

Unique origin of Andaman Islanders: insight from autosomal loci

K. Thangaraj · G. Chaubey · A. G. Reddy ·
V. K. Singh · L. Singh

Received: 9 April 2006 / Accepted: 1 June 2006 / Published online: 19 August 2006
© The Japan Society of Human Genetics and Springer-Verlag 2006

Abstract Our mtDNA and Y chromosome studies lead to the conclusion that the Andamanese “Negrito” mtDNA lineages have survived in the Andaman Islands in complete genetic isolation from other South and Southeast Asian populations since the initial settlement of the region by the out-of-Africa migration. In order to obtain a robust reconstruction of the evolutionary history of the Andamanese, we carried out a study on the three aboriginal populations, namely, the Great Andamanese, Onge and Nicobarese, using autosomal microsatellite markers. The range of alleles (7–31.2) observed in the studied population and heterozygosity values (0.392–0.857) indicate that the selected STR markers are highly polymorphic in all the three populations, and genetic variability within the populations is significantly high, with a mean gene diversity of 77%. The Andaman “Negrito” populations do not show particular affinities either with the African populations or with the Indian populations, confirming their unique origin. In contrast, Nicobarese show close affinities with the Southeast Asian populations, suggesting their recent entry in the Islands.

Keywords Andamanese · Onge · Nicobarese · STRs · Microsatellites · Autosomes · Enigmatic tribes

K. Thangaraj · G. Chaubey · A. G. Reddy · V. K. Singh ·
L. Singh (✉)
Centre for Cellular and Molecular Biology,
Uppal Road, Hyderabad 500 007, India
e-mail: lalji@ccmb.res.in

G. Chaubey
Institute of Molecular and Cell Biology,
University of Tartu and Estonian Biocentre,
Tartu 51010, Estonia

Introduction

The “Negrito” populations found scattered in parts of southern India, the Andaman Islands, Malaysia and the Philippines are considered to be the relic of early modern humans and hence assume considerable anthropological and genetic importance. Their gene pool is slowly disappearing either due to assimilation with adjoining populations, as in the case of the Semang of Malaysia and Aeta of the Philippines, or due to their population collapse, as is evident in the case of the aboriginals of the Andaman Islands. The Andaman and Nicobar Islands are inhabited by six enigmatic indigenous tribal populations, of which four have been characterized traditionally as “Negritos” (the Jarawa, Onge, Sentinelese and Great Andamanese) and two as “Mongoloid” (the Nicobarese and Shompens) populations. Although the Negrito and the Mongoloid racial stocks are phenotypically similar to Africans and Southeast Asians, respectively, their origin was a mystery until recently. Our previous study on the vanishing tribal populations of the Andaman and Nicobar Islands was the first comprehensive molecular study using both the Y chromosome and mitochondrial (mtDNA) markers that provided deep insight into their origin and migration (Thangaraj et al. 2003). Analysis of the Onge and Greater Andamanese complete mtDNA sequences revealed the existence of two unique clades, which we proposed as M31 and M32 (Thangaraj et al. 2005a). Most importantly, the information from the complete mtDNA sequences showed that there is not a single coding region mutation that would support the relatedness of these two new haplogroups with any of the known Indian or East Asian mtDNA haplogroups. We suggested that both the no-

vel haplogroups, M31 and M32, are not reported anywhere and might have originated ‘in-situ’ at different time scales on this island, and we proposed that the Andamanese Negrito mtDNA lineages might have survived in the islands in complete genetic isolation from other South and Southeast Asian populations since the initial settlement of the region by the out-of-Africa migration (Thangaraj et al. 2005a).

However, all the above studies are based on the uniparently inherited mitochondrial or Y chromosomal DNA information. Hence, in order to obtain a robust reconstruction of the evolutionary history of these islanders, we carried out a study on three of the six aboriginal populations (the Great Andamanese, Onge and Nicobarese) employing autosomal microsatellite markers. These markers have been analysed extensively for inferring genetic relationships among continental populations (Deka et al. 1999) as well as between geographically contiguous populations (Reddy et al. 2001, 2005; Pandu et al. 2006; Thangaraj et al. 2006a). The present analysis, using autosomal microsatellite markers on the Andaman and Nicobar Islanders, reveals the molecular diversity present among these isolated, vanishing population groups and explicates their genetic relationship with other populations.

Materials and methods

Blood samples were collected from 97 enigmatic Andaman islanders including: 15 Great Andamanese,

46 Onge and 36 Nicobarese (Fig. 1). The actual population size of these tribal populations is very small; their origin has been discussed elsewhere (Thangaraj et al. 2003, 2005a). The Great Andamanese and Onge were hunter-gatherers, but presently they are in a government settlement and receive food grains from the government. However, the Onge are still maintaining their forest-dwelling life, while the Nicobarese are culturally reverted agriculturists. DNA was isolated from the above samples following the protocol published elsewhere (Thangaraj et al. 2002). DNA samples were quantified by spectrophotometer and size fractionated in 0.8% agarose gel. To amplify nine STR loci (Amp/profiler plus kit) in a multiplex reaction following the manufacturer’s instructions (Applied Biosystems, USA), 1.0 ng of genomic DNA was used. Two positive and two negative controls were used along with every set of PCR reactions. PCR amplicons were analyzed in an ABI Prism 3700 Genetic Analyzer (Applied Biosystems 2001) using GeneScan and Genotyper software (Perkin Elmer) to obtain the allelic description, which is now available elsewhere (Thangaraj et al. 2006b).

Statistical analysis

The allele frequencies were computed by a single-gene counting method. Arlequin software version 2.00 was used to obtain observed and expected heterozygosity locus-by-locus AMOVA, gene diversity and the exact test for Hardy–Weinberg equilibrium probabilities

Fig. 1 Map of Andaman Nicobar Islands, India, showing sampled area (arrow marks)



(Schneider et al. 1997). The P values are obtained for Hardy–Weinberg equilibrium. Locus-by-locus AMOVA analysis has been done by grouping into (1) linguistic and (2) geographical groups. Pairwise genetic distances between the studied populations were computed as a linearization of $F_{ST}/(1-F_{ST})$ (Slatkin 1995) using ARLEQUIN 2.00 software (Schneider et al. 1997). These linearized distance values were used to create a MDS plot (Fig. 2) based on microset data using SPSS 11.5 (Chicago, IL). In order to clarify the genetic affinity between the studied populations, neighbor-joining (NJ) (Saitou and Nei 1987) phylogeny was constructed using the D_A distance (Nei and Roychoudhury 1993), which has been found to be effective for analyzing closely related populations.

Results and discussion

Allele distribution patterns exhibit considerable variation among the natives Andaman and Nicobar Islanders. The Great Andamanese had a wider range of alleles compared to the Onge and Nicobarese, most probably due to the admixture, as our earlier study revealed that their original paternal genome was replaced by the mainland male labors brought in by the English dynasty (Thangaraj et al. 2003). However, they make a unique group with the Onge in the MDS plot and NJ tree (Figs. 2, 3), suggesting that their original alleles are so unique that they are sufficient to group them with the Onge. The number of alleles (7–31.2) observed in the studied population and heterozygosity values (0.392–0.857) indicate that the selected STR markers are highly polymorphic even in the highly

endogamous populations. Therefore, these markers are suitable for generating variation among the highly inbreeding populations. The average observed heterozygosity (Table 1) was almost equal for all three of the populations, and they are much lower than 0.08–0.88, the average observed heterozygosity for mainland Indians (Kashyap et al. 2003; Reddy et al. 2005), reflecting the aboriginals' considerable isolation and inbreeding. The exact test for Hardy–Weinberg equilibrium (Table 1) suggests that most of the loci in the studied populations are in equilibrium ($P > 0.005$). The relatively long branches for the aboriginals of the Andaman and Nicobar Islands coupled with their low gene diversities (Table 2) further accentuate the effect of inbreeding (Nei and Roychoudhury 1993) among them. The genetic variability within the populations is significantly high, with a mean gene diversity of 77%. Analysis of molecular variance (AMOVA), presented in Table 3, revealed that as a single group, a great extent of genetic variation (89.13%) was present within the populations of the Islands. To determine how the residual genetic variance was compartmentalized, we grouped the populations into (1) linguistic (Greater Andamanese and Onge in one group, Nicobarese with South East Asians in the second group and Dravidian mainland Indians in the third group) and (2) geographical groups (Andaman Islanders in one group, mainland Indian populations in another and South East Asians in a third group). The genetic variance between the groups was statistically significant only in the linguistic apportionment. The inter-population genetic relationship among Andaman Islanders and other populations was determined using multidimensional analysis. The plot (Fig. 2) of multidimensional scaling depicts population configurations in accordance with

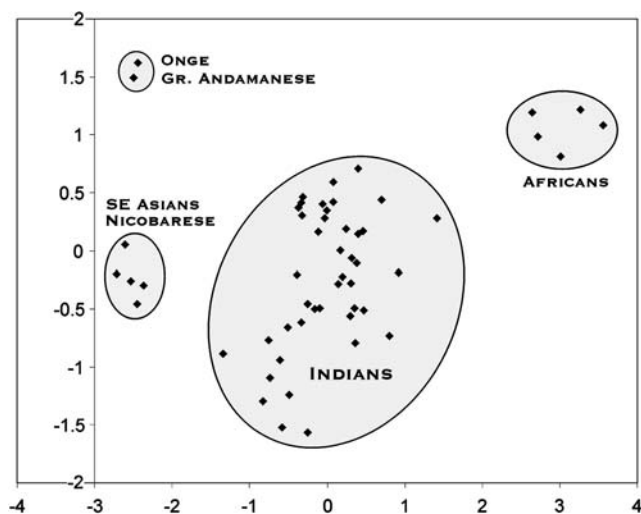


Fig. 2 MDS plot of Andaman Islanders with world populations

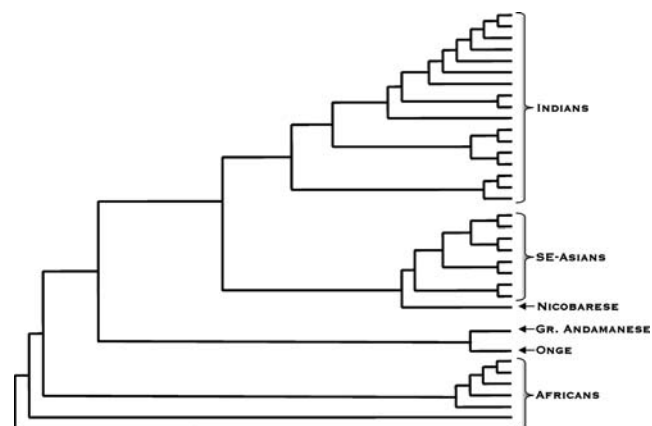


Fig. 3 Neighbor-joining phylogeny of aboriginal Andaman and Nicobar populations with global populations constructed based on nine microsatellite markers

Table 1 Heterozygosity values of the STR makers among the tribal populations of the Andaman and Nicobar Islands

Locus	Great Andamanese	Onge	Nicobarese
D3S1358	0.467 (0.22824±0.00115)	0.667 (0.63359±0.00145)	0.392 (0.03623±0.00051)
VWA	0.650 (0.66179±0.00157)	0.740 (0.69506±0.00150)	0.643 (0.18294±0.00093)
FGA	0.725 (0.53903±0.00175)	0.815 (0.06410±0.00066)	0.786 (0.02088±0.00026)
D8S1179	0.775 (0.42811±0.00157)	0.629 (0.54312±0.00143)	0.676 (0.03851±0.00030)
D21S11	0.825 (0.11793±0.00106)	0.444 (0.02591±0.00053)	0.643 (0.02806±0.00037)
D18S51	0.400 (0.23398±0.00116)	0.778 (0.53134±0.00084)	0.857 (0.03203±0.00047)
D5S818	0.384 (0.60564±0.00200)	0.629 (0.72463±0.00141)	0.750 (0.40047±0.00166)
D13S317	0.615 (0.03372±0.00055)	0.630 (0.04979±0.00070)	0.464 (0.02185±0.00047)
D7S820	0.775 (0.55357±0.00151)	0.555 (1.00000±0.00000)	0.393 (0.00741±0.00021)

P values for Hardy–Weinberg equilibrium are given in the parentheses

Table 2 Gene diversity among the tribal populations of Andaman and Nicobar Islands

Populations	Gene diversity
G. Andamanese	0.684802±0.372653
Onge	0.638600±0.344268
Nicobarese	0.659686±0.403016

their ethnic affiliations. The Great Andamanese and Onge were found to separate strictly from other populations, but the Nicobarese showed a relation with the Mongoloid populations, corroborating our previous studies (Thangaraj et al. 2005a, b).

D_A distances were employed for construction of the neighbor-joining phylogeny (Saitou and Nei 1987) represented in Fig. 3. The D_A distances were found to be extremely high, indicating the occurrence of population bottlenecks in the aboriginals. Genetic distance values are known to increase rapidly when bottlenecks have occurred (Chakraborty and Nei 1977), probably the reason for the occurrence of the Negrito populations of the Andaman Islands as extreme outliers in the neighbor-joining phylogeny (Fig. 3) after Africans. The Nicobarese exhibited an affinity towards other Indo-Mongoloid groups as reported earlier (Kashyap et al. 2003; Thangaraj et al. 2005a, b). Consistent with the populations that have experienced a recent reduction

of their effective population size, the Negritos of the Andaman Islands exhibited a reduction of allele numbers and gene diversity.

The extent of isolation, genetic drift and recent reduction in effective population size among Negrito populations of the Andaman Islands is reflected in the absence of allele diversity and the large genetic distances they demonstrate from other mainland Indian populations. Their phenotypic, linguistic and other genetic attributes corroborate substantial periods of isolation. The mitochondrial study (Endicott et al. 2003) showed that in the haplogroup M background, these populations are genetically similar to Asian populations. The Onge, genetically equidistant from the African and mainland Indian populations, are probably the only remnants of early modern humans who have preserved their genetic identity through extreme seclusion from external intrusion. The Nicobarese, the Mongoloid population inhabiting the Nicobar Islands, exhibit similarities with other Asian Mongoloid and Austro-Asiatic/Mundari speakers. The present study strongly corroborates our earlier conclusions based on mtDNA and the Y chromosome (Thangaraj et al. 2003, 2005a, b). The Negritos of the Andaman Islands do not show a particular association either with the African populations or with the mainland Indian populations. Our findings also reveal that these

Table 3 Locus-by-locus AMOVA and *Fst* analysis

Locus	Among the group % Variation	Among the population % Variation	Within the population % Variation	FSC (<i>P</i> value)	FST (<i>P</i> value)	FCT (<i>P</i> value)
D3S1358	19.87974	3.41118	76.70908	0.04258 (0.02737)	0.23291 (0.00000)	0.19880 (0.24145)
VWA	1.52515	3.10688	95.36798	0.03155 (0.07429)	0.04632 (0.00880)	0.01525 (0.24438)
FGA	1.90965	2.12552	95.96482	0.02167 (0.13783)	0.04035 (0.00684)	0.01910 (0.51711)
D8S1179	1.25387	5.38800	93.35814	0.05456 (0.01173)	0.06642 (0.00000)	0.01254 (0.25122)
D21S11	-2.84316	14.97227	87.87090	0.14558 (0.00000)	0.12129 (0.00000)	0.02843 (0.48485)
D18S51	-4.58879	20.66148	83.92732	0.19755 (0.00000)	0.16073 (0.00000)	0.04589 (0.48876)
D5S818	0.97240	7.80122	91.22638	0.07878 (0.00098)	0.08774 (0.00000)	0.00972 (0.25709)
D13S317	3.12667	9.00483	87.86850	0.09295 (0.00000)	0.12131 (0.00000)	0.03127 (0.24731)
D7S820	20.02630	-2.08650	82.06020	0.02609 (0.64027)	0.17940 (0.00000)	0.20026 (0.24633)

aboriginal populations of the Andaman and Nicobar Islands have sustained their shared ancestry and illustrate that they harbor deeply rooted genetic evidence.

Acknowledgments We thank the anonymous subjects of Andaman Island, India, who volunteered to give their blood samples.

References

- Applied Biosystems, PCR Protocol for AmpFISTR Profiler Plus (2001) Applied Biosystems Human Identification Group, Applied Biosystems, USA
- Chakraborty R, Nei M (1977) Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. *Evolution* 31:347–356
- Deka R, Shriver MD, Yu LM, Heidreich EM, Jin L, Zhong Y, McGarvey ST, Agarwal SS, Bunker CH, Miki T, Hundrieser J, Yin S, Raskin S, Barrantes R, Ferrell RE, Chakraborty R (1999) Genetic variation at 23 microsatellite loci in 16 human populations. *J Genet* 78:99–121
- Endicott P, Gilbert MT, 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
- Kashyap VK, Sarkar N, Sahoo S, Sarkar BN, Trivedi R (2003) Genetic variation at 15 microsatellite loci in human populations of India. *Curr Sci* 85:464–473
- Nei M, Roychoudhury AK (1993) Evolutionary relationships of human populations on a global scale. *Mol Biol Evol* 10:927–943
- Pandu G, Gandhi KP, Sharma JD, Chaubey G, Thangaraj K (2006) Genetic profile of nine STR loci among Goud and Padmashali populations of Andhra Pradesh, India. *Forensic Sci Int* 157:201–205
- Reddy BM, Sun G, Luis JR, Crawford MH, Hemam NS, Deka R (2001) Genomic diversity at 13 short tandem repeat loci in a substructured caste population, Golla, of southern Andhra Pradesh, India. *Hum Biol* 73:175–190
- Reddy BM, Naidu VM, Madhavi VK, Thangaraj K, Langstieh BT, Venkataramana P, Kumar V, Singh L (2005) STR data for the AmpF/STR Profiler Plus loci among 27 population of different social hierarchy from southern part of Andhra Pradesh, India. *Forensic Sci Int* 149:81–97
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic tree. *Mol Biol Evol* 14:406–425
- Schneider S, Rossie D, Excoffier L (1997) Arlequin Ver 2.000, a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva. <http://www.anthropologie.unige.ch/arlequin>
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462
- Thangaraj K, Joshi MB, Reddy AG, Gupta NJ, Chakravarty B, Singh L (2002) CAG repeat expansion in the androgen receptor gene is not associated with male infertility in Indian populations. *J Androl* 23:813–816
- Thangaraj K, Singh L, Reddy AG, Rao VR, Sehgal SC, Underhill PA, Pierson M, Frame IG, Hagelberg E (2003) Genetic affinities of the Andaman islanders, a vanishing human population. *Curr Biol* 13:86–93
- Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, Singh L (2005a) Reconstructing the origin of Andaman islanders. *Science* 309:996
- Thangaraj K, Sridhar V, Kivisild T, Reddy AG, Chaubey G, Singh VK, Kaur S, Agarawal P, Rai A, Gupta J, Mallick CB, Kumar N, Velavan TP, Suganthan R, Udaykumar D, Kumar R, Mishra R, Khan A, Annapurna C, Singh L (2005b) Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations. *Hum Genet* 116:507–517
- Thangaraj K, Chaubey G, Singh VK, Reddy AG, Pavate PP, Singh L (2006a) Genetic profile of nine autosomal STR loci among Halakki and Kunabhi populations of Karnataka, India. *J Forensic Sci* 51:190–192
- Thangaraj K, Chaubey G, Reddy AG, Singh VK, Singh L (2006b) Autosomal STR data on the enigmatic Andaman islanders. *J Forensic Sci* Apr 28 [Epub ahead of print]