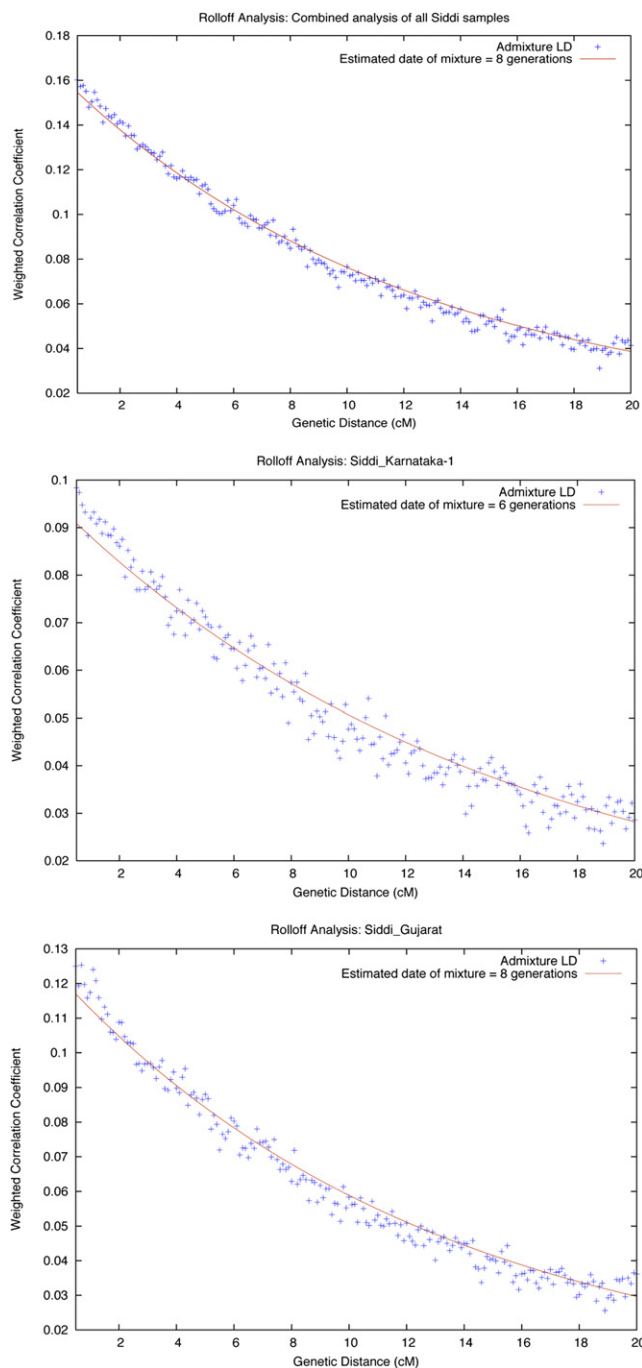
	African Ancestry	Non-African Ancestry
<b>African Ancestral Population = West Africans (YRI)</b>		
Siddi_Gujarat	66.90% $\pm$ 0.59%	33.10% $\pm$ 0.59%
Siddi_Karnataka-1	70.90% $\pm$ 0.65%	29.10% $\pm$ 0.65%
Siddi_Karnataka-2	62.30% $\pm$ 0.99%	37.70% $\pm$ 0.99%
<b>African Ancestral Population = East Africans (LWK)</b>		
Siddi_Gujarat	70.50% $\pm$ 0.66%	29.50% $\pm$ 0.66%
Siddi_Karnataka-1	74.40% $\pm$ 0.71%	25.60% $\pm$ 0.71%
Siddi_Karnataka-2	64.80% $\pm$ 1.11%	35.20% $\pm$ 1.11%

Admixture proportion estimates are based on a regression-style method<sup>11</sup> for which the ancestral populations shown in the table were used. ICP includes combined data from 16 Indian groups and the Portuguese and represents the ancestral non-African population.

decay of the LD in the admixed individuals and weighting it by the allele frequency differentiation in the ancestral populations (such that the statistics are only sensitive to admixture LD), we can precisely estimate the time since the admixture event. Simulations have suggested that this method is robust for data from poor surrogates of ancestral populations and can estimate the date of admixture up to 300 generations ago.<sup>12</sup>

Applying *ROLLOFF* to the Siddis (combining data from all three groups—Siddi\_Karnataka-1, Siddi\_Karnataka-2, and Siddi\_Gujarat—to increase the power), we observed an approximately exponential decay of the weighted correlation with distance, which provides strong evidence of admixture (Figure 2). By using the least-squares method to fit an exponential distribution to this pattern, we estimated an average date of ~eight generations, or 200 years (if one assumes a generation size of 25 years<sup>13</sup>). This approximately coincides with the historical date of arrival of most African ancestors of the Siddis to India. To show that combining the data from the admixed group does not substantially change the results, we ran *ROLLOFF* separately for each admixed group and obtained qualitatively similar results (within two standard errors) for Siddi\_Gujarat and Siddi\_Karnataka-1. Because of the limited number of samples, we were not able to perform analysis for the Siddi\_Karnataka-2 group (*ROLLOFF* analysis requires at least four samples). In addition, changing the African ancestral group to East African Luhya did not change the estimated date of admixture (Figure S1).

To gain insight into the most likely source of the African ancestry in Siddis, we examined paternally inherited Y-chromosomal biallelic markers as well as maternally inherited mtDNA markers. Analysis of data from uniparentally inherited markers can provide information about population genetic relatedness, including probable ancestral source populations and information related to admixture events. We genotyped 32 Y-chromosomal biallelic markers (viz. M94, M60, M182, M168, M130, M145, M96, M75, M2, M89, M82, M304, M172, M9,



**Figure 2. ROLLOFF Analysis of Siddis**

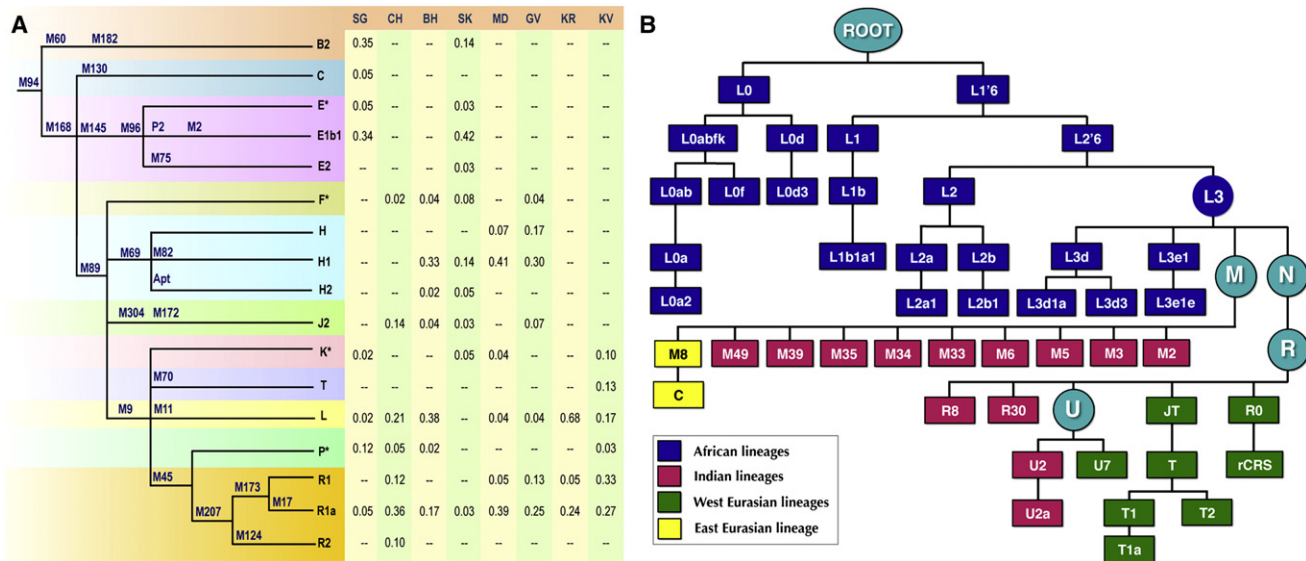
We analyzed 12 Siddi samples from Karnataka and Gujarat and estimated admixture linkage disequilibrium by computing the LD between all pairs of markers and weighting it by the frequency differentiation between the ancestral populations (YRI and ICP). We observed an approximately exponential decay of LD with distance and found an average estimated date of admixture of  $8 \pm 1$  generations. This corresponds to a time of around 200 years (if we assume a generation interval of 25 years). The estimated dates of admixture for Siddi\_Karnataka-1 and Siddi\_Gujarat are  $6 \pm 1$  and  $8 \pm 1$  generations, respectively. Standard errors were computed with a weighted block jackknife as described in Moorjani et al.<sup>12</sup>

M70, M11, M45, M207, M173, M17, M124, M201, M170, M70, M147, M189, M214, M52, M33, M356, P36, and P2) in 125 Siddis and 268 individuals (all males) from nearby Indian groups. We combined our data with published data from 2,301 individuals belonging to 56 different groups from the African subcontinent and 667 individuals from 16 populations from Gujarat, Karnataka, Maharashtra, and Andhra-Pradesh in India (Document S2).<sup>14–26</sup>

We observed that the Y-chromosomal haplogroups B2-M182 and E1b1a-M2, which are characteristic of African ancestry, were present at high frequencies in the Siddis but not in other Indians. Moreover, about 70% of the Siddi male lineages fall into haplogroups generally characteristic of African populations (Figure 3A), thus confirming the results from the autosomal DNA markers (Figure 1B). The remaining 30% were C\*-M130- and M89-derived Indian or Near-Eastern lineages (H1a-M82, H2-Apt-H2, J2-M172, L-M11, and P\*-M45). The populations neighboring the Siddis were found to harbor only these Asian-specific haplogroups. It is interesting to note that none of the African paternal lineages were observed among the neighboring Indian groups, whereas Indian-specific lineages were detected in Siddi individuals. This suggests primarily unidirectional paternal gene flow from Indian populations to the Siddis (Figure 2B).

To learn more about the source of the African paternal lineages, we performed PCA with a merge of our Y-chromosomal dataset (Siddis and neighboring Indian groups) with data from 2,301 individuals from 56 African populations (Document S2). A plot of the first and second PCs showed that the Siddis cluster with Bantu-speaking populations of sub-Saharan Africa (Figure S2A). Previous studies have proposed that the E3a (currently known as E1b1a), E2, and B2 haplogroups are associated with the Bantu expansions within Africa.<sup>21,22,27</sup> The presence of these haplogroups in the Siddis suggests that their ancestors might have been part of this expansion. To investigate this possibility, we typed 17 Y-STRs by using multiplex PCR and the Y-filer kit (Applied Biosystems, Foster City, USA) in reaction volumes of 10  $\mu$ l with 1U of AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, USA), 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 250  $\mu$ M dNTPs, 3.0  $\mu$ M of each primer (forward primers were fluorescently labeled), and 1 ng of DNA template. Thermal cycling conditions were as follows: (1) 95°C for 11 min, (2) 30 cycles as follows: 94°C for 1 min, 61°C for 1 min, and 72°C for 1 min, (3) 60°C for 80 min, and (4) 25°C hold. The PCR amplicons along with GS500 LIZ (as a size standard) were run in the ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, USA). The raw data were analyzed with the GeneMapper v4.0 software program (Applied Biosystems, Foster City, US).

We excluded two DYS385 loci from the current analyses because they could not be distinguished via the typing method employed, and we renamed locus DYS389I as DYS389b, whereas we calculated DYS389a by subtracting DYS389I from DYS389II. We constructed median-joining



**Figure 3. Y-Chromosomal and mtDNA Haplogroups in Siddis**

(A) Y-chromosomal haplogroup frequencies in the populations analyzed. Abbreviations are as follows: SG, Siddis from Gujarat; CH, Charan; BH, Bharwad; SK, Siddis from Karnataka; MD, Medar; GV, Gram Vokkal; KR, Korova; and KV, Kare Vokkal. The African-specific B and E1b1a haplogroups in India were found exclusively among the Siddi population.

(B) Distribution of mtDNA haplogroups in Siddis. Details of diagnostic mutations that define haplogroups are shown in Figure S5.

networks with ten common loci (Figure S3) for the two major African haplogroups (E1b1a-M2 and B2-M182) that are present at high frequencies in Siddis. We supplemented our dataset with other published data that included African samples.<sup>28,29</sup> The TMRCA (time to most recent common ancestor) was estimated with the  $\rho$  statistic (the mean number of mutations from the assumed root), for which a 25-year generation time was used, and the TD statistic (for both, a mutation rate of  $6.9 \times 10^{-4}$  per STR per generation was assumed),<sup>30</sup> The majority of the Siddis haplotypes were found shared on otherwise Bantu-specific branches and were present all over the tree (Figure S3). In addition, the Gujarat and Karnataka Siddis were highly diverged and did not share any haplotypes. These results support the autosomal observation of high  $F_{st}$  differentiation among Siddis from Gujarat and Karnataka. Although the majority of the Siddi haplotypes were scattered in the network, we found that all haplogroup B2 Gujarat Siddis formed a cluster and coalesced to their most recent common ancestor  $2.4 \pm 1$  Kya (thousand years ago). The sharing of haplotypes suggests relatedness among the samples. This is similar to the results seen in the autosomal analyses of the Siddi\_Gujarat and Siddi\_Karnataka-2 samples. The male effective population size was estimated with BATWING<sup>31,32</sup>, via a demographic model that assumes a period of constant size followed by exponential growth (the prior probabilities for the other parameters used in the model were set as previously described).<sup>26</sup> A random subset of 40 samples was analyzed after  $10^6$  to  $10^8$  MCMC cycles, and we obtained the same posterior probability for effective population size (N) as that obtained after  $10^7$  cycles. The effective population size of the African ancestors of Siddis brought to India during

the slave trade was estimated as  $\sim 1,400$  individuals (Table S3 and Figure S4).

To gain insight into the maternal lineages and to test the directionality of maternal gene flow in the Siddis, we assayed the hypervariable region I (HVRI) of mtDNA in 153 Siddis and 269 individuals from the nearby Indian populations (accession numbers JN022021–JN022442). These data were compared with those from the revised Cambridge Reference Sequence (rCRS),<sup>33</sup> and variations were scored. We assigned haplogroups on the basis of HVRI variations and further confirmed these by genotyping the coding-region mutations published to date.<sup>34</sup> The mtDNA haplogroup distributions in the Siddis are shown in Figure 3B, Document S3, and Figure S5.

PCA plots of the combined dataset (Document S3) showed that the African-specific mtDNA haplogroups were present at high frequency in the Siddis; these results were similar to the observations from the autosomal and paternal lineages (Figure S2B). The African-specific haplogroup L was present at a frequency of 53% and 24% in Siddis from Gujarat and Karnataka, respectively. Previous studies have suggested that the L0a, L2a, L3b, and L3e haplogroups are associated with the Bantu expansion.<sup>35–39</sup> Haplogroup L2a (including L2a1) was observed in the Siddis along with rare sublineages of L2, which further supports the conclusion that the ancestors of the Siddis were most likely African Bantus (Figure S5). The L0d lineage, which is now largely confined to the Khoisan-speaking South African populations but which was possibly more widespread in the past,<sup>40</sup> was also observed in two Siddi individuals from Gujarat state. The presence of Indian-specific sublineages of M and N (R and U, which include M2, M3, M5, M6, M33, M35, M39, M57, R8,



**Figure 4. Migration History of the Siddis**

Dotted arrows represent the expansion of Bantu speakers with agriculture within Africa. Agriculture in Africa started from the central western part of the continent and proceeded toward the east and south of the subcontinent. The shaded gray area represents the Portuguese territory, and the lines between Africa and India represent the path that the Portuguese might have used during the 15<sup>th</sup>–19<sup>th</sup> centuries to supply African slaves to Indian rulers on the western coast of India.

R30, and U2 haplogroups) is indicative of recent admixture with indigenous Indian populations (Figure S5).<sup>26</sup> In addition, haplogroup T, which is widespread in southern and Western Europe<sup>41</sup> and is also present at a low frequency in some South Asian groups<sup>42</sup> was present in four Siddi individuals (Figure S5). This suggests maternal gene flow from a West Eurasian ancestral source—perhaps Portuguese or Indian. Consistent with the Y-chromosomal results, there is no evidence of African haplogroups in the neighboring Indian populations, thus confirming the hypothesis of unidirectional gene flow to Siddi individuals from contemporary Indian populations (Figure 1B).

In order to further explore the evidence of sub-Saharan ancestry, we analyzed data for the *G6PD* (MIM 305900) variants in Siddis along with 26 ethnic populations from India. The A– variant, which provides protection against malarial infection and is estimated to have a sub-Saharan African origin between 3,840 to 11,760 YBP,<sup>43</sup> was observed only in Siddis (10%) and not in any other Indian populations (Table S4). This further strengthens the evidence for the sub-Saharan ancestry of the Siddis.

In conclusion, our combined analysis of genetic variation in the Siddis, involving high-resolution sex-linked and autosomal markers, provides strong evidence of African ancestry together with unidirectional gene flow from local Indian groups to the Siddis. The directionality of gene flow supports the complex genetic structuring among Indian populations, which are highly influenced by social norms. We have traced the likely ancestry of Indian Siddis to sub-Saharan African Bantus. The ancestry proportions based on the analysis of autosomal and Y-chromosomal markers are similar, whereas mtDNA markers reveal more likely South Asian lineages among Siddi individuals. The model that emerges from our results is as follows: During the course of the Bantu expansion, African farmers settled in East Africa. Later, during the 15<sup>th</sup> to 17<sup>th</sup> centuries, this region was predominantly ruled by the Portuguese. They brought some Africans to India as slaves and sold them to local Nawabs and Sultans, whose descendants, admixed with neighboring populations, comprise the present-day Siddi population of India (Figure 4).

## Supplemental Data

Supplemental Data include four tables, five figures, and two datasets and can be found with this article online at <http://www.cell.com/AJHG/>.

## Acknowledgments

We thank all the donors, who have voluntarily donated their blood samples for this study. K.T. was supported by the UK-India Education and Research Initiative (RG-4772), the Indian Council of Medical Research, and the Council of Scientific and Industrial Research, Government of India. C.T.S. was supported by The Wellcome Trust, and G.C. was supported by the Centre of Excellence of the Estonian Biocentre. D.R. and P.M. were supported by the National Science Foundation. L.S. was supported by a Bose Fellowship from the Department of Science and Technology, Government of India and a Bhatnagar fellowship from the Council of Scientific and Industrial Research, Government of India. *ROLLOFF* software is available upon request from P.M.

Received: February 1, 2011

Revised: May 15, 2011

Accepted: May 31, 2011

Published online: July 7, 2011

## Web Resources

The URLs for data presented herein are as follows:

Online Mendelian Inheritance in Man (OMIM), <http://www.omim.org>

Phylotree, <http://www.phylotree.org/>

## Accession Numbers

The sequences of the mtDNA hypervariable region I (HVRI) from 153 Siddis and 269 individuals from the nearby Indian populations were deposited in the GenBank database under accession numbers JN022021–JN022442.

## References

- Lodhi, A. (1992). African settlements in India. *Nordic Journal of African Studies* 1, 83–86.
- Bhattacharya, D. (1970). Indians of African origin. *Cah. Etud. Afr.* 10, 579–582.
- Gauniyal, M., Chahal, S.M., and Kshatriya, G.K. (2008). Genetic affinities of the Siddis of South India: an emigrant population of East Africa. *Hum. Biol.* 80, 251–270.
- Thangaraj, K., Ramana, G.V., and Singh, L. (1999). Y-chromosome and mitochondrial DNA polymorphisms in Indian populations. *Electrophoresis* 20, 1743–1747.
- Ramana, G.V., Su, B., Jin, L., Singh, L., Wang, N., Underhill, P., and Chakraborty, R. (2001). Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *Eur. J. Hum. Genet.* 9, 695–700.
- Gauniyal, M., Aggarwal, A., and Kshatriya, G.K. (2011). Genomic Structure of the Immigrant Siddis of East Africa to Southern India: A Study of 20 Autosomal DNA Markers. *Biochem. Genet.* Published online January 28, 2011. 10.1007/s10528-011-9419-7.
- International HapMap 3 Consortium, Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Bonnen, P.E., de Bakker, P.I.W., Deloukas, P., et al. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58.
- Nelson, M.R., Bryc, K., King, K.S., Indap, A., Boyko, A.R., Novembre, J., Briley, L.P., Maruyama, Y., Waterworth, D.M., Waeber, G., et al. (2008). The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am. J. Hum. Genet.* 83, 347–358.
- Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. (2009). Reconstructing Indian population history. *Nature* 461, 489–494.
- Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2, e190.
- Patterson, N., Petersen, D.C., van der Ross, R.E., Sudoyo, H., Glashoff, R.H., Marzuki, S., Reich, D., and Hayes, V.M. (2010). Genetic structure of a unique admixed population: Implications for medical research. *Hum. Mol. Genet.* 19, 411–419.
- Moorjani, P., Patterson, N., Hirschhorn, J.N., Keinan, A., Hao, L., Atzmon, G., Burns, E., Ostrer, H., Price, A.L., and Reich, D. (2011). The history of african gene flow into southern europeans, levantines, and jews. *PLoS Genet.* 7, e1001373.
- Rootsi, S., Zhivotovsky, L.A., Baldovic, M., Kayser, M., Kutuev, I.A., Khusainova, R., Bermisheva, M.A., Gubina, M., Fedorova, S.A., Ilumae, M.A., et al. (2008). A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur. J. Hum. Genet.* 15, 204–221.
- Scozzari, R., Torroni, A., Semino, O., Cruciani, F., Spedini, G., and Santachiara Benerecetti, S.A. (1994). Genetic studies in Cameroon: Mitochondrial DNA polymorphisms in Bamileke. *Hum. Biol.* 66, 1–12.
- Scozzari, R., Cruciani, F., Santolamazza, P., Sellitto, D., Cole, D.E., Rubin, L.A., Labuda, D., Marini, E., Sacca, V., Vona, G., and Torroni, A. (1997). mtDNA and Y chromosome-specific polymorphisms in modern Ojibwa: implications about the origin of their gene pool. *Am. J. Hum. Genet.* 60, 241–244.
- Scozzari, R., Cruciani, F., Pangrazio, A., Santolamazza, P., Vona, G., Moral, P., Latini, V., Varesi, L., Memmi, M.M., Romano, V., et al. (2001). Human Y-chromosome variation in the western Mediterranean area: Implications for the peopling of the region. *Hum. Immunol.* 62, 871–884.
- Cruciani, F., Santolamazza, P., Shen, P., Macaulay, V., Moral, P., Olckers, A., Modiano, D., Holmes, S., Destro-Bisoli, G., Coia, V., et al. (2002). A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am. J. Hum. Genet.* 70, 1197–1214.
- Luis, J.R., Rowold, D.J., Regueiro, M., Caeiro, B., Cinnioglu, C., Roseman, C., Underhill, P.A., Cavalli-Sforza, L.L., and Herrera, R.J. (2004). The Levant versus the Horn of Africa: Evidence for bidirectional corridors of human migrations. *Am. J. Hum. Genet.* 74, 532–544.
- Semino, O., Santachiara-Benerecetti, A.S., Falaschi, F., Cavalli-Sforza, L.L., and Underhill, P.A. (2002). Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am. J. Hum. Genet.* 70, 265–268.
- Semino, O., Magri, C., Benuzzi, G., Lin, A.A., Al-Zahery, N., Battaglia, V., Maccioni, L., Triantaphyllidis, C., Shen, P.,

- Oefner, P.J., et al. (2004). Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: Inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am. J. Hum. Genet.* 74, 1023–1034.
21. Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonn -Tamir, B., Bertranpetit, J., Francalacci, P., et al. (2000). Y chromosome sequence variation and the history of human populations. *Nat. Genet.* 26, 358–361.
  22. Plaza, S., Salas, A., Calafell, F., Corte-Real, F., Bertranpetit, J., Carracedo, A., and Comas, D. (2004). Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Hum. Genet.* 115, 439–447.
  23. Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Tolk, H.V., Stepanov, V., et al. (2003). The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* 72, 313–332.
  24. Thanseem, I., Thangaraj, K., Chaubey, G., Singh, V.K., Bhaskar, L.V.K.S., Reddy, B.M., Reddy, A.G., and Singh, L. (2006). Genetic affinities among the lower castes and tribal groups of India: Inference from Y chromosome and mitochondrial DNA. *BMC Genet.* 7, 42.
  25. Trivedi, R., Sahoo, S., Singh, A., Bindu, G., Banerjee, J., Tandon, M., Gaikwad, S., Rajkumar, R., Sitalaximi, T., and Ashma, R. (2008). Genetic imprints of Pleistocene origin of Indian populations: A comprehensive phylogeographic sketch of Indian Y-chromosomes. *Int. J. Hum. Genet.* 8, 97–118.
  26. Thangaraj, K., Naidu, B.P., Crivellaro, F., Tamang, R., Upadhyay, S., Sharma, V.K., Reddy, A.G., Walimbe, S.R., Chaubey, G., Kivisild, T., and Singh, L. (2010). The influence of natural barriers in shaping the genetic structure of Maharashtra populations. *PLoS ONE* 5, e15283.
  27. Brehm, A., Pereira, L., Bandelt, H.J., Prata, M.J., and Amorim, A. (2002). Mitochondrial portrait of the Cabo Verde archipelago: The Senegambian outpost of Atlantic slave trade. *Ann. Hum. Genet.* 66, 49–60.
  28. Berniell-Lee, G., Calafell, F., Bosch, E., Heyer, E., Sica, L., Mougouia-Daouda, P., van der Veen, L., Hombert, J.M., Quintana-Murci, L., and Comas, D. (2009). Genetic and demographic implications of the Bantu expansion: Insights from human paternal lineages. *Mol. Biol. Evol.* 26, 1581–1589.
  29. Tofanelli, S., Bertocini, S., Castr , L., Luiselli, D., Calafell, F., Donati, G., and Paoli, G. (2009). On the origins and admixture of Malagasy: New evidence from high-resolution analyses of paternal and maternal lineages. *Mol. Biol. Evol.* 26, 2109–2124.
  30. Zhivotovskiy, L.A., Underhill, P.A., Cinnioglu, C., Kayser, M., Morar, B., Kivisild, T., Scozzari, R., Cruciani, F., Destro-Bisol, G., Spedini, G., et al. (2004). The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* 74, 50–61.
  31. Beerli, P., and Felsenstein, J. (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc. Natl. Acad. Sci. USA* 98, 4563–4568.
  32. Beerli, P., and Felsenstein, J. (1999). Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152, 763–773.
  33. Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 23, 147.
  34. van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* 30, E386–E394.
  35. Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M.J., and Amorim, A. (2001). Prehistoric and historic traces in the mtDNA of Mozambique: Insights into the Bantu expansions and the slave trade. *Ann. Hum. Genet.* 65, 439–458.
  36. Salas, A., Richards, M., De la Fe, T., Lareu, M.V., Sobrino, B., S nchez-Diz, P., Macaulay, V., and Carracedo, A. (2002). The making of the African mtDNA landscape. *Am. J. Hum. Genet.* 71, 1082–1111.
  37. Salas, A., Richards, M., Lareu, M.V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V., and Carracedo, A. (2004). The African diaspora: Mitochondrial DNA and the Atlantic slave trade. *Am. J. Hum. Genet.* 74, 454–465.
  38. Beleza, S., Gusm o, L., Amorim, A., Carracedo, A., and Salas, A. (2005). The genetic legacy of western Bantu migrations. *Hum. Genet.* 117, 366–375.
  39. Wood, E.T., Stover, D.A., Ehret, C., Destro-Bisol, G., Spedini, G., McLeod, H., Louie, L., Bamshad, M., Strassmann, B.I., Soodyall, H., and Hammer, M.F. (2005). Contrasting patterns of Y chromosome and mtDNA variation in Africa: Evidence for sex-biased demographic processes. *Eur. J. Hum. Genet.* 13, 867–876.
  40. Gonder, M.K., Mortensen, H.M., Reed, F.A., de Sousa, A., and Tishkoff, S.A. (2007). Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol. Biol. Evol.* 24, 757–768.
  41. Crispim, D., Canani, L.H., Gross, J.L., Tschiedel, B., Souto, K.E.P., and Roisenberg, I. (2006). The European-specific mitochondrial cluster J/T could confer an increased risk of insulin-resistance and type 2 diabetes: an analysis of the m.4216T > C and m.4917A > G variants. *Ann. Hum. Genet.* 70, 488–495.
  42. Chaubey, G., Metspalu, M., Kivisild, T., and Villems, R. (2007). Peopling of South Asia: investigating the caste-tribe continuum in India. *Bioessays* 29, 91–100.
  43. Tishkoff, S.A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J., et al. (2001). Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* 293, 455–462.