ORIGINAL ARTICLE

Rajni Trivedi · T. Sitalaximi · Jheelam Banerjee
Anamika Singh · P.K. Sircar · V.K. Kashyap

Molecular insights into the origins of the Shompen, a declining population of the Nicobar archipelago

Abstract The Shompen, one of the most isolated and poorly understood contemporary hunter–gatherer populations, inhabit Great Nicobar Island, the southernmost island of the Nicobar archipelago. Morphological imprints in the Shompen were interpreted to favour a mixed Indo-Chinese, Malay, Negrito and Dravidian origin. Analyses of the mitochondrial, Y-chromosomal and autosomal gene pool of contemporary Shompen have revealed low diversity, illustrating a founder effect in the island population. Mitochondrial sequence analyses revealed the presence of two haplogroups of R lineage: B5a, and a newly defined clade, R12. Y-chromosomal analyses demonstrated the occurrence of a single lineage found predominantly in Austro-Asiatic speakers across Asia. With the different types of genetic markers analysed, the Shompen exhibit varying levels of genetic relatedness with the Nicobarese, and Austro-Asiatic speakers of mainland India and Southeast Asia. These genetic analyses provide evidence that the Shompen, an offshoot of the Nicobarese, are descendants of Mesolithic hunter–gatherers of Southeast Asian origin, deriving from at least two source populations.

Keywords Nicobar · Shompen · Austro-Asiatic · Mitochondrial · Y-chromosomal · Autosomal

R. Trivedi · T. Sitalaximi · J. Banerjee · A. Singh
V.K. Kashyap
National DNA Analysis Centre,
Central Forensic Science Laboratory,
30 Gorachand Road, Kolkata,
700014, West Bengal, India

P.K. Sircar
Andaman Adim Janjati Vikas Samiti,
Government of India, Port Blair, India

Present address: V.K. Kashyap (✉)
National Institute of Biologicals, A-32,
Sector 62, Institutional Area, Noida,
201307 Uttar Pradesh, India
E-mail: nibindia@vsnl.net
Tel.: +91-120-2403014
Fax: +91-120-2400027

Introduction

The Shompen, an isolated aboriginal population (population size: ~200 individuals), is confined to the southernmost Great Nicobar Island of the Andaman and Nicobar archipelago located in the Bay of Bengal. Following a hunting–gathering way of life, the Shompen have remained in seclusion, in contrast to the Nicobarese, the other indigenous population that is dispersed over all 12 inhabited islands of the Nicobar archipelago. Although the Shompen and Nicobarese are ethnically described as "Mongoloids", their evolutionary histories have been debated due to their evident differences in physical appearance (Man 1886; Temple 1901). The Shompen have a stark phenotypic resemblance to African populations with a trace of Mongoloid traits. Juxtaposed to this, the languages spoken by these two groups are distinct and unintelligible even to each other. These facts have stimulated two primary lines of inquiry for population biologists. First, to trace the temporal and geographic origins of the Shompen. Second, with respect to intertribal differentiation, to determine whether the Shompen and Nicobarese share a common origin and have differentiated due to long-term geographic segregation, or whether the differences result from their separate origin from diverse populations that arrived in successive waves to inhabit the Andaman and Nicobar Islands. There are several theories regarding the origin of the Shompen, alluding to their descent from Malay, Burmese and Chinese (Rizvi 1990).

Languages are known to carry imprints of ancestries and migrational histories of populations (Cavalli-Sforza et al. 1988). The language isolates of the Shompen and Nicobarese have both been grouped into the Mon-Khmer branch of the Austro-Asiatic linguistic family (Khasi is the third Indian language that derives from the Mon-Khmer branch) (Ruhlen 1991). The Shompen speak 'Shompen', which has 25 consonants and 35 vowels. The language is polysyllabic, having words of foreign origin from Portuguese and Malay. Linguistic

Table 1 Mitochondrial DNA (mtDNA) and Y-chromosomal diversity in Shompen. *n* Number of individuals, *SD* standard deviation, *MPD* mean pairwise difference

Population	*n* mtDNA/Y	Number of mtDNA/Y-haplogroups	Number of mtDNA/Y-haplotypes	mtDNA sequence diversity±SD/Y-haplotype diversity±SD	mtDNA nucleotide diversity±SD	mtDNA MPD±SD/Y-MPD±SD
Shompen	29/12	2/1	3/10	0.512±0.063/0.879±0.075	0.012±0.007	4.280±2.183/1.333±0.889

and genetic evolution do not necessarily go hand in hand, although the two can be correlated in certain instances (Cavalli-Sforza et al. 1988; Sajantila and Pääbo 1995), leaving open the possibility that the two Austro-Asiatic-speaking populations, the Nicobarese and the Shompen, share common genetic origins.

Recent molecular studies on the Nicobarese population (Endicott et al. 2003; Thangaraj et al. 2003) have revealed their Southeast Asian origin (Prasad et al. 2001); however, the Shompen remain poorly understood due to inaccessibility due to their isolation. Earlier studies on the Shompen were confined to conventional genetic markers (Agarwal 1966), which did not help in unravelling the position of the Shompen with respect to modern human diversity. To address the issues of origin and antiquity of the Shompen, we carried out a comprehensive analyses of mitochondrial control and coding region polymorphisms; 20 Y-chromosomal short tandem repeats (Y-STRs), 35 lineage defining Y-chromosomal single nucleotide polymorphisms (Y-SNPs) and 15 autosomal short tandem repeats (STRs) in the Shompen and compared the results with available data on related populations of India, East Asia, Southeast Asia and Oceania.

Materials and methods

DNA sample

Blood samples were collected from 33 unrelated Shompen individuals after ethical clearance from appropriate bodies. Morphometric indices and disease status were recorded at the time of sample collection. DNA was isolated by the organic extraction method (Sambrook et al. 1989).

Analysis of autosomal STRs

Fifteen STRs were typed for 33 unrelated individuals using an AmpF/STR Identifiler kit (Applied Biosystems, Foster City, CA). The amplified products were electrophoresed in a 6% denaturing polyacrylamide gel on ABI Prism 377 DNA Sequencer (PE Applied Biosystems, Foster City, CA). Sizing of DNA fragments was performed using GeneScan analysis software (version 3.7), and allele designation was assigned using Genotyper DNA fragment analysis (version 3.7) software (Applied Biosystems, Foster City, CA).

Mitochondrial analysis

Of the 33 DNA samples, 29 were first amplified for selected regions of the mitochondrial DNA (mtDNA). They were assayed for the presence of the intergenic 9 bp deletion along with mtDNA restriction fragment length polymorphisms using standard primers and protocols

(Torroni et al. 1993, 1996). Sequencing of mitochondrial hypervariable segment I (HVS-I) (16024–16383) and hypervariable segment II (HVS-II) (57–372) (GenBank accession numbers DQ094084–DQ094112 for HVS-I sequences and DQ094113–DQ094141 for HVS-II sequences; http://www.ncbi.nlm.nih.gov/Genbank/) was carried out according to the technical booklet on mtDNA Sequencing (PE Applied Biosystems). Sequencing for informative coding region mutations was also carried out in few selected samples from B5a and R9b lineages using primers described by Torroni et al. (2001).

Analysis of Y-chromosomal polymorphisms

Binary polymorphism typing

Thirty-five SNPs distributed throughout the Y-chromosome were typed hierarchically in 12 Shompen males. The M9 (C–G) mutation was detected by *Hinf*I restriction digestion of a 164 bp PCR product obtained using standard primers and protocols (Underhill et al. 1997). Sequencing of the other polymorphisms was carried out using standard primer sets (Underhill et al. 1997).

Analysis of Y-STRs

Twenty Y-chromosomal STRs comprising three trinucleotide repeats (DYS426, DYS392 & DYS388), 12 tetranucleotide repeats (DYS389I, DYS389II, DYS439, DYS437, DYS391, DYS385a, DYS385b, DYS390, DYS393, DYS19, DYSH4, DYS460), two pentanucleotide repeat polymorphisms (DYS447, DYS438), two dinucleotide markers (YCAa, YCAb) and one hexanucleotide marker, DYS448, were analysed by an in-house-developed novel 20plex PCR reaction, performed for all the microsatellite markers in a final reaction volume of 10 μl containing 10 ng genomic DNA. The chosen Y-STRs were amplified using primers from the Y-STR database (http://www.ystr.charite.de) and the reaction conditions were standardised in our laboratory (S. Sahoo and V.K. Kashyap, unpublished). Amplified products were run on ABI Prism 3100 Genetic Analyzer using LIZ-500 (PE Applied Biosystems, Foster City, CA) as the internal lane size standard.

Statistical analysis

Allele frequencies, and expected and observed heterozygosities (Guo et al. 1992) for autosomal STRs were calculated using DNATYPE software (Windows 95/NT version 1998, University of Texas, Houston, TX). D_A distances (Nei et al. 1983) were employed for constructing the neighbour-joining phylogeny (Saitou and Nei 1987). $\delta\mu^2$ distances (Goldstein et al. 1995) were computed for estimating divergence time, T, using the formula, $T = \delta\mu^2/2\omega$, where ω is the effective mutation rate.

Mitochondrial sequences were edited between positions 57–372 for HVS-II and positions 16024–16383 for HVS-I and sequence traces of Y-chromosome were compared with respective control consensus sequences using BIOEDIT Software (North Carolina State University, NC). Diversity indices and mean number of pairwise differences between sequences/haplotypes for mitochondrial and Y-haplotypes were calculated using Arlequin software (Schneider et al. 2000). Median-joining networks for mitochondrial and Y-STR haplotypes were constructed by NETWORK v 4.1 software (Bandelt et al. 1999). ρ statistics (Forster et al. 1996) was applied to calculate coalescence estimates taking into account the mutation rate of one mutation every

Fig. 1 Median-joining network depicting the identified mtDNA lineages in Shompen. Substitutions are shown relative to the CRS. Node areas are proportional to haplotype frequencies. Character change is specified only for transversions
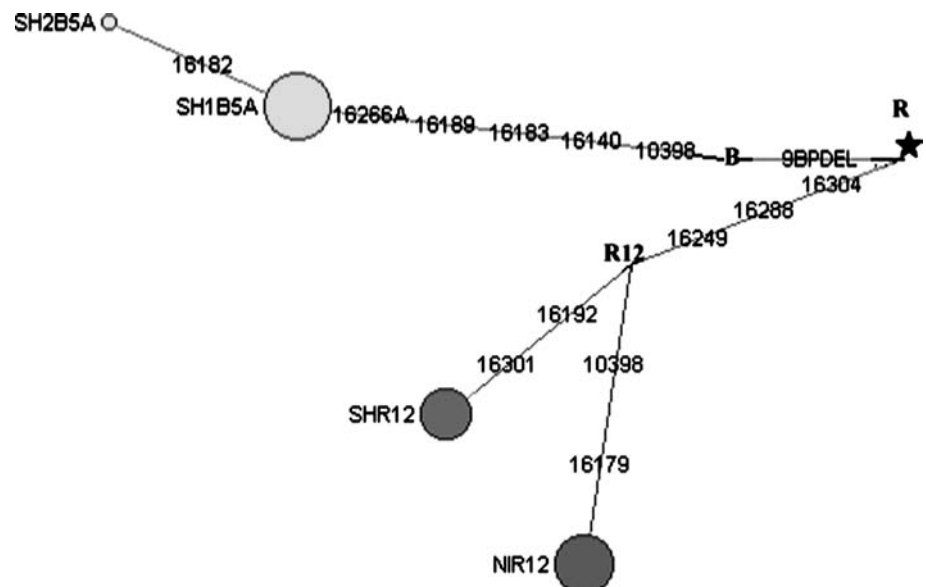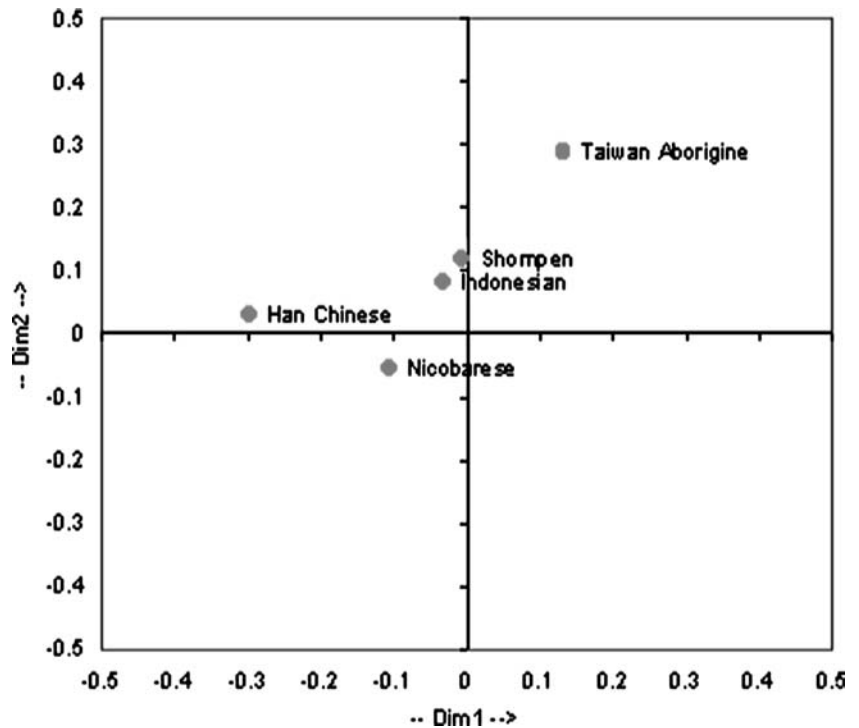
Fig. 2 Multidimensional scaling (MDS) plot of F_{ST} distances between similar mitochondrial haplotypes present among Shompen, Nicobarese and East/Southeast Asian populations

20,180 years. The mutational age was then converted into years by multiplication with the mutation rate. The respective standard errors were computed as described by Saillard et al. (2000). F_{ST} distances for mitochondrial haplotypes and Y-STRs were calculated using Arlequin software and corresponding MDS plots constructed with SPSS v7.1 software. Y-STR variance, $v = T\mu\eta_2$ (Kittles et al. 1998), was calculated to estimate the coalescence age using the mutation rate described by Zhivotovsky et al. (2004).

## Results

Mitochondrial DNA diversity

mtDNA polymorphisms were analysed in 29 individuals. Three haplotypes were identified, with a total of 14 polymorphic sites. The haplotype diversity was found to be low, at 0.51 (Table 1). Analyses of coding region polymorphisms established the absence of haplogroup M, the predominant lineage found in Indian and in other Asian populations. On further analyses, 19 individuals were found to harbour the intergenic COII/tRNA^{Lys} 9 bp deletion, which, along with the control region polymorphisms, placed them under the B5a lineage. A new R clade, which we propose to designate as R12, defined by transitions at nucleotide positions 16249, 16288 and 16304, was identified in ten Shompen individuals; the ancestral sequence of this clade occurs among the Nicobarese (Prasad et al. 2001). The major haplogroups prevalent among Shompen were similar to those identified earlier among Nicobarese (authors' unpublished data). In particular, the mutations in

HVS-I between Shompen and Nicobarese B5a lineages were identical with differences present only in HVS-II. The R12 lineage, however, was found to be more divergent between Nicobarese and Shompen, with three mutational differences in HVS-I alone (Fig. 1). Although mutation at position 16304 indicates the sequences belong to the R9b lineage, the absence of R9 defining mutations in the coding region at positions 3970 and 13928 in both populations implied a new clade, designated as R12 in the present study. The occurrence of only two lineages, coupled with their low diversity in the Shompen, in addition to the high diversity of these two lineages in the Nicobarese, indicates a founder effect in the Shompen. An MDS plot (Fig. 2) constructed using other related B5a lineages present among Nicobarese (authors' unpublished data), Han Chinese (Yao et al. 2002), Indonesians (Redd et al. 1995), Taiwanese aborigines (Melton et al. 1998) and R12 lineages prevalent among the Nicobarese (authors' unpublished data) and Indonesians (Lum et al. 1994) reveals that the Shompen are closer to Indonesians than to Nicobarese.

Y-chromosomal diversity

A single major Y-chromosomal haplogroup, O2a, defined by M95 marker was identified in 12 Shompen males. The O2a lineage was also found in Nicobarese individuals. This lineage is also reported among Central Asians, as well as East and Southeast Asians (Su et al. 1999; Kayser et al. 2000, 2001, 2003; Su et al. 2000; Underhill et al. 2000; Wells et al. 2001). Given the high frequency of occurrence of the O2a lineage among the

Fig. 3 MDS plot of R_{ST} distances of Y-chromosomal haplotype frequencies for Shompen and related populations
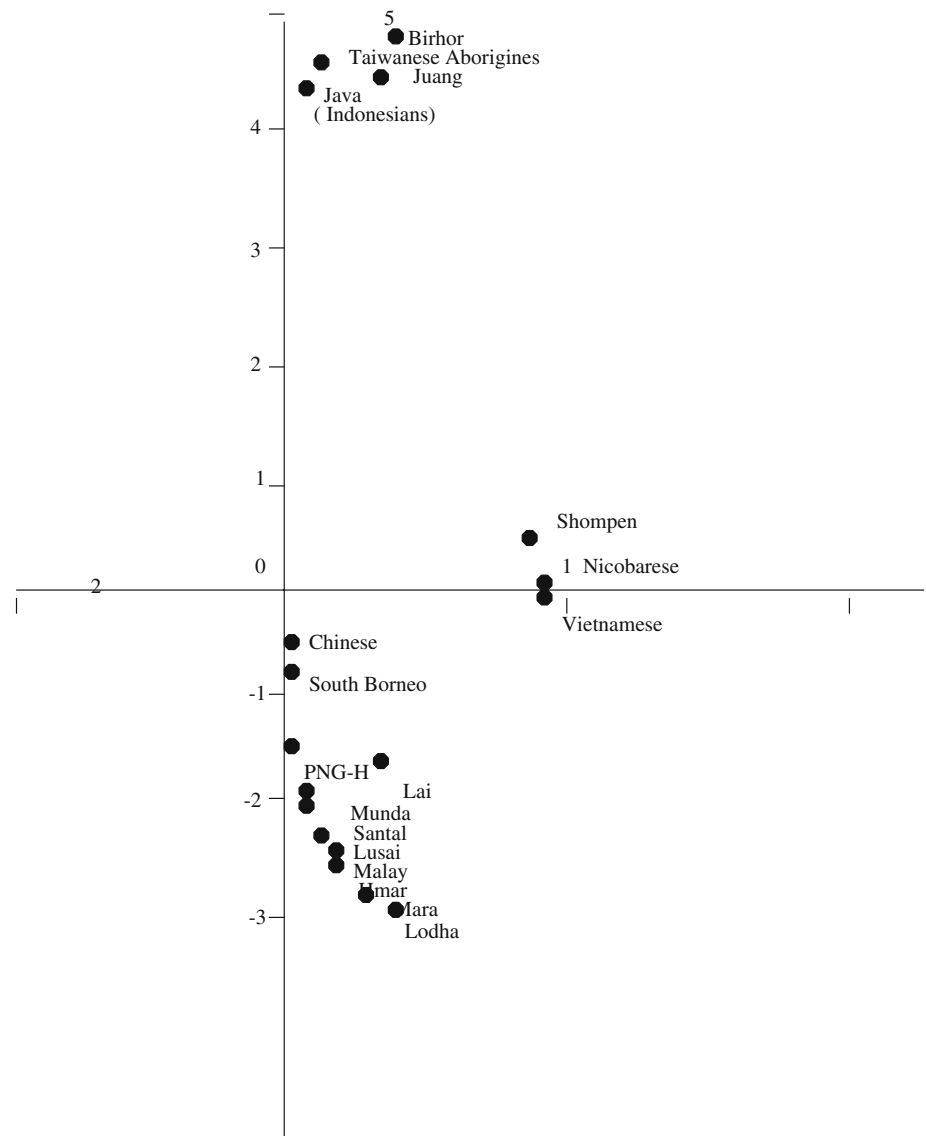
Austro-Asiatic linguistic gene pool, it has been suggested that this haplogroup might be related to the spread of Austro-Asiatic speakers (Kivisild et al. 2003) in the Asian continent.

In order to further investigate the correlation of the O2a lineage with Austro-Asiatic language speakers, 20 Y-STRs were analysed in the Shompen; 11 loci were found to be monomorphic. Of the ten Y-STR haplotypes identified in Shompen, eight were unique. The haplotype diversity and mean pairwise differences were low, at 0.879 and 1.333, respectively (Table 1). Haplotype data for 20 Y-STRs are presented in Table 2. A multidimensional scaling (MDS) plot of R_{ST} distances of Y-chromosomal haplotype frequencies for Shompen and related populations (Kayser et al. 2001; 2003, authors' unpublished data) (Fig. 3) demonstrate the affinity of the Shompen to Austro-Asiatic speakers, and to Nicobarese and Vietnamese, rather than to mainland

Indian Austro-Asiatic populations that share comparable O2a lineage frequencies. The reduced median-joining network based on Y-STR haplotypes (Fig. 4), produced distinct groups of mainland Indian Austro-Asiatic speakers (with two individuals sharing a Shompen haplotype), Shompen, Nicobarese and East and Southeast Asian populations.

Autosomal microsatellite diversity

Distribution of alleles across the 15 microsatellite loci in 33 Shompen individuals revealed low polymorphism with an average heterozygosity of 61.9% (Table 3). Phylogenetic analysis employing common shared microsatellites in related populations (Linacre et al. 2001; Bagdonavicius et al. 2002; Alves et al. 2004) demonstrated the Shompen clustering away from mainland

Table 2 Haplotype data for 20 Y-chromosomal microsatellites in the Shompen

Haplogroup	Haplotype	DYS438	DYS439	DYS389II	DYS437	DYS389I	DYS391	DYS385a	DYS385b	DYS390	YCAIIa	YCAIIb	DYS393	DYS426	DYS392	DYS19	DYS388	H4	DYS460	DYS448	DYS447	Frequency
O2a	H1	10	16	18	13	12	7	14	17	25	15	17	14	10	12	15	16	10	9	21	27	1
O2a	H2	10	17	18	13	12	7	15	17	25	15	17	14	10	12	15	16	10	10	20	27	1
O2a	H3	10	17	18	13	12	7	15	17	25	15	17	14	10	12	15	16	10	11	20	27	1
O2a	H4	10	17	18	13	12	7	15	17	25	15	17	14	10	13	15	12	10	11	20	27	1
O2a	H5	10	17	18	13	12	9	15	17	25	15	17	14	10	12	15	16	10	11	20	27	2
O2a	H6	10	17	18	13	12	9	15	17	25	15	17	14	10	12	15	16	10	10	20	27	1
O2a	H7	10	17	18	13	12	9	15	17	25	15	17	14	10	13	15	12	10	11	20	27	1
O2a	H8	10	17	18	13	12	9	15	17	25	15	17	14	10	12	15	12	10	10	20	27	2
O2a	H9	10	17	18	13	12	10	15	17	25	15	17	14	10	12	15	12	10	10	20	27	1
O2a	H10	10	17	19	13	13	9	15	17	25	15	17	14	10	12	15	12	10	10	20	27	1

Indian populations with greater proximity to Southeast Asian populations (Fig. 5).

Discussion

Although both the Shompen and the Nicobarese probably derive from the same founding population, considerable genetic differentiation has contributed to their current recognition as two distinct populations, as evident from the extensive analyses performed on mitochondrial, Y-chromosomal and autosomal markers. Occupying the most remote island of the Nicobar archipelago, the Shompen have existed in physical isolation for a considerable period of time. This feature is evident from the existence of marked linguistic differences between the Shompen and Nicobarese languages. On the other hand, the Nicobarese, spread across 12 inhabited islands of the Nicobar archipelago as well as in Little Andaman have adapted well to existing conditions and constitute an expanding population, maintaining contacts with settlers and traders from mainland India and other Asian countries.
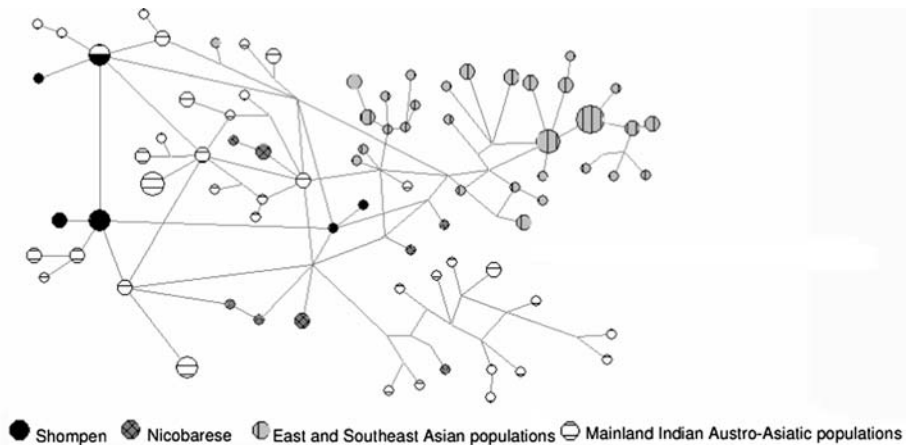
Mitochondrial sequence imprints

MtDNA analysis of the Shompen reveals the occurrence of only two haplogroups, B5a and R12, which are also the predominant lineages amongst Nicobarese, suggesting common ancestry of the two populations. The presence of both lineages in the two populations suggests either that the founding population of the Nicobar archipelago harboured these two different lineages or that they were brought in by two distinct founders. Mutational differences in the haplotypes of the Shompen and Nicobarese could have arisen following subsequent genetic differentiation of these two groups. The HVS-I sequences of the B5a haplotype were identical in both Shompen and Nicobarese, differing only at two additional HVS-II mutations (positions 309.1 and 309.2) in the Nicobarese. An identical match to the Shompen haplotype profile was found in a Han Chinese sample (Yao et al. 2002), while one with a single-mutational step difference was reported in an Indonesian sample (Redd et al. 1995) and another with a two-mutational step difference was found in a Taiwanese aborigine (Melton et al. 1998). The B5a haplogroup is reported predominantly in populations of China, Thailand and Taiwan. The R12 lineage identified by the HVS-I sequence motif 16249, 16288 and 16304 further differs by additional mutations in both Shompen and Nicobarese (Fig. 1). This lineage is also more diverse in the Nicobarese (Prasad et al. 2001). It is probable that populations harbouring a high frequency of this lineage are yet to be sampled, especially across Southeast Asia. The estimated coalescence time for the R12 lineage is $21,000 \pm 13,000$ years. The MDS plot (Fig. 2) reveals that the Shompen maternal lineage is genetically closer to that of Indonesians than that of other compared populations.

Table 3 Distribution of allele frequency at 15 autosomal short tandem repeats (STRs) (indicated in *bold*) in the Shompen population. *Obs Het* Observed heterozygosity

Allele	D5S818	FGA	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	THO1	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51
6	—	—	—	—	—	—	—	0.125	—	—	—	—	—	—	—
7	0.018	—	—	—	0.018	—	—	0.018	—	—	—	—	—	—	—
8	—	—	—	—	**0.643**	—	—	—	**0.732**	—	—	—	—	0.393	—
9	0.018	—	—	—	0.054	0.214	—	**0.393**	0.179	**0.393**	—	—	—	0.161	—
9.3	—	—	—	—	—	—	—	0.089	—	—	—	—	—	—	—
10	**0.482**	—	0.036	—	0.036	—	—	0.375	0.071	—	—	—	—	—	—
10.2	—	—	—	—	—	0.036	—	—	—	—	—	—	—	—	—
11	0.125	—	—	—	0.018	0.214	—	—	—	0.179	—	—	—	**0.446**	—
12	0.107	—	**0.536**	—	0.053	**0.536**	—	—	—	0.179	—	—	—	—	0.500
13	0.214	—	—	—	0.179	—	—	—	0.018	0.250	—	0.036	—	—	0.036
13.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.018
14	0.036	—	0.393	—	—	—	—	—	—	—	—	0.161	**0.643**	—	0.071
15	—	—	0.036	—	—	—	0.393	—	—	—	—	—	—	—	**0.339**
15.2	—	—	—	—	—	—	—	—	—	—	—	**0.696**	—	—	—
16	—	—	—	—	—	—	0.196	—	—	—	—	—	0.036	—	0.036
16.2	—	—	—	—	—	—	—	—	—	—	—	0.089	—	—	—
17	—	—	—	—	—	—	**0.411**	—	—	—	0.018	—	0.125	—	—
18	—	—	—	—	—	—	—	—	—	—	0.018	—	—	—	—
18.2	—	—	—	—	—	—	—	—	—	—	—	0.018	—	—	—
19	—	—	—	—	—	—	—	—	—	—	**0.643**	—	0.196	—	—
20	—	0.018	—	—	—	—	—	—	—	—	0.179	—	—	—	—
20.2	—	0.179	—	—	—	—	—	—	—	—	—	—	—	—	—
21	—	**0.321**	—	—	—	—	—	—	—	—	0.054	—	—	—	—
21.2	—	0.125	—	—	—	—	—	—	—	—	—	—	—	—	—
22	—	0.196	—	—	—	—	—	—	—	—	0.089	—	—	—	—
24	—	0.143	—	—	—	—	—	—	—	—	—	—	—	—	—
25	—	0.018	—	—	—	—	—	—	—	—	—	—	—	—	—
29	—	—	—	0.125	—	—	—	—	—	—	—	—	—	—	—
29.2	—	—	—	0.054	—	—	—	—	—	—	—	—	—	—	—
30	—	—	—	0.500	—	—	—	—	—	—	—	—	—	—	—
30.2	—	—	—	0.071	—	—	—	—	—	—	—	—	—	—	—
31	—	—	—	**0.214**	—	—	—	—	—	—	—	—	—	—	—
31.2	—	—	—	0.018	—	—	—	—	—	—	—	—	—	—	—
32	—	—	—	0.018	—	—	—	—	—	—	—	—	—	—	—
Obs Het	0.607	0.857	0.679	0.536	0.464	0.714	0.714	0.750	0.393	0.571	0.607	0.536	0.643	0.571	0.643

Fig. 4 Reduced median-joining network of haplogroup O2a individuals, based on their Y-short tandem repeat (STR) haplotypes. Different populations of origin are differentially shaded as indicated

● Shompen ⊗ Nicobarese ◑ East and Southeast Asian populations ○ Mainland Indian Austro-Asiatic populations

Y-chromosomal signature

The Y-chromosomal haplogroup O2a was the only lineage present among the Shompen. The frequency of the O2a lineage is reported highest among the Mon-Khmer-speaking Malayan aborigines (Orang-Asli; Su et al. 2000), the Bulang of southern China (Su et al.1999), Vietnamese (Kayser et al. 2003), and among Austronesian-speaking Taiwanese aborigines (Yami; Su et al. 2000) and Indonesians (Kayser et al. 2003). MDS analysis with Y-STRs (Fig. 3) has revealed that the Shompen are closer to Nicobarese and Vietnamese than to mainland Indian Austro-Asiatic populations. Additional data from other Asian populations would be required to trace the spread of this lineage in Asia. The Shompen and Nicobarese clusters were found to be distinct from the mainland Indian and East Asian/ Southeast Asian clusters (Fig. 4), illustrating the lack of geneflow in contemporary populations following their early splits. The single haplotype shared between the Shompen and mainland Indians could

be due to low resolution resulting from the low number of Y-STR markers available for comparison. The age of the Shompen O2a cluster as estimated from Y-STR variance was ∼3,000 years.

Synthesis of mitochondrial, Y-chromosomal and autosomal evidences

Based on their physical appearance, the Shompen have been considered to be of mixed Indo-Chinese, Malay, Negrito and Dravidian origin (Rizvi 1990). Our data, based on mitochondrial, Y-chromosomal, and autosomal analyses, however, support the Southeast Asian origin of this population. Although the current dataset has not revealed the occurrence of any haplogroups specific to Dravidian populations, this does not exclude such an occurrence in a larger sample size. With only a single Y-chromosomal lineage identified in the Shompen, it is plausible that the other lineages have been lost as a result

Fig. 5 Neighbour-joining phylogeny constructed using D_A distances based on autosomal microsatellite analysis, demonstrating the relationship of the Shompen to mainland Indian, East Asian and Southeast Asian populations
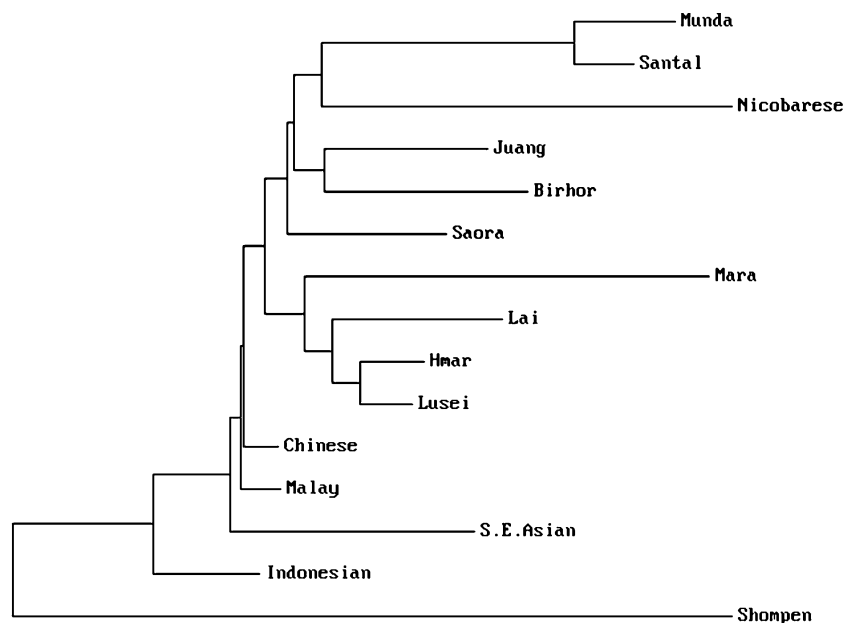
of the founder effect in this small population. The time of divergence of the Shompen from the Nicobarese deduced from the $\delta\mu^2$ distance of the autosomal data was ~14,000 years ago. It is likely that the mitochondrial B5a lineage was introduced later, independent of the older R12 lineage, probably along with bearers of the Y-chromosomal O2a lineage. The R12 haplogroup has also been observed among Great Andamanese (authors' unpublished data). However, its low frequency in the Great Andamanese does not support its origin in the Andaman Islands, ruling out the genetic contribution of Andaman Negritos to the current set of Shompen analysed.

With the Shompen demonstrating proximity to Southeast Asian populations with all genetic systems employed in this study, we infer that they trace their origins to mainland or island Southeast Asia with no discernible admixture in their contemporary gene pool.

References

Agarwal HN (1966) ABO blood groups and the sickle cell investigation among the Shompen of Great Nicobar. J Anthropol Soc 1:149–150

Alves C, Gusmao L, Damascero A (2004) Contribution for an African autosomic database (AmpFlSTR identifiler and PowerPlex 16 system) and a report on genotypic variations. Forensic Sci Int 139:201–205

Bagdonavicius A, Turbett GR, Buckleton JS, Walsh SJ (2002) Western australian sub-population data for the thirteen AmpFlSTR® Profiler Plus™ and COfiler™ STR loci. J Forensic Sci 47:1149–1153

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain JL (1988) Reconstruction of human evolution; bringing together genetic, archaeological and linguistic data. Proc Natl Acad Sci USA 85:6002–6006

Endicott P, Thomas M, Gilbert P, Stringer C, Lalueza-Fox C, Willerslev E, Hansen AJ, Cooper A (2003) The genetic origins of the Andaman Islanders. Am J Hum Genet 72:178–184

Forster P, Harding R, Torroni A, Bandelt HJ (1996) Origin and evolution of native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935–945

Goldstein DB, Ruis-Linares A, Cavalli-Sforza LL, Feldman ML (1995a) An evaluation of genetic distance for use with microsatellite loci. Genetics 139:463–471

Goldstein DB, Ruis-Linares A, Cavalli-Sforza LL, Feldman ML (1995b) Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci USA 92:6723–6727

Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. Biometrics 48:361–372

Kayser M, Brauer S, Weiss G, Underhill PA, Roewer L, Schiefenhovel W, Stoneking M (2000) Melanesian origins of Polynesians Y-chromosomes. Curr Biol 10:1237–1246

Kayser M, Brauer S, Weiss G, Schiefenhovel W, Underhill P, Stoneking M (2001) Independent histories of human Y-chromosome from Melanesia and Australia. Am J Hum Genet 68:173–190

Kayser M, Brauer S, Weiss G, Schiefenhovel W, Underhill P, Shen P, Oefner P, Tommaseo-Ponzerta M, Stoneking M (2003) Reduced Y-chromosome, but not mitochondrial DNA diversity in human populations from West New Guinea. Am J Hum Genet 72:281–302

Kittles RA, Perola M, Peltonen M, Bergen AW, Aragon RA, Virkkunen M, Linnoila M, Goldman D, Long JC (1998) Dual origins of Finns revealed by Y-chromosome haplotype variation. Am J Hum Genet 62:1171–1179

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

Linacre AM, Phillips P, Wally A, Goodwin W (2001) STR data for the SGM Plus loci from three Indonesian populations. Forensic Sci Int 123:232–234

Lum K, Rickards O, Ching C, Cann RL (1994) Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. Hum Biol 66:567–590

Man EH (1886) A brief account of the Nicobar Islands with special reference to the inland tribes of Great Nicobar. J Anthropol Inst 15:428–451

Melton T, Clifford S, Martinson J, Batzer M, Stoneking M (1998) Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. Am J Hum Genet 63:1807–1823

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol 19:153–170

Prasad BV, Ricker CE, Watkins WS, Dixon ME, Rao BB, Naidu JM., Jorde LB, Bamshad M (2001) Mitochondrial DNA variation in Nicobarese islanders. Hum Biol 73:715–725

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, Stoneking M (1995) Evolutionary history of the COII/tRNA Lys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12:604–615

Rizvi SNH (1990) The Shompen, a vanishing tribe of the Great Nicobar Island. In: Singh KS (ed) Seagull Books, Calcutta

Ruhlen W (1991) A guide to the World's languages, vol 1. Classification. Stanford University Press, Stanford, CA, pp 172–184

Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby S (2000) MtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet 67:718–726

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic tree. Mol Biol Evol 4:406–425

Sajantila A, Pääbo S (1995) Language replacement in Scandinavia. Nat Genet 11:359

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning—a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schneider S, Kueffer J-M, Roessli D, Excoffier L (2000) ARLEQUIN: a software for population genetic data analysis. University of Geneva, Geneva, Switzerland

Seah LH, Jeevan NH, Othman MI, Jaya P, Ooi YS, Wong PC, Kee SS (2003) STR data for the AmpFlSTR Identifiler loci in 3 ethnic groups (Malay, Chinese, Indian) of the Malaysian population. Forensic Sci Int 138:134–137

Singh KS (1998) India's communities national series People of India. Oxford University Press, Oxford

Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J et al (1999) Y-chromosome evidence for a north-ward migration of modern humans into eastern Asia during last Ice Age. Am J Hum Genet 65:1718–1724

Su B, Xiao C, Deka R, Sielstad MT, Kangwanpong D, Xiao J, Lu D, Underhill P, Cavalli-Sforza L, Chakraborty R et al (2000). Y-chromosome haplotypes reveal prehistorical migrations to the Himalayas. Hum Genet 107:582–590

Temple RC (1901) Census of India, 1901: Andaman and Nicobar Islands. Government of India Publication, Calcutta

Thangaraj K, Singh L, Reddy AG, Rao VR, Sehgal SC, Underhill PA, Pierson M, Frame IG, Hagelberg E (2003) Genetic affinities of Andaman islanders, a vanishing human population. Curr Biol 13:86–93

Torroni A, Schurr TB, Cabell ME, Brown MD, Neel JV, Larsen M, Smith DG (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Schozzari R, Obinu D, Savontaus M-L, Wallace DC (1996) Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835–1850

Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M et al (2001) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? Am J Hum Genet 69:1348–1356

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Detection of numerous Y-chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7:996–1005

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Betranpetit J, Francalacci P et al (2000) Y-chromosome sequence variation and history of human populations. Nat Genet 26:358–361

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Lin J, Su B, Pitchappan R, Shanmugalakshmi S et al (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci USA 98:10244–10249

Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP (2002) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70:635–651

Zhivotovsky LA (2001) Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. Mol Biol Evol 18: 700–709

Zhivotovsky LA, Underhill PA, Cinnioğlu 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