Am. J. Hum. Genet. 72:178–184, 2003

Report

The Genetic Origins of the Andaman Islanders

Phillip Endicott,[1] M. Thomas P. Gilbert,[1] Chris Stringer,[2] Carles Lalueza-Fox,[3] Eske Willerslev,[4] Anders J. Hansen,[4] and Alan Cooper[1]

[1]Henry Wellcome Ancient Biomolecules Centre, Department of Zoology, University of Oxford, Oxford; [2]Department of Palaeontology, The Natural History Museum, London; [3]Secció Antropologia, Departament de Biologia Animal, Universitat de Barcelona, Barcelona; and [4]Department of Evolutionary Biology, Zoological Institute, University of Copenhagen, Copenhagen

Mitochondrial sequences were retrieved from museum specimens of the enigmatic Andaman Islanders to analyze their evolutionary history. D-loop and protein-coding data reveal that phenotypic similarities with African pygmoid groups are convergent. Genetic and epigenetic data are interpreted as favoring the long-term isolation of the Andamanese, extensive population substructure, and/or two temporally distinct settlements. An early colonization featured populations bearing mtDNA lineage M2, and this lineage is hypothesized to represent the phylogenetic signal of an early southern movement of humans through Asia. The results demonstrate that Victorian anthropological collections can be used to study extinct, or seriously admixed populations, to provide new data about early human origins.

The 200 islands of the Andaman archipelago lie in an arc between Burma and Indonesia in the Bay of Bengal. Their inhabitants possess an extremely distinctive phenotype, typified by a small average height, gracile build, dark pigmentation, and unusual hair morphology. Victorian anthropologists noted phenotypic similarities to the pygmoid peoples of Africa and suggested a recent African origin (e.g., Dobson 1875). However, to differing extents, these physical characteristics are present in populations scattered throughout Asia and Near Oceania. These groups are often referred to as "Negritos" (from the Spanish diminutive for Negro) and predominantly follow a mobile, hunting/foraging mode of subsistence. This has led to the alternative suggestion that they might represent an ancient substratum of humanity in Asia, predating later migrations and agrarian expansion events (Coon 1966; Molnar 1983; Cavalli-Sforza et al. 1994). Populations matching this description still exist in southern India, Sri Lanka, Malaysia, Indonesia, Papua New Guinea, and China. Many others appear to have survived being replaced or assimilated until at least the advent of European colonialism (Wallace 1883; Forbes 1884; Wollaston 1912).

Because of their island location, the Andamanese had little sustained contact with the outside world until the mid-19th century. This is reflected in their languages, which are linguistic isolates with striking morphological features (Wurm 1971). Although there is a limited affinity—on lexical, structural, and typological grounds only—with a few small isolates in Papua New Guinea and eastern Indonesia (Greenberg 1971; Wurm and McElhanan 1975; Ruhlen 1991), there is no widely accepted interpretation of the relationship of the Andamanese languages to the extant linguistic families of the region. The archaeological evidence for an early occupation of the Andamans is scant because of the mobile lifestyle of the inhabitants and the limited number of excavations that have taken place. At present, the oldest confirmed radiocarbon date is just >2,000 years old, and there are no artifacts to suggest contact or trade with the world outside the archipelago (Cooper 1993).

History confirms the insular existence of the Andaman Islanders over the past 2 millennia, subject only to visits from slave raiders and resource gatherers (Cooper 1988). Seafarers have steadfastly avoided contact because of the fearsome reputation of the inhabitants, preferring instead the long-established port of call in the neighboring Nicobars (Cooper 1988). The British founded a penal colony at Port Blair in 1858, with disastrous consequences for

Received August 15, 2002; accepted for publication October 7, 2002; electronically published December 11, 2002.

Address for correspondence and reprints: Dr. Alan Cooper, Henry Wellcome Ancient Biomolecules Centre, Department of Zoology, University of Oxford, Oxford, OX1 3PS England. E-mail: alan.cooper@zoology.ox.ac.uk

0002-9297/2003/7201-0019$15.00

the indigenous population, whose numbers declined rapidly because of disease and social disruption (Man 1932; Portman 1990; Temple 1903). Of the 13 linguistically defined groups known in the mid-19th century (fig. 1), only three survive (Jarawa, Sentinelese, and Onge). This triad was connected with the Greater Andamanese language clade on a typological—rather than a cognatic—basis, suggesting a historical separation of considerable depth (Radcliffe-Brown 1914; Portman 1990).

The linguistic division appears to have been matched by phenotypic variation, since E. H. Man described the Jarawa as not bearing the least resemblance to their neighbors (the Aka-Bea), being taller, differently proportioned, and having another type of hair morphology (Barnard-Davis 1867). Conversely, the inhabitants of North Sentinel Island were connected to the Jarawa on the basis of physical, as well as linguistic, similarities (Portman 1990). Subsequent studies classified all 13 speech communities and demonstrated that the Onge, Jarawa, and Sentinelese could also be differentiated from the Greater Andamanese by material culture forms and social practice (Man 1932; Portman 1990). These sociocultural, phenotypic, and linguistic delineations are reinforced by the geographical distributions of populations encountered by the British in 1858 (fig. 1). The Onge and Sentinelese are on islands to the south, whereas the Jarawa appear to form a wedge among the Aka-Bea, with whom they were perpetually at war.

The epigenetic data suggest that the Andaman Islanders originated from either two colonization events or a single founding population that has been subdivided for an extended length of time. The geographic and social isolation of the Andaman Islands means that these hypotheses can be tested using genetic data, but it is unfortunate that the majority of Andamanese peoples and their languages are long dead. The Sentinelese and Jarawa (neither of whom is currently thought to number >100) survive because of their continuing hostility toward colonialism, and are, therefore, inaccessible. The Onge have been through a sustained population bottleneck, and the 10 Greater Andamanese speech communities are now represented by <40 admixed descendants. Consequently, the accessible living populations do not represent either the genetic diversity or the population substructure that existed prior to colonial occupation. However, although powerless to prevent the decline of the Andamanese, Victorian anthropologists were keen to preserve a record of physical variation among humans and obtained collections of skeletal material. These collections now represent a unique scientific resource for genetic research by analysis of ancient DNA (aDNA).

Ancient human DNA studies are extremely problematic because of the extreme risk of contamination of samples and laboratories with modern human material, and it is critically important that a number of criteria be followed

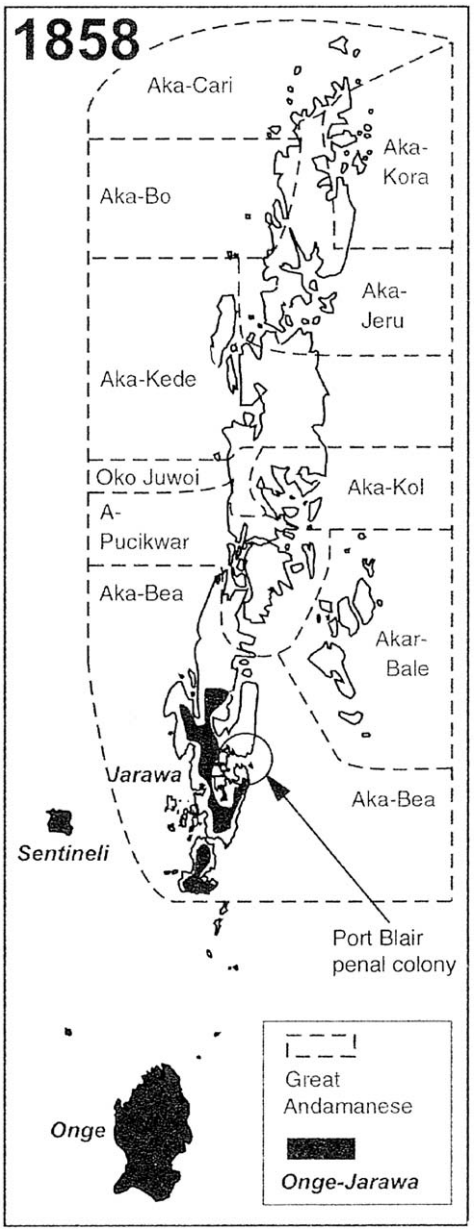

Figure 1 Map of the Andaman Islands showing the distribution of the known communities of 1858. The two clades (*black* and *white*) are defined by linguistic affinities. The term "Jarawa" is here extended to the inhabitants of Sentinel Island, on the basis of their similar phenotype and a similarity in language. The Greater Andamanese practiced body tattooing, whereas the Onge and Jarawa did not. There were also differences in material culture.

(Cooper and Poinar 2000). We used a physically isolated, dedicated aDNA laboratory in the Oxford University Museum of Natural History and extracted DNA from single teeth as described by Gilbert et al. (2003*b* [in this issue]) using glove boxes, complete protective clothing and breathing masks, and all appropriate controls (Cooper and Poinar 2000). Samples from collections at the

Natural History Museum (NHM), London, were obtained from nine Aka-Bea, two Jarawa, and one individual from the adjoining Nicobar Islands. Four teeth were also sent directly from the NHM to aDNA laboratories in Copenhagen and Barcelona for independent, blind replication. GenBank accession numbers and NHM numbers are as follows: Aka-Bea 1, GenBank AY191257, NHM 1905.11.25.1; Aka-Bea 2, GenBank AY191258, NHM 8.0302; Aka-Bea 3, GenBank AY191255, NHM 8.0306; Aka-Bea 4, GenBank AY191261, NHM 1905.11.25.3; Aka-Bea 5, GenBank AY191259, NHM 8.0209; Aka-Bea 6, GenBank AY191256, NHM IM20.4; Aka-Bea 7, GenBank AY191262, NHM 1905.11.25.2; Aka-Bea 8, GenBank AY191260, NHM 1865.5.26.1; Aka-Bea 9, GenBank AY191263, NHM 1956.17.10.2; Jarawa 1, GenBank AY191264, NHM 8.0501; Jarawa 2, GenBank AY191265, NHM 8.051; and Nicobar 1, GenBank AY191266, NHM 1886.5.14.1.A. All samples were chosen from a resource of >50 specimens for their visually assessed state of preservation. The accompanying provenance indicates that all except the Jarawa—whose remains date from 1925—died within the first 50 years of British occupation. The Greater Andamanese practiced infanticide on any miscegenated progeny (Man 1932), and the Jarawa were hostile to outsiders. Consequently, these samples have a high probability of being free from admixture.

mtDNA is commonly used in aDNA studies because of the high copy number in living cells (up to 5,000–10,000 times that of single-copy nuclear sequences) and its effectively haploid, nonrecombining nature (Avise et al. 1987). The hypervariable region (HVR-1) of the mitochondrial control region (D-loop) is routinely used to study human phylogeographic patterns because of the rapid evolutionary rate (≥10 times that of the protein-coding region of mtDNA [Vigilant et al. 1991; Richards and Macaulay 2001]). The high average mutation rate of the D-loop and a marked heterogeneity in rates between sites can lead to irresolvable genealogies. However, comparisons of large amounts of mitochondrial protein-coding and control region data indicate that these inconsistencies tend to be due to a lack of genealogical resolution rather than incorrect topology (Finnilä et al. 2001; Richards and Macaulay 2001). To obtain maximum resolution from the shorter aDNA sequences, it is important to have a well-characterized contemporary phylogeny in which they can be placed. Consequently, SNPs from the protein-coding region were used to place the ancient specimens within one of the known monophyletic clades (haplogroups); HVR-1 data was used to assign membership of mtDNA lineages within these clades.

HVR-1 was amplified and sequenced from each specimen as described by Gilbert et al. (2003*b* [in this issue]), using four sets of overlapping primer pairs (L15996/H16139, L16131/H16218, L16209/H16365, and L16287/H16410) (from Handt et al. [1996], except for L15996 [5′-CTCCACCATTAGCACCCAAAGC-3′]). In addition, two protein-coding region SNPs that are specific to the Asian mtDNA M haplogroup were sequenced with primers L10353F (5′-GCCCTAAGTCTGGCCTATGAG-3′) and H10442R (5′-TGAGTCGAAATCATTCGTTTTG-3′). All numbering refers to the Cambridge reference sequence (CRS) of Anderson et al. (1981). Sequences were cloned to detect contaminants and artifacts associated with postmortem template modification (Gilbert et al. 2003*b* [in this issue], 2003*a* [in this issue]). Several samples were treated with the *Escherichia coli* uracil-N-glycosylase prior to amplification, to cleave deaminated nucleotides and confirm the presence of ancient damaged templates (Gilbert et al. 2003*b* [in this issue]). For full details on regions amplified, see table 1.

HVR-1 sequence ambiguities, consistent with studies indicating a correlation between mutation rates in vivo and postmortem hotspots of template damage (Gilbert et al. 2003*b* [in this issue]), were resolved by analysis of multiple clones. Known contaminants (e.g., researchers and previously amplified sequences) and artifacts caused by the amplification of damaged templates from a low copy number were carefully and systematically eliminated. No between-sample contamination or jumping PCR (Pääbo et al. 1990; Gilbert et al. 2003*a* [in this issue]) artifacts were detected in the sequences, and clones identified a single consensus sequence with scattered singleton substitutions. Only a limited amount of sequence diversity was detected between 15996 and 16218, so efforts were concentrated on the most informative region between 16209 and 16410. The teeth selected for independent replication were chosen to represent two haplotypes to control for cross contamination, and both replications matched the Oxford sequences (table 1). Tests in Copenhagen and Oxford showed that large amplifications (>250 bp) resulted in the putative Andaman sequences being replaced by obvious modern contaminants.

The protein-coding region sequences produced clear results, with no ambiguities, and showed that all 12 specimens possessed 10400T and 10398G, placing them in the Asian M haplogroup (Passarino et al. 1996; Macaulay and Richards 2000) and ruling out an African origin. The HVR-1 data separate them into two lineages, identified on the Indian mainland (Bamshad et al. 2001) as M4 and M2 (fig. 2), by the presence of 16311C and 16319A, respectively. The high mutation rate of 16311 may have given rise to paraphyletic lineages within M, and it is possible that the Andaman M4 is one of these. This would not, however, affect the main findings of this study. The Andamanese M2 contains two haplotypes, 16223T/16319A/16357C and 16223T/16319A/16344T/16357C (the former is presumed ancestral to the latter). Neither of the Andaman M2 variants is present in cur-

Table 1

MtDNA Haplotypes in 11 Andamanese Individuals and 1 Nicobar Individual

	Substitution relative to CRS							
	Coding Region		D-Loop (Control Region)					
Sample	10398G	10400T	16223T	16311C	16319A	16344T	16357C	Haplotype
Aka-Bea 1	C	C	CRU		CRU	CRU	CR	M2
Aka-Bea 2	D	D	CU		CU	CU	C	M2
Aka-Bea 5	C	C	C		C	C	C	M2
Aka-Bea 8	C	C	R		CR	CR	CR	M2
Aka-Bea 3	C	C	C		C		C	M2
Aka-Bea 6	D	D	U		U		D	M2
Aka-Bea 4	C	C	CR	CR				M4
Aka-Bea 7	C	C	CU	CU				M4
Aka-Bea 9	D	D	DR	R				M4
Jarawa 1	D	D	N	C				M4
Jarawa 2	C	C	C	C				M4
Nicobar	C	C	CU	CU				M4

Note.—C = Cloned, D = Direct Sequence, R = Replication Clones, U = UNG Clones, N = No sequence obtained.

rent Indian data sets, suggesting that the Andaman sequences represent novel members of M2 and an extension to its known geographical distribution (fig. 3). Within the small data set, there appears to be no obvious relationship between linguistic groups and mitochondrial haplotypes (table 1), although more non–Aka-Bea samples are required for statistical significance.

Other studies, utilizing HVR-1 data, have calculated M2 and M4 to coalesce at 63,000 ± 6,000 years ago and 32,000 ± 7,500 years ago, respectively (Kivisild et al. 1999*b*). The other known major components of Indian M are calculated to be <40,000 years old (Kivisild et al. 1999*b*). The relatively basal position of the Andaman M2 haplotypes in the median joining network (fig. 2), together with their absence from Indian data sets, suggests that 16344T and 16357C have developed in situ, after an early colonization. Otherwise, these M2 variants should have been preserved in later mainland population expansions, including M2a and M2b, thought to have occurred around 17,000–32,000 years ago (Kivisild et al. 1999*b*). Alternatively, it is possible that the haplotypes have become extinct in India or are present at a low frequency and have not yet been sampled, but, in each case, an early settlement of the Andaman Islands by an M2-bearing population is implied. It is conceivable that a haplotype located in Andhra Pradesh (Kivisild et al. 1999*a*) (16223T/16274A/16319A/16357C) could represent an ancestral sequence to the Andaman M2. This would require a back mutation of 16274A, but it is equally likely, given the absence of an intermediary sequence, that this belongs to M2b with 16357C occurring twice. The Andaman M4 haplotype is also basal within its provisional lineage but is still present among populations in India (fig. 3), suggesting it was subject to the late Pleistocene population expansions and, therefore, consistent with its date of coalescence. Thus, there is a potential chronological separation between the origin of the 223T/319A/357C haplotype outside mainland India—and the development of the 223T/311C haplotype in mainland Asia. This suggests that there may have been at least two distinguishable founding events for the Andaman Islands, and this hypothesis can be further evaluated by a second method of population genetic inference.

The total population of the Andaman Islands in 1858 is thought to have been >3,500 (Temple 1903; Man 1932). For a conservative approach, this study takes the upper limit, calculated by the carrying capacity of the Andaman Islands, of ~5,000 (Erickson and Beckerman 1975). The effective population size (N_{ef}) for breeding females is approximated to 10% of N (500) and one generation taken to be 20 years. Using the methods of Tavaré (1984) and the upper CI for M4, the probability of two or more lineages surviving after 39,500 years when $n = 11$ is unlikely ($P < .05$) unless there has been population substructure. This is the oldest possible date for a single colonization provided by the CIs for the ages of M2 and M4. Given the substantially older date of coalescence for M2 and the current lack of a source population for the Andaman haplotypes, the data is consistent with at least two founding events. Alternative explanations include a recent immigration bearing the M2 and/or M4 lineages, although that is not consistent with the distinctive languages of the Andaman Islanders. The current lack of data for Myanmar and Malaysia makes it difficult to evaluate such potential ancestral populations. There is no data on the communities farther east that have a limited linguistic affinity with the Andamanese languages, but there is an unconfirmed suggestion that the Onge used to raid the Nicobars during the 18th century (Man 1932), and our Nicobar sample

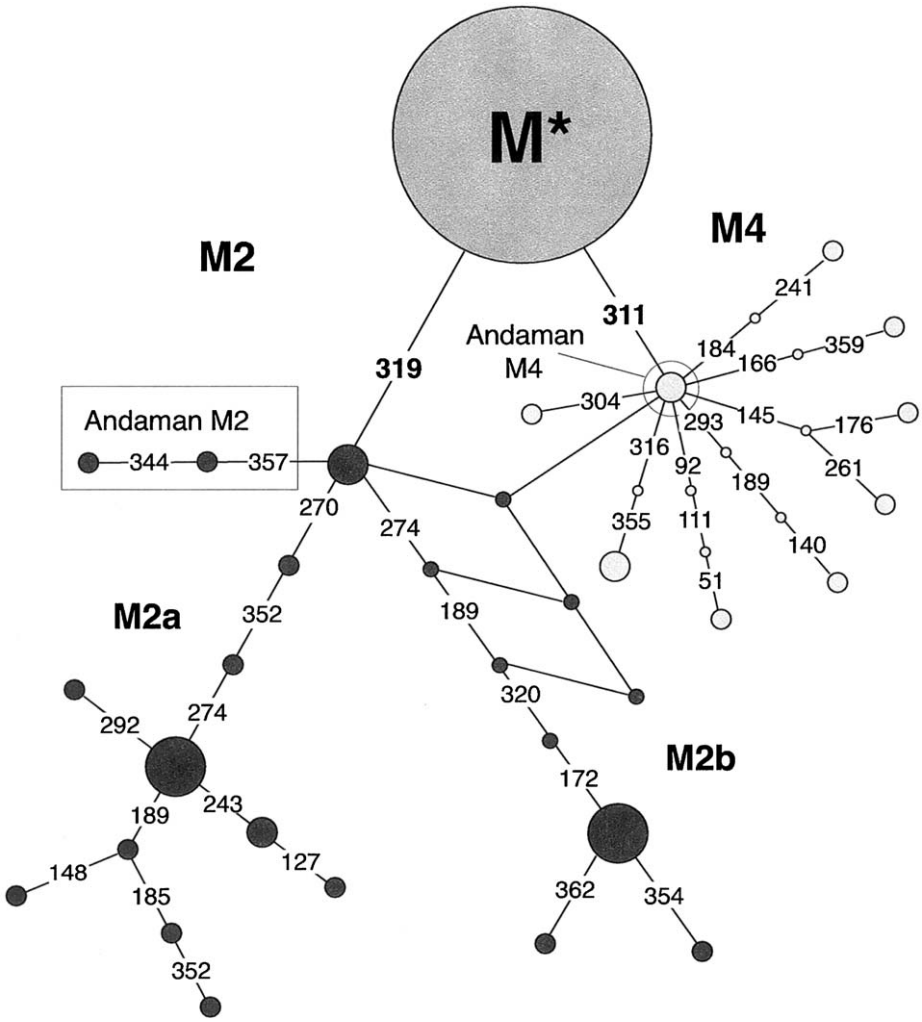

Figure 2 Median joining network (after Kivisild et al. 1999*b*) showing the known components of the M2 and M4 mtDNA lineages belonging to Asian haplogroup M in mainland India and the Andaman Islands. Substitution positions minus 16,000 are shown and are relative to the CRS (Anderson et al. 1981). The reticulation is shown with equal weighting, although the higher mutation rate of 311 and the date of coalescence for M4 favor the assignment of these three variants to M2. The diameters of the nodes are an approximate indication of the population totals inferred from sampling to date. The "starlike" radiations of M4, M2a, and M2b are consistent with late Pleistocene population expansion events (17,000–32,000 years ago). The linear nature of M2 prior to these demographic signatures suggests a greater antiquity, confirmed by dates modeled on the coalescent process (63,000 ± 6,000 years ago compared with 32,000 ± 7,500 years ago for M4).

is assigned to M4. However, a recent study of contemporary mtDNA from the Nicobars revealed only one M4 haplotype and no M2 haplotypes (Prasad et al. 2001). So, unless there has been wholesale mtDNA replacement in the Nicobars during the last 150 years, there is currently no evidence of a source population for mtDNA admixture with the Andamans.

Detailed data on the distribution of mtDNA lineages within India provides evidence of a potential early route for human colonization in south Asia. There is a geographic cline (fig. 3) in the presence of haplogroup M among ethnic (noncaste) populations of India ($n = 169$), which decreases from ~71% of the total in the south to ~55% in the north (extrapolated from Mountain et al. [1995] and Roychoudhury et al. [2001]). This distribution is matched by a similar trend among caste populations ($n = 598$) of ~63% to ~50% (extrapolated from Kivisild et al. [1999*a*] and Mountain et al. [1995]). Both sets of figures support the idea that haplogroup M may represent the phylogenetic signature of an early, southern colonization route in Asia (Distell 1999; Kivisild et al. 1999*a*; Quintana-Murci et al. 1999). Of the known M lineages in India, M2 is the largest (Bamshad et al. 2001) and is distributed in a similarly pronounced cline. Within caste populations, M2 represents ~6% of M in the north but averages ~13% in the south. However, among ethnic pop-

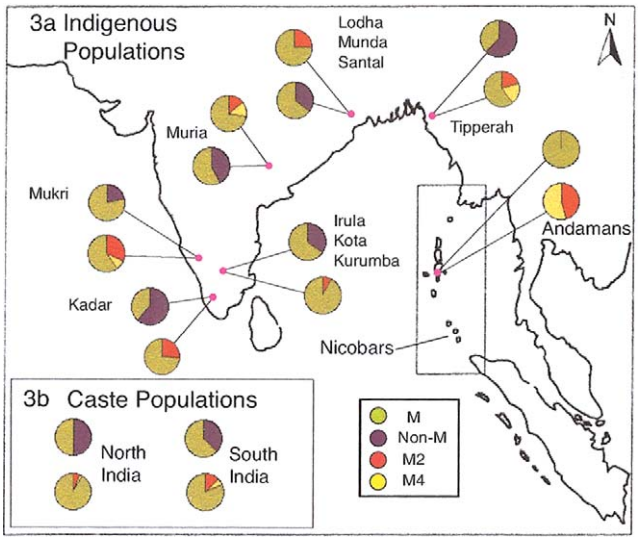

Figure 3 Distribution map for mtDNA lineages and haplogroups present in India and the Andaman Islands. The noncaste populations are considered individually, whereas the figures for castes are separated north/south along the borders of Goa, Karnataka, and Andhra Pradesh. The upper pie charts depict the percentage of M versus non-M haplotypes, whereas the lower ones indicate the percentage of M2 and M4 expressed as a proportion of all M haplotypes. The figures for ethnic populations from West Bengal and Tamil Nadu have been conflated in each case for the purposes of clarity, although they do conceal the fact that neither the Irula or Kota samples contained any M2. Despite this, the average of M2 in the total noncaste population (n = 169) exceeded 20%, compared to ~10% for caste populations (n = 598). New data sets for additional Indian ethnic groups confirm these distributions (n = 528, M2 = 23% [P. Majumder, personal communication]). This asymmetry is not reflected in the data for M4, which displays a more even distribution (Kivisild et al. 1999*b*).

ulations, the M2 component averages >20% of M and is geographically relatively constant. In contrast, although M4 also displays a cline south to north, in general, it has a more even demographic distribution (Kivisild et al. 1999*b*). The high frequency of M2 is consistent with its greater age, and its distribution suggests that many of the populations viewed as the autochthons of India because of their cultural inheritance (Majumder 2001; Roychoudhury et al. 2001) may also be genetic descendants of the early settlers of southern Asia. New data sets (P. Majumder, personal communication) confirm the absence of any Andaman M2 haplotypes among the ethnic populations of India (n = 528), despite a frequency of ~23% for the M2 lineage within M overall (n = 351, distributions not shown).

The early colonization of the Andaman archipelago by bearers of the M2 lineage supports the growing evidence of an early movement of humans through southern Asia and indicates that phenotypic similarities with African groups are convergent. It also suggests that early human migrants were capable of reaching all the islands of southeast Asia and, therefore, Near Oceania by the late Pleistocene. Such a dispersal is consistent with the scattered distribution of Negrito populations. All lines of evidence—social, cultural, historical, archaeological, linguistic, phenotypic, and genetic—support the conclusion that the Andaman Islanders have been isolated for a substantial period of time. It is not currently possible to categorically distinguish between two or more founding events and a single colonization followed by extensive population subdivision; a more detailed mtDNA phylogeny of south and southeast Asia may permit future work to differentiate between these two hypotheses. Whichever turns out to be correct, the implications for understanding the population dynamics of prehistory are profound. These findings illustrate the importance of sampling human biodiversity prior to significant modern admixture and extirpations and show that sequences derived from aDNA can have a significant role in the interpretation of contemporary human genetic distributions.

Acknowledgments

This project was funded by National Environmental Research Council (small grant NER/B/S/2000/00812). A.C. and T.G. are funded by the Wellcome Trust. E.W. and A.J.H. were supported by the Villiumkann Rasmussen Fonden, Denmark. We thank the Natural History Museum in London and its staff, especially Louise Humphrey and Robert Krusynski, for access to the specimens and for assistance with their archives. Harry Persaud of the British Museum also provided access to important archival material. Jaume Bertranpetit provided laboratory facilities and reagents for replication in Barcelona. Vincent Macaulay, Gil McVean (Oxford Department of Statistics), and George Weber of the Andaman Association provided helpful information. We also thank the sequencing staff in the Department of Zoology and the librarians of Oxford University. Partha Majumder generously provided details on unpublished Indian mtDNA data.

Electronic-Database Information

Accession numbers and URL for data presented herein are as follows:

GenBank, http://www.ncbi.nlm.nih.gov/ (for accession nos. AY191255 [Aka-Bea 3], AY191256 [Aka-Bea 6], AY191257 [Aka-Bea 1], AY191258 [Aka-Bea 2], AY191259 [Aka-Bea 5], AY191260 [Aka-Bea 8], AY191261 [Aka-Bea 4], AY191262 [Aka-Bea 7], AY191263 [Aka-Bea 9], AY191264 [Jarawa 1], AY191265 [Jarawa 2], and AY191266 [Nicobar 1])

References

Anderson S, Bankier AT, Arrel B, de Brujin M, Coulson A, Drouin J, Eperon I, Nierlich D, Roe B, Sanger F, Schreier P, Smith A, Staden R, Young I (1981) Sequence and organisation of the human mitochondrial genome. Nature 290:457–465

Avise J, Arnold J, Ball R, Bermingham E, Lamb T, Neigle J, Reeb C , Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Ann Rev Ecol Syst 18:489–522

Bamshad M, Kivisild T, Watkins S, Dixon M, Ricker C, Rao B, Naidu J, Prasad B, Reddy P, Rasanayagam A, Papiha S, Villems R, Redd A, Hammer M, Nguyen S, Carroll M, Batzer M, Jorde L (2001) Genetic evidence on the origins of Indian caste populations. Genome Res 11:994–1004

Barnard-Davis J (1867) Thesaurus craniorum: catalogue of the skulls of the various races of man, in the collection of Joseph Barnard-Davis. Printed for the subscribers, London

Cavalli-Sforza L, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, NJ

Coon CS (1966) The living races of man. Cape, London

Cooper A, Poinar H (2000) Ancient DNA: do it right or not at all. Science 289:1139

Cooper Z (1988) Analysis of the contacts with the Andaman Islands during the last two millennia. South Asian Studies 5:133–147

——— (1993) The origins of the Andaman Islanders: local myth and archaeological evidence. Antiquity 67:394–399

Distell TR (1999) The southern route to Asia. Curr Biol 9: R929–R928

Dobson GE (1875) On the Andamans and the Andamanese. J R Anthropol Soc VI, London

Ericksen P, Beckerman S (1975) Population determinants in the Andaman Islands. Mankind 10:105–107

Finnalä S, Lehtonen M, Majamaa K (2001) Phylogenetic network for European mtDNA. Am J Hum Genet 68:1475–1484

Forbes HG (1884) On the ethnology of Timor Laut. J R Anthropol Soc XIII, London

Gilbert MTP, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, Lynnerup N, Cooper A (2003*a*) Characterization of genetic miscoding lesions caused by postmortem damage. Am J Hum Genet 71:48–61 (in this issue)

Gilbert MTP, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, Lynnerup N, Cooper A (2003*b*) Distribution patterns of postmortem damage in human mitochondrial DNA. Am J Hum Genet 71:32–47 (in this issue)

Greenberg JH (1971) The Indo-Pacific hypothesis. In: Sebeok TA (ed) Current trends in linguistics, vol. 8: Linguistics in Oceania. Mouton, The Hague and Paris, pp 807–871

Handt O, Krings M, Ward R, Pääbo S (1996) The retrieval of ancient human DNA sequences. Am J Hum Genet 59:368–376

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins W, Dixon M, Papiha S, Mastana S, Mir M, Ferak V, Villems R (1999*a*) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. Curr Biol 9:1331–1334

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Parik J, Papiha S, Villems R (1999*b*) The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the old world. In: Papiha SS, Deka R, Chakraborty R (eds) Genomic diversity: applications in human population genetics. Kluwer/Plenum, New York

Macaulay V, Richards M (2000) Genetic data and the colonisation of Europe: genealogies and founders. In: Renfrew C, Boyle K (eds) Archaeogenetics DNA and the population prehistory of Europe. MacDonald Institute for Archaeological Research, Cambridge, United Kingdom

Majumder PP (2001) Ethnic populations of India as seen from an evolutionary perspective. J Biosci 26:533–545

Man EH (1932) The Andaman Islanders. Royal Anthropological Institute of Great Britain and Ireland, London (originally published 1885)

Molnar S (1983) Human variation, races, types, and ethnic groups. Prentice-Hall, Englewood Cliffs, NJ

Mountain J, Herbert J, Bhattacharyya S, Underhill P, Ottolenghi C, Gadgil M, Cavalli-Sforza L (1995) Demographic history of India and mtDNA sequence diversity. Am J Hum Genet 56:979–992

Pääbo S, Irwin D, Wilson A (1990) DNA damage promotes jumping between templates during enzymatic amplification. J Biol Chem 265:4718–4721

Passrino G, Semino O, Bernini LF, Santachiara-Benerecetti AS (1996) Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. Am J Hum Genet 59:927–934

Portman MV (1990) A history of our relations with the Andamanese, compiled from the records of the government of India. Asian Educational Services, New Delhi and Madras (originally published 1899)

Prasad B, Ricker C, Watkins W, Dixon M, Rao B, Naidu J, Jorde L, Bamshad M (2001) Mitochondrial DNA variation in Nicobarese Islanders. Hum Biol 73:715–725

Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benerecetti A (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. Nat Genet 23:437–441

Radcliffe-Brown AR (1914) Notes on the languages of the Andaman Islanders. Anthropos 9:36–52

Richards M, Macaulay V (2001) The mitochondrial gene tree comes of age. Am J Hum Genet 68:1315–1320

Roychoudhury S, Roy S, Basu A, Banerjee R, Vishwanathan H, Rani M, Sil S, Mitra M, Majumder P (2001) Genomic structures and population histories of linguistically distinct tribal groups of India. Hum Genet 109:339–350

Ruhlen M (1987) A guide to the world's languages. Stanford University Press, Stanford, CA

Tavaré S (1984) Line-of-descent and genealogical processes. Theor Popul Biol 26:119–164

Temple R (1903) Census of India 1901, vol 3. Government of India Central Publications, Calcutta

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson A (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507

Wallace AR (1883) Australasia. Stanford, London

Wollaston AFR (1912) Pygmies and Papuans, stone-age today in Dutch New Guinea. Smith, Elder and Co, London

Wurm SA (1971) Sebeok TA (ed) Current trends in linguistics, vol. 8: Linguistics in Oceania. Mouton, The Hague and Paris, pp 807–871

Wurm SA, McElhanan K (1975) New Guinea area languages and language study, vol 1: Papuan languages and the New Guinea linguistic scene. Research School of Pacific and Asian Studies, Australian National University, Canberra