Unraveling the Population History of Indian Siddis

Ranajit Das[1,*,†] and Priyanka Upadhyai[2,†]

[1]Manipal Centre for Natural Sciences (MCNS), Manipal University, Karnataka, India

[2]Department of Medical Genetics, Kasturba Medical College, Manipal University, Karnataka, India

[†]Both authors contributed equally to this work.

*Corresponding author: E-mail: ranajit.das@manipal.edu.

Accepted: May 10, 2017

Abstract

The Siddis are a unique Indian tribe of African, South Asian, and European ancestry. While previous investigations have traced their ancestral origins to the Bantu populations from subSaharan Africa, the geographic localization of their ancestry has remained elusive. Here, we performed biogeographical analysis to delineate the ancestral origin of the Siddis employing an admixture based algorithm, Geographical Population Structure (GPS). We evaluated the Siddi genomes in reference to five African populations from the 1000 Genomes project, two Bantu groups from the Human Genome Diversity Panel (HGDP) and five South Indian populations. The Geographic Population Structure analysis localized the ancestral Siddis to Botsawana and its present-day northeastern border with Zimbabwe, overlapping with one of the principal areas of secondary Bantu settlement in southeast Africa. Our results further indicated that while the Siddi genomes are significantly diverged from that of the Bantus, they manifested the highest genomic proximity to the North-East Bantus and the Luhyas from Kenya. Our findings resonate with evidences supporting secondary Bantu dispersal routes that progressed southward from the east African Bantu center, in the interlacustrine region and likely brought the ancestral Siddis to settlement sites in south and southeastern Africa from where they were disseminated to India, by the Portuguese. We evaluated our results in the light of existing historical, linguistic and genetic evidences, to glean an improved resolution into the reconstruction of the distinctive population history of the Siddis, and advance our knowledge of the demographic factors that likely contributed to the contemporary Siddi genomes.

Key words: Siddi, geo-location, GPS, admixture, Portuguese slave trade, Bantu expansion.

Introduction

India and its adjoining areas in South Asia are a panorama of astounding ethnic, linguistic, and genetic diversity interlaced with distinctive sociocultural practices. Contemporary India is a conglomeration of 4,635 anthropologically well-defined populations, including 532 tribes (Narang et al. 2010; Tamang et al. 2012). The Indian subcontinent has witnessed numerous waves of immigration and gene-flow from various parts of the world, in prehistoric and historic times that have been instrumental in shaping its population structure and demography (Barnabas et al. 1996; Kivisild et al. 1999; Metspalu et al. 2004; Misra 2001; Ratnagar 1995). The Siddis are a unique tribal group of African ancestry predominantly found in the Indian states of Gujarat, Karnataka, Andhra Pradesh, and Telengana (Lodhi 1992). The earliest evidences of their migration date back to ~1100 A.D and indicate that the Siddi immigrants settled on the western coast of India (Bhattacharya 1969; Gauniyal et al. 2008). By the 13th century a greater influx of Siddis occurred and they were imported by regional Indian kings and princes to serve as slaves and soldiers (Shah et al. 2011). Subsequently, during the 16th–19th centuries the Siddis were transported in large numbers to India as slaves by the Portuguese (Bhattacharya 1970; Nevet 1981). Several previous investigations have suggested that the Siddi genomes are predominantly closest to that of the Africans (Thangaraj et al. 1999). Further work not only confirmed the African ancestry of the Siddi people, but additionally revealed that they have assimilated considerable fractions of nonAfrican admixture components in their genomes (Gauniyal et al. 2008, 2011; Ramana et al. 2001). This is not unexpected given that the Siddis shared long periods of contact with both the South Indians (South Asians) and the Portuguese (Europeans). In congruence, subsequent studies delineated three ancestral components, African, South Asian, and European in the Siddi genomes (Gauniyal et al. 2008, 2011; Narang et al. 2011; Shah et al. 2011).

© The Author(s) 2017. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Further analyses traced the ancestry of Siddis to the Bantu speaking populations from subSaharan Africa (Narang et al. 2011; Shah et al. 2011). The Bantu expansion had swept out of west-central Africa, ~5000 years before present (YBP) and was one of the most defining cultural events that led to the spread of Bantu languages and genomes over the majority of the continent (Ehret and Posnansky 1982; Li et al. 2014; Newman 1995; Pereira et al. 2001). Given the complex and intricate demographic history of Africa, the genetically and culturally diverse Indian populace the precise geographic origin of the Siddi ancestry has remained elusive. In the current study we sought to ascertain the biogeographical affinity and map the ancestral origins of the Siddis. Further we aimed to expand upon the findings of previous studies (Narang et al. 2011; Shah et al. 2011) to build a cohesive and comprehensive understanding of the Siddi population history.

To facilitate our understanding of the Siddi population history, we performed a genome-wide analysis of the Siddis ($N = 14$), wherein samples had been procured from the Indian states of Gujarat ($N = 8$) and Karnataka ($N = 6$) (Moorjani et al. 2013) and assessed them in reference to seven African and five South Indian populations. We obtained five African populations from the 1000 Genomes project, namely Yorubans from Ibadan, Nigeria (YRI) ($N = 108$), Luhyas from Webuye, Kenya (LWK) ($N = 99$), Gambians from the Western Divisions in Gambia (GWD) ($N = 113$), Mendes from Sierra Leone (MSL) ($N = 85$), and Esans from Nigeria (ESN) ($N = 99$) (Genomes Project et al. 2015). And supplemented our analysis by evaluating the Siddis in reference to two Bantu populations, namely North-East Bantus from Kenya ($N = 12$) and Bantus from South Africa ($N = 8$) available in the Human Genome Diversity Panel (HGDP) (Li et al. 2008; Rosenberg et al. 2002). We also obtained 26 South Indian samples corresponding to five populations, namely Kattunayakkan ($N = 5$), Kuruchiyan ($N = 5$), Paniya ($N = 5$), Hallaki ($N = 7$), and Velama ($N = 4$) (Moorjani et al. 2013). Overall, we analyzed 564 samples evaluating a total of 90,872 single nucleotide polymorphisms (SNPs) that are common to all data sets under assessment. We mapped the likely geographical coordinates corresponding to the ancestral origins of the Siddi tribe by applying an admixture based method, Geographical Population Structure (GPS) (Elhaik et al. 2014) that has been successfully employed to trace the accurate biogeographical origin of various modern populations (Das et al. 2016; Marshall et al. 2016). This approach relies on extrapolating the genomic similarity between the query and reference populations to infer the potential biogeographical affiliation of the former using the geographic locations corresponding to the latter. Together with existing historical, linguistic and genetic evidences our findings provide a greater resolution into the genetic relatedness between populations, the fine-scale population structure and recapitulate the population history of the Siddi tribe.

Materials and Methods

Data Sets

Our analysis utilized publically available data sets from the 1000 Genomes project, phase 3 (Genomes Project et al. 2015), the Human Genome Diversity Panel (HGDP) data set 2 (Li et al. 2008; Rosenberg et al. 2002) and previously published data (Moorjani et al. 2013). File conversions and manipulations were performed using EIG v4.2 (Price et al. 2006), VCF tools (Danecek et al. 2011) and PLINK v1.07 (http://zzz.bwh.harvard.edu/plink/) (Purcell et al. 2007). All three data sets were made compatible with each other and merged together using PLINK.

Population Clustering and Admixture Analysis

Population stratification was estimated using the –cluster command in PLINK. Multidimensional scaling analysis was performed in PLINK using –mds command alongside the –read-genome flag. The multidimensional (mds) plot was generated using R v3.2.3 (https://www.r-project.org/) and it comprised of genome data pertaining to 564 individuals from Africa and India. The genetic ancestry of all individuals was estimated using an unsupervised clustering algorithm, ADMIXTURE (Alexander et al. 2009). The optimum number of ancestral components (K) was discerned by minimizing the cross-validation error (CVE) (Alexander et al. 2009) implemented in ADMIXTURE v1.3 using a –cv flag to the ADMIXTURE command line. For the data set including all 564 individuals, the lowest CVE was estimated for $K = 7$ (supplementary fig. S1, Supplementary Material online). Similarly, we also assessed the ancestry of 524 African and 14 Siddi individuals using ADMIXTURE and employed 73,629 Africa specific SNPs (MAF > 0.1) (Kent et al. 2011) to interrogate the African ancestry of the Siddis. Here, the lowest CVE was estimated for $K = 3$ (supplementary fig. S2, Supplementary Material online).

The ancient splits and genetic relatedness between the populations was inferred using TreeMix v1.13 (Pickrell and Pritchard 2012). Nine Onge genomes (Moorjani et al. 2013) were employed to root the maximum likelihood tree generated by TreeMix.

To further interrogate the genetic structure of the Siddi genomes we computed the three-population "f_3 statistic" (Reich et al. 2009) implemented in TreeMix v1.13. The test was in the form f_3 (Siddi; source 1, source 2) where a significant negative value of the f_3 statistic was an unequivocal signal of admixture. Various combinations of African and South Indian populations were employed as the source populations (supplementary table S1, Supplementary Material online). All feasible combinations of the f_3 statistic were computed in blocks of 500 SNPs using a –k 500 flag to the "threepop" command line.

Tracing the Biogeographical Origin of Siddis

Biogeographical analysis was performed using the Geographic Population Structure (GPS) algorithm as described previously (Das et al. 2016; Elhaik et al. 2014; Marshall et al. 2016). GPS determines the biogeographical affinity of a sample by correlating its admixture signature with that of present-day reference samples of known biogeographical affinity and subsequently converting the genetic distances into geographic distances. To infer the geographical coordinates (latitude and longitude) of an individual given *K* admixture proportions, GPS requires a reference population set of *N* populations with both *K* admixture proportions and two geographical coordinates (longitude and latitude). All admixture proportions were calculated as illustrated previously (Elhaik et al. 2014). GPS predictions should be considered as the last location where admixture had occurred or the *geographical origin*. For individuals of mixed ancestries, GPS coordinates represent the mean geographic locations of their immediate parental populations. In the present study we used the African populations from the 1000 Genomes project and HGDP as reference and interpreted their admixture fractions and geographic locations (latitudnal and longitudnal coordinates) to determine the biogeographical ancestry of the Siddi people.

Dating the Time of Admixture Events

The time of admixture events was estimated using ALDER v.1.02 (Loh et al. 2013) employing a generation time of 25 years. The input files ("EIGENSTRAT" format) for ALDER analysis were generated using EIG v. 4.2 (Price et al. 2006). The Siddi populations were used as the "admixpop" (admixed population) and African and South Indian populations were used as the "refpops" (reference populations) during the ALDER analysis.

Results

Clustering of Populations

The multidimensional scaling (MDS) plot revealed three distinct clusters; the Siddis formed an isolated cluster widely separated from the South Indian groups (Kattunayakkan, Kuruchiyan, Paniya, Hallaki, and Velama) and distinguished from the cluster of the African populations (fig. 1). While the Siddis emerged as significantly differentiated from most African populations, it is noteworthy that they appeared to be genomically proximal to the North-East Bantus and Luhyas (LWK) from Kenya.

Using unsupervised clustering, as implemented in ADMIXTURE (Alexander et al. 2009) we investigated the Siddi, African, and South Indian groups further and estimated the ancestry of each individual. At $K = 7$, ESN, LWK, and GWD groups were largely assigned homogeneously to unique populations (fig. 2). As may be surmised we found that

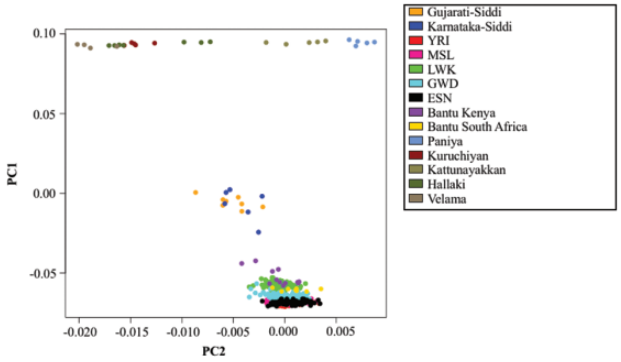

Fig. 1.—A multidimensional scaling plot of the Siddis, South Indian populations (Moorjani et al. 2013) and African groups from 1,000 Genomes project and HGDP. In this scatter plot each point represents an individual. Multidimensional scaling analysis was performed in PLINK and the plot was generated in R package. The orange and blue circles designate Siddi populations from Gujarat and Karnataka, respectively. The first eigenvector illustrates the separation of Siddis from South Indian populations, while the second eigenvector explains the variation among the African populations.

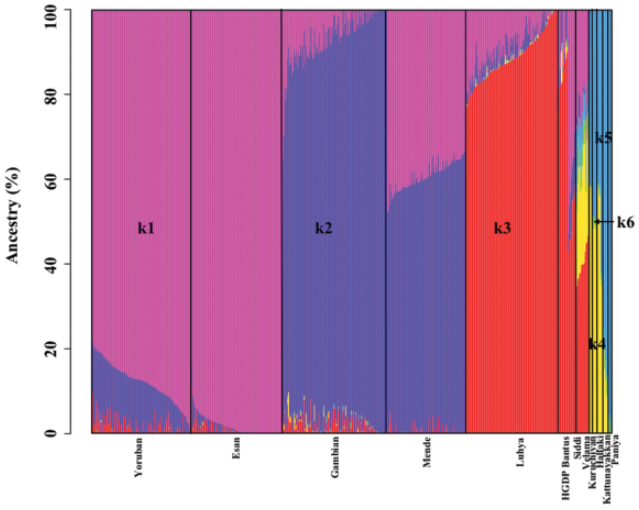

Fig. 2.—An admixture plot showing the ancestry components of the Indian Siddis along with the Indian and African populations ($N = 564$). Percent ancestry is plotted on the *Y*-axis. The ancestral components in the genomes under evaluation was estimated using ADMIXTURE v1.3. A model with seven ancestral components ($K = 7$) was determined as the most parsimonious to account for the variation and similarities of the genome-wide genotype data from 564 individuals. Each individual is represented by a vertical line partitioned into colored segments whose lengths are proportional to the genetic contributions of the ancestral components to the genome of the individual. Population labels were added after each individual's ancestry had been estimated. To note *k*1, *k*2, *k*3, *k*4, *k*5, and *k*6 represent putative Nigerian, Gambian, Bantu, and Indian ancestral components, respectively.

populations from the same or overlapping geographic location, such as the Nigerians (YRI and ESN) showed a very high degree of genomic similarity. Similar results were obtained for the Bantu groups, including Luhyas (LWK), North-East Bantus from Kenya, and Bantus from South Africa. The Siddis were discerned to possess the highest fraction of the Bantu (*k3*), followed by Indian (*k4*) and Nigerian (*k1*) ancestral components. In addition, they also appeared to contain minor fractions of *k5* and *k6* ancestral components likely inherited from the Indians.

An in-depth understanding of the African component of ancestry in the Siddis may be achieved by analyses of specific genomic regions in the latter that are exclusively of African origin. To this end we assessed the ancestry of African and Siddi individuals using 73,629 Africa specific SNPs (MAF > 0.1) (Kent et al. 2011). At $K = 3$, three distinct ancestral components became evident, namely Nigerian (*k1*), Gambian (*k2*), and Bantu (*k3*) (fig. 3). Consistent with previous results Siddis were found to be genomically closest to North-East Bantus and the Luhyas (LWK) from Kenya and Bantus from South Africa. Siddi genomes were ascertained to have the highest fraction of Bantu component (*k3*) with minor fractions of Nigerian (*k1*) and Gambian (*k2*) ancestral components.

We employed TreeMix v1.13 (Pickrell and Pritchard 2012) to investigate the pattern of population splits and mixtures amongst the Siddi, the African and the Indian populations. Our findings confirmed a high degree of genetic relatedness between the Siddis and other Bantu populations, in particular the Bantus from Kenya and the Luhyas (LWK) (fig. 4).

Given the evidences of African and Indian ancestry in the Siddi genomes we computed the statistics of the form f_3 (Siddis; $Source_1$, $Source_2$) and implemented it in TreeMix v1.13. A significantly negative f_3 statistic, indicates that the allele frequencies of the test population (Siddis) are likely an intermediate between the two source (ancestral) populations. This statistic was found to be significantly negative ($Z \leq -3$) for various combinations of an African and another South Indian population, consistent with signatures of significant admixture between these populations in the Siddi genomes (supplementary table S1, Supplementary Material online).

Tracing the Biogeographical Origin of Siddi Populations

A strong correspondence between genetic and geographic distances in worldwide populations has been established through several previous studies (Cavalli-Sforza and Menozzi 1994; Elhaik et al. 2014). Presently we sought to determine the geographic location corresponding to the ancestral origin of the Siddis by applying the GPS algorithm that has been used to trace the origins of several modern-day populations and is demonstrated to be more suitable for assessing biogeographical affinity in contrast to principal component analysis (PCA) (Das et al. 2016; Elhaik et al. 2014; Marshall et al. 2016).

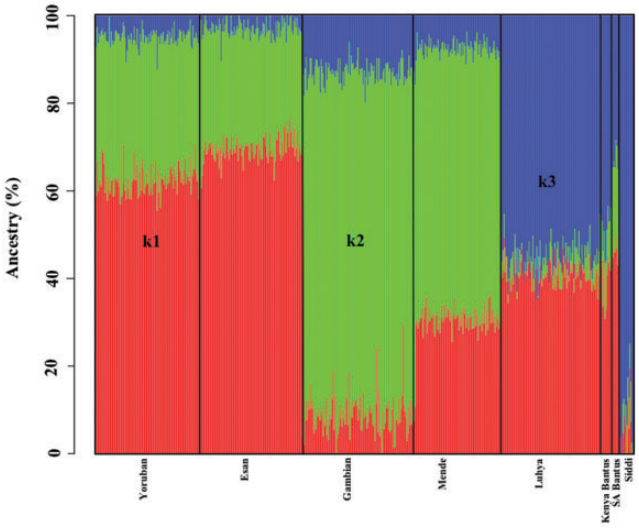

Fig. 3.—An admixture plot showing the ancestry components of the Indian Siddis with African populations ($N = 538$) evaluating only the African specific SNPs (MAF > 0.1). Percent ancestry is plotted on the *Y*-axis. The ancestral components in evaluated genomes were estimated using ADMIXTURE v1.3. A model with three ancestral components ($K = 3$) was the most parsimonious to explain the variation and similarities of the genome wide genotype data on the 538 individuals. Each individual is represented by a vertical line partitioned into colored segments whose lengths are proportional to the genetic contributions of the ancestral components to the genome of the individual. Population labels were added after each individual's ancestry had been estimated. To note *k1*, *k2*, and *k3* represent putative Nigerian, Gambian, and Bantu ancestral components, respectively.

Population expansion is accompanied by genetic exchanges with other groups modifying the admixture signature of the migrant group, while its isolation preserves its original admixture signals. Therefore, the GPS predictions correspond to the biogeographical affinity or the last geographic location where significant admixture had occurred in relation to the reference populations. In the current study, GPS assigned 12 out of 14 Siddi samples to Botswana while the remaining were positioned at its present day northeastern border with Zimbabwe (fig. 5). The postulated secondary expansion of the Bantus is regarded to have progressed southwards from the principal Bantu center in the Great Lakes region of east Africa, reaching most of southern and southeastern Africa, ~1000 YBP (Li et al. 2014; Pereira et al. 2001; Phillipson 1993). Consistent with this the ancestral origins of the Siddis, as predicted by GPS overlapped with that of likely Bantu settlements in southeastern Africa, from where ancestral African Siddi groups were purportedly transported to India during the 13th–19th centuries (Gauniyal et al. 2008; Nevet 1981; Shah et al. 2011).

Determination of the Time of Admixture Events

To quantitatively estimate the date for the admixture between ancestral Siddis and South Indian populations we employed

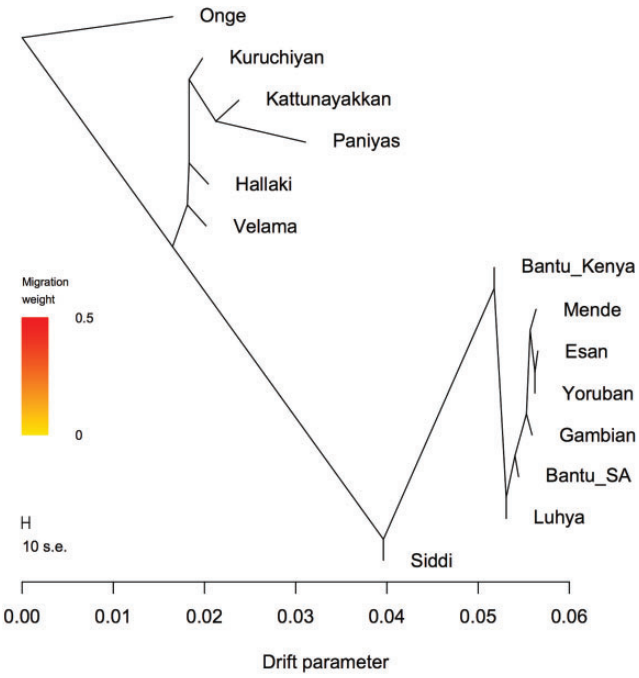

Fig. 4.—A Maximum Likelihood tree examining the genetic relatedness between Indian Siddis, the South Indians, and African populations. The population tree was constructed using TreeMix v1.13. The tree was rooted using Onge, a nonAfrican Andamanese population. The horizontal axis depicts the drift parameter. The scale bar shows ten times the average standard error of the entries in the sample covariance matrix. This confirmed the high genetic relatedness between Siddis and the Bantu populations, especially Bantus from Kenya and Luhyas (LWK).

ALDER v.1.02 (Loh et al. 2013) and traced back the admixture event to have occurred ~200 YBP or ~8 generations ago, assuming a generation time of 25 years (supplementary table S2, Supplementary Material online). Our estimation is concordant with that proposed previously using the less accurate ROLLOFF method (Shah et al. 2011) and is in close correspondence with ethnohistorical evidences of ancestral Bantu speaking Siddi groups being imported to India as slaves during the 16th and 17th centuries, by the Portuguese (Bhattacharya 1970; Nevet 1981).

Discussion

The Siddis are a unique tribal group settled in India whose ancestry is composed of African, South Asian, and European components (Bhattacharya 1970; Gauniyal et al. 2008, 2011; Narang et al. 2011; Ramana et al. 2001; Shah et al. 2011; Thangaraj et al. 1999). Several genetic studies have suggested that they are most closely related to Africans (Gauniyal et al. 2008, 2011) and have traced their ancestry to Bantu language speakers from subSaharan Africa (Narang et al. 2011; Shah et al. 2011). The Bantu populations refer to 300–600 African ethnic groups, who speak Bantu languages belonging to the Bantoid subgroup of Benue-Congo branch in the Niger-Congo language family and predominantly occupy western-central, eastern and southern Africa (Butt 2006; Nurse 2006). Bantu languages consist of three major groups, including northwestern, eastern and western Bantu languages (Guthrie 1948; Holden 2002; Vansina 1995). The Bantu expansion is regarded as a crucial demographic event in the history of subSaharan Africa and coincided with the spread of agriculture and iron metallurgy to southern and central regions of the continent (Diamond and Bellwood 2003; Newman 1995; Phillipson 1993). Linguistic, archaeological and genetic evidences are consistent in suggesting that proto-Bantu groups originated in the vicinity of the Cross River valley near the present-day border between Nigeria and Cameroon and their expansion occurred in multiple migratory waves, ~5000 YBP (Ehret and Posnansky 1982; Huffman 1982; Li et al. 2014; Pakendorf 2011; Pereira et al. 2001; Phillipson 1993; Salas et al. 2002). A southwest migration led to the dispersal of the proto-Bantu people to the Congo rainforest, while another concomitant southward migration likely progressed along the Atlantic coast, ~3500 YBP. The eastward wave of Bantu dispersal brought them to the interlacustrine region around the Great Lakes of east Africa, in present-day Uganda, Kenya and Tanzania, ~3000 YBP. A subsequent wave of Bantu dispersal initiated from the eastern Bantu center, ~1700 YBP southwards through Zimbabwe, Botswana. and reached Mozambique and South Africa (Li et al. 2014; Newman 1995; Phillipson 1993).

Several studies have suggested that multiple migratory routes of Bantu populations seemingly converged, overlapped, and split during distinct periods of time leading to the spread of Bantu genomes and languages to a great majority in subequatorial Africa (Ansari Pour et al. 2013; de Filippo et al. 2011, 2012; Montano et al. 2011; Plaza et al. 2004). Given the extensive and intricate nature of population expansions that spanned across Africa the precise geographic deduction of the ancestral origin of Siddi genomes had so far remained elusive (Narang et al. 2011). In the present study we sought to map the biogeographical origin of the Indian Siddi groups using the GPS algorithm. GPS localized the biogeographical affinity of the Siddi genomes to Botswana and its present day northeastern border with Zimbabwe (fig. 5), coinciding with some of the principal areas of secondary Bantu settlement in southern and southeastern Africa, from where ancestral African Siddi groups were purportedly transported to India during the 13th–19th centuries (Gauniyal et al. 2008; Nevet 1981; Shah et al. 2011). Further we extended and expanded upon previous analyses to achieve a deeper insight into the genetic relatedness between populations, utilizing a higher number of African and Bantu genomes for evaluation alongside the Siddis, in comparison to previous studies (Narang et al. 2011; Shah et al. 2011). Of the three Bantu populations included in our analyses, the Indian Siddis were revealed to share the greatest genomic proximity with the North-East Bantus and Luhyas from Webuye (LWK) from Kenya (figs. 1–3). This was further supported by the results

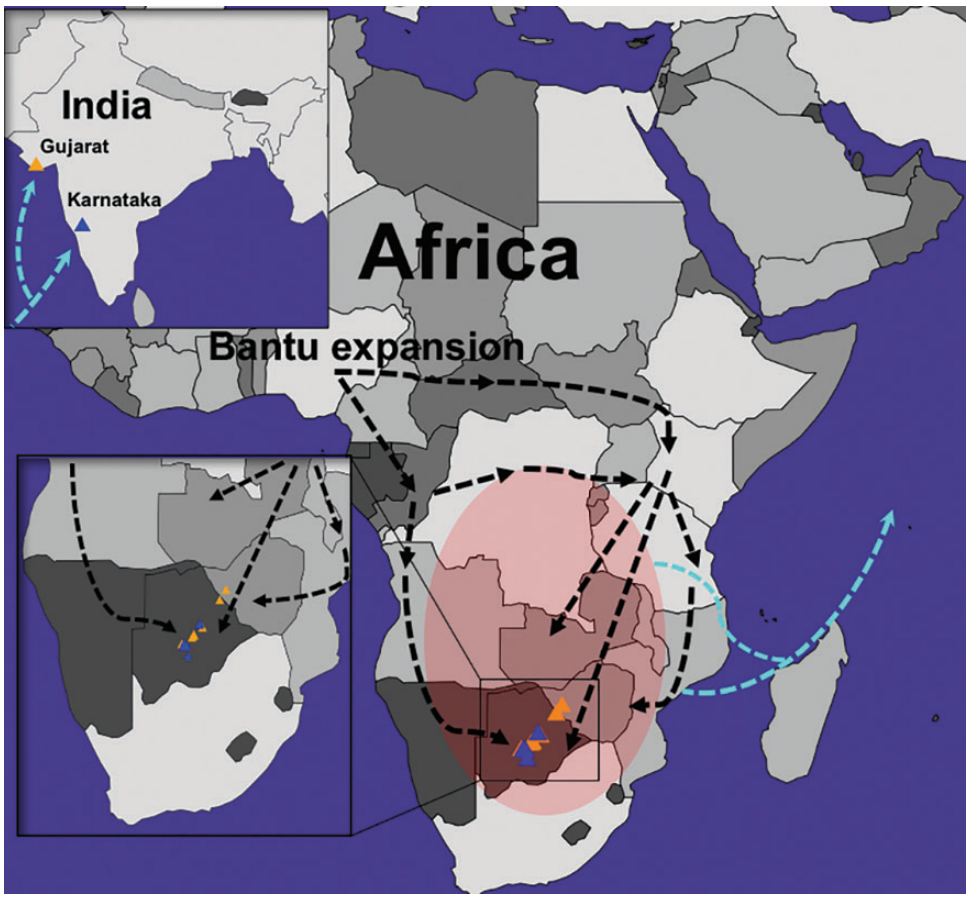

Fig. 5.—A map depicting the GPS predicted locations of the Indian Siddi genomes. The blue and orange triangles depict Siddi populations from the Indian states of Karnataka and Gujarat, respectively. Bantu expansion routes from west-central and eastern Bantu centers to south and southeast Africa are depicted by broken black lines (not to scale), adapted from Li et al. (2014). The red oval represents the region of secondary Bantu settlements in southern Africa that overlaps with the biogeographical origins of the Siddis (blue and orange triangles) and includes locations from where ancestral Siddi people were likely transported to India during the 13th–19th centuries (Gauniyal et al. 2008; Nevet 1981; Shah et al. 2011) (depicted by broken cyan lines).

of our population tree analysis employing the TreeMix algorithm (fig. 4). Notably while previous studies discerned a predominant African component in the Siddi ancestry, their genetic relatedness when compared to East (LWK) and West (YRI) African genomes had remained unresolved. While one investigation determined the Siddis to be close to both the YRI and Bantu groups (Narang et al. 2011) and similar findings emerged from another with regards to the whole genome data, however, based on Y haplogroup analysis Siddis were ascertained to be genetically closest to the Bantus (Shah et al. 2011). In contrast, our present findings unequivocally suggest the Siddis to be genetically closest, to the North-East Bantu and LWK groups of Kenya. The proximity of the African component of the Siddi ancestry to the two Bantu groups from Kenya is in concordance with secondary Bantu dispersal routes that putatively progressed southwards from the east African interlacustrine Bantu center, one wave of migration likely occurred along the shores of Lake Malawi via eastern Zimbabwe and Botswana reaching north Transvaal (northern South Africa), while the other advanced along the Ruvuma river that forms the border between Tanzania and Mozambique, reaching present-day Natal on the southeastern coast of Africa (Li et al. 2014; Pereira et al. 2001). It has also been suggested that southeastern and southwestern Bantu speaking groups mixed, on the basis of an overlapping occupation of present day Zambia (Huffman 1989). Together the southward Bantu expansions are likely to have brought the ancestral Siddi groups to settlement sites in southeast Africa from where they were disseminated to India, by the Portuguese (fig. 5).

Historically southeast Africa had emerged as an important source of slaves from the 16th century onwards, when individuals from Mozambique and Madagascar constituted a major proportion of the slaves being shipped by the Portuguese, primarily to former European colonies, such as the Caribbean and Brazil in the Americas (Thomas 1998). Several lines of evidences suggest a concordant time-line, during the 16th–19th centuries, when ancestral Siddi people were

transported by the Portuguese from Mozambique and its neighboring areas in southeast Africa, and eventually sold to regional Indian princes and chieftains, to serve as slaves and soldiers (Bhattacharya 1970; Nevet 1981). We examined the time-scale of the admixture event in the Siddi population history by applying the ALDER algorithm that employs an improved weighted linkage disequilibrium based method to investigate admixture (Loh et al. 2013). Our analysis dates the admixture between Siddis and South Asians to ~200 YBP or eight generations ago (supplementary table S2, Supplementary Material online), using a generation time of 25 years that is congruent with an earlier proposed timeline calculated using the less accurate ROLLOFF algorithm (Loh et al. 2013; Shah et al. 2011). The time-scale of admixture appears to be concordant with historical evidences suggesting that a large number of African ancestors of the Siddis were imported to India from Mozambique between 1680 and 1720 A.D. (Gauniyal et al. 2008; Nevet 1981). We also examined the African and Indian ancestry in Siddi genomes by computing the three-population f_3 statistic that further ratified the significant admixture between these ancestral populations in the Siddi genome (supplemetary table S1, Supplementary Material online).

Our analyses uncover the likely geographic locations corresponding to the ancestral origin of the Siddi genomes in Africa. It shines light on the genetic relatedness between populations, expanding upon the findings from previous investigations (Narang et al. 2011; Shah et al. 2011) to fine-tune and extend the prevailing understanding of the Siddis, a unique tribe of African origin from India. Taken together with existing historical, linguistic, and genetic evidences, our findings enable an improved resolution into the reconstruction of the distinctive population history of the Siddis, and advance our knowledge of the demographic factors that likely contributed to the contemporary Siddi genomes.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Literature Cited

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664.

Ansari Pour N, Plaster CA, Bradman N. 2013. Evidence from Y-chromosome analysis for a late exclusively eastern expansion of the Bantu-speaking people. Eur J Hum Genet. 21:423–429.

Barnabas S, Apte RV, Suresh CG. 1996. Ancestry and interrelationships of the Indians and their relationship with other world populations: a study based on mitochondrial DNA polymorphisms. Ann Hum Genet. 60:409–422.

Bhattacharya DK. 1969. Anthropometry of a Negroid population in India: Siddis of Gujarat. J Anthropol Soc Nippon. 77:254–256.

Bhattacharya D. 1970. Indians of African origin. Cah Etud Afr. 10:579–582.

Butt JJ. 2006. The greenwood dictionary of world history. Westport, USA: Greenwood Publishing Group.

Cavalli-Sforza LL, Menozzi PAP. 1994. The history and geography of human genes. Princeton: Princeton University Press.

Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Das R, Wexler P, Pirooznia M, Elhaik E. 2016. Localizing Ashkenazic jews to primeval villages in the ancient Iranian lands of Ashkenaz. Genome Biol Evol. 8:1132–1149.

de Filippo C, et al. 2011. Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups. Mol Biol Evol. 28:1255–1269.

de Filippo C, Bostoen K, Stoneking M, Pakendorf B. 2012. Bringing together linguistic and genetic evidence to test the Bantu expansion. Proc Biol Sci. 279:3256–3263.

Diamond J, Bellwood P. 2003. Farmers and their languages: the first expansions. Science 300:597–603.

Ehret C, Posnansky M. 1982. The archaeological and linguistic reconstruction of African history. Berkeley, CA: University of California Press.

Elhaik E, et al. 2014. Geographic population structure analysis of worldwide human populations infers their biogeographical origins. Nat Commun. 5:3513.

Gauniyal M, Aggarwal A, Kshatriya GK. 2011. Genomic structure of the immigrant Siddis of East Africa to southern India: a study of 20 autosomal DNA markers. Biochem Genet. 49:427–442.

Gauniyal M, Chahal SM, Kshatriya GK. 2008. Genetic affinities of the Siddis of South India: an emigrant population of East Africa. Hum Biol. 80:251–270.

Genomes Project C, et al. 2015. A global reference for human genetic variation. Nature 526:68–74.

Guthrie M. 1948. The classification of the Bantu languages. London, UK: Oxford University Press for the International African Institute.

Holden CJ. 2002. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. Proc Biol Sci. 269:793–799.

Huffman TN. 1982. Archaeology and the ethnohistory of the African Iron Age. Annu Rev Anthropol. 11:133–150.

Huffman TN. 1989. Iron age migrations: the ceramic sequence in southern Zambia: excavations at Gundu and Ndonde. Johannesburg, South Africa: Witwatersrand University Press.

Kent JW Jr, et al. 2011. Do rare variant genotypes predict common variant genotypes?. BMC Proc. 5(9 Suppl):S87.

Kivisild T, et al. 1999. Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. Curr Biol. 9:1331–1334.

Li JZ, et al. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319:1100–1104.

Li S, Schlebusch C, Jakobsson M. 2014. Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. Proc Biol Sci. 281:pii:20141448.

Lodhi A. 1992. African settlements in India. Nord J Afr Stud. 1:83–86.

Loh PR, et al. 2013. Inferring admixture histories of human populations using linkage disequilibrium. Genetics 193:1233–1254.

Marshall S, Das R, Pirooznia M, Elhaik E. 2016. Reconstructing Druze population history. Sci Rep.s 6:35837.

Metspalu M, et al. 2004. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet. 5:26.

Misra VN. 2001. Prehistoric human colonization of India. J Biosci. 26:491–531.

Montano V, et al. 2011. The Bantu expansion revisited: a new analysis of Y chromosome variation in Central Western Africa. Mol Ecol. 20:2693–2708.

Moorjani P, et al. 2013. Genetic evidence for recent population mixture in India. Am J Hum Genet. 93:422–438.
Narang A, et al. 2011. Recent admixture in an Indian population of African ancestry. Am J Hum Genet. 89:111–120.
Narang A, et al. 2010. IGVBrowser: a genomic variation resource from diverse Indian populations. Database (Oxford) 2010:baq022.
Nevet A. 1981. John the Britto. Bangalore, India: Loyal Mandira.
Newman JL. 1995. The peopling of Africa: a geographic interpretation. New Haven, Conn: Yale University Press.
Nurse D. 2006. Bantu languages. In: Bown K, editor. Encyclopedia of Language and Linguistics. Amsterdam: Elsevier.
Pakendorf B. 2011. Molecular perspectives on the Bantu expansion: a synthesis. Lang Dyn Change 1:50–88.
Pereira L, et al. 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.
Phillipson DW. 1993. African archaeology. Cambridge: Cambridge University Press.
Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8:e1002967.
Plaza S, et al. 2004. Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. Hum Genet. 115:439–447.
Price AL, et al. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 38:904–909.
Purcell S, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81:559–575.
Ramana GV, et al. 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.
Ratnagar S. 1995. Archaeological perspectives of early Indian societies. In: Thapar R, editor. Recent perspectives of early Indian history. Mumbai, India: Popular Prakashan. p. 1–52.
Reich D, Thangaraj K, Patterson N, Price AL, Singh L. 2009. Reconstructing Indian population history. Nature 461:489–494.
Rosenberg NA, et al. 2002. Genetic structure of human populations. Science 298:2381–2385.
Salas A, et al. 2002. The making of the African mtDNA landscape. Am J Hum Genet. 71:1082–1111.
Shah AM, et al. 2011. Indian Siddis: African descendants with Indian admixture. Am J Hum Genet. 89:154–161.
Tamang R, Singh L, Thangaraj K. 2012. Complex genetic origin of Indian populations and its implications. J Biosci. 37:911–919.
Thangaraj K, Ramana GV, Singh L. 1999. Y-chromosome and mitochondrial DNA polymorphisms in Indian populations. Electrophoresis 20:1743–1747.
Thomas H. 1998. The slave trade: the history of the Atlantic slave trade. London: Macmillan Publishers Ltd.
Vansina J. 1995. New linguistic evidence and the Bantu expansion. J Afr Hist. 36:173–195.

Associate editor: Partha Majumder