
Volume 85

Issue 1 *Special Issue on Revisiting the "Negrito"
Hypothesis*

Article 7

2013

The Andaman Islanders in a Regional Genetic Context: Reexamining the Evidence for an Early Peopling of the Archipelago from South Asia

Gyaneshwer Chaubey

Estonian Biocenter, Tartu, Estonia

Phillip Endicott

Musée de l'Homme, Paris, France, endicott@mnhn.fr

Follow this and additional works at: <http://digitalcommons.wayne.edu/humbiol>



Part of the [Biological and Physical Anthropology Commons](#), and the [Genetics and Genomics Commons](#)

Recommended Citation

Chaubey, Gyaneshwer and Endicott, Phillip (2013) "The Andaman Islanders in a Regional Genetic Context: Reexamining the Evidence for an Early Peopling of the Archipelago from South Asia," *Human Biology*: Vol. 85: Iss. 1, Article 7.
Available at: <http://digitalcommons.wayne.edu/humbiol/vol85/iss1/7>

The Andaman Islanders in a Regional Genetic Context: Reexamining the Evidence for an Early Peopling of the Archipelago from South Asia

Abstract

The indigenous inhabitants of the Andaman Islands were considered by many early anthropologists to be pristine examples of a "negrito" substrate of humanity that existed throughout Southeast Asia. Despite over 150 years of research and study, questions over the extent of shared ancestry between Andaman Islanders and other small-bodied, gracile, dark-skinned populations throughout the region are still unresolved. This shared phenotype could be a product of shared history, evolutionary convergence, or a mixture of both. Recent population genetic studies have tended to emphasize long-term physical isolation of the Andaman Islanders and an affinity to ancestral populations of South Asia. We reexamine the genetic evidence from genome-wide autosomal single-nucleotide polymorphism (SNP) data for a shared history between the tribes of Little Andaman (Onge) and Great Andaman, and between these two groups and the rest of South and Southeast Asia (both negrito and non-negrito groups).

Keywords

Southeast Asia, South Asia, Negrito, Andamanese, Autosomes

The Andaman Islanders in a Regional Genetic Context: Reexamining the Evidence for an Early Peopling of the Archipelago from South Asia

GYANESHWER CHAUBEY¹ AND PHILLIP ENDICOTT^{2*}

Abstract The indigenous inhabitants of the Andaman Islands were considered by many early anthropologists to be pristine examples of a “negrito” substrate of humanity that existed throughout Southeast Asia. Despite over 150 years of research and study, questions over the extent of shared ancestry between Andaman Islanders and other small-bodied, gracile, dark-skinned populations throughout the region are still unresolved. This shared phenotype could be a product of shared history, evolutionary convergence, or a mixture of both. Recent population genetic studies have tended to emphasize long-term physical isolation of the Andaman Islanders and an affinity to ancestral populations of South Asia. We reexamine the genetic evidence from genome-wide autosomal single-nucleotide polymorphism (SNP) data for a shared history between the tribes of Little Andaman (Onge) and Great Andaman, and between these two groups and the rest of South and Southeast Asia (both negrito and non-negrito groups).

Situated on the eastern edge of the Indian Ocean, formed from part of a chain of submerged mountains stretching from Myanmar to Sumatra, the Andaman archipelago is the most western part of Island Southeast Asia (ISEA) but falls under the administration of India due to its colonial history. Although the islands formed one landmass at the time of the last glacial maximum, they were not part of the Sunda Shelf, and it is likely that human settlement always required a sea crossing of some kind. The inhabitants of the archipelago are perhaps the most enigmatic indigenous people in Southeast Asia, and their origin has been a subject of speculation since they first entered European consciousness after the permanent colonization of the islands by the British in 1858.

Early attempts to make sense of human physical differences by classifying people into various “races” placed the Andaman Islanders into a group called

¹Estonian Biocenter, Riia 23b, Tartu, 51010, Estonia.

²Musée de l’Homme, 61 rue Buffon, 75005, Paris, France.

*Correspondence to: Phillip Endicott, Musée de l’Homme, 61 rue Buffon, 75005, Paris, France. E-mail: endicott@mnhn.fr.

Human Biology, February–June 2013, v. 85, no. 1–3, pp. 153–172.

Copyright © 2013 Wayne State University Press, Detroit, Michigan 48201-1309

“negritos.” This term is from the Spanish diminutive for black and was first used to describe Philippine groups of a visually similar phenotype, characterized by short stature, dark skin pigmentation, and tight curly hair. People answering this description were found, usually as populations of mobile resource procurers, across mainland Southeast Asia and ISEA, and envisaged to be relict populations of an early substrate of humanity, pushed into marginal environments by the encroachment of agriculturalists since the onset of the Neolithic (Quatrefages 1895; Radcliffe-Brown 1922).

Whether there is any basis for grouping these peoples together by this limited definition of phenotype is investigated elsewhere in the issue (see Benjamin, Bulbeck, Migliano et al., Stock), but as defined by stature, hair morphology, and pigmentation, negritos are today found in the Andaman Islands, the Philippines, Malaysia, and Thailand. Among these, the groups from the Andaman archipelago are unique in retaining languages that predate the expansion and adoption of Austronesian in the Philippines (Reid this issue), and Austroasiatic (Aslian) in Malaysia (Dunn this issue) during the mid to late Holocene. These linguistic differences, due to recent language shift, render it difficult to detect whether these regional populations once shared a common cultural heritage. However, Blust (this issue) suggests that some clues linking Malaysian and Philippine negritos—and possibly Andamanese—are retained in various versions of the “Thunder-God” cultural complex.

The simple dichotomy between Andamanese languages and the rest of Southeast Asia is an oversimplification because the archipelago was itself divided linguistically, with the tribes of Great Andaman speaking languages that bore very little resemblance to the Onge-Jarawa group of South Andaman and southern Great Andaman (Portman 1884). Nevertheless, the preservation of linguistic isolates in the Andamans, combined with a long history of resistance to outside contact (Cooper 1989), led many nineteenth-century observers to speculate that the Andamanese had been cut off in their island home since prehistoric times (Quatrefages 1895; Radcliffe-Brown 1922). It followed, therefore, that they might represent the “negrito” race in its pristine state and perhaps hold important clues to human ancestry in general (for a critical read, see Bulbeck this issue). Although there was no concept of a human origin in Africa at this time, the possible sharing of some aspects of phenotype with African pygmies inevitably led to speculation of an African origin (Radcliffe-Brown 1922). Today, it is widely believed that all humans originated in Africa, so the discussion has turned to the timing of the first settlement of the Andaman archipelago and how closely it is linked to the arrival of humans in South and Southeast Asia.

The archaeological record of the Andaman Islands, however, is scant and does not currently extend beyond the first millennium BC (Cooper 2002). Some of this lacuna could possibly be related to the dramatic rises in sea level experienced during the terminal Pleistocene and early Holocene, which also drowned Sundaland and wiped out the archaeological record for a key area the size of the India (Higham this issue). On the other hand, if the archaeology is taken at face value, it

is conceivable that the ancestors of the Andamanese arrived as recently as the late Holocene, perhaps as resource procurers for a vibrant regional trade network (Morrison 2007). Certainly, the presence of ceramics and pigs in the Andamans bears testament to contact with the outside world during the last millennia (Bulbeck this issue). Nevertheless, the nineteenth-century popular notion of long-term isolation, a stone-age people abroad in the present, is still predominant among scientific and popular writers alike.

Early mitochondrial DNA (mtDNA) studies of the Andamanese (for a classical genetic markers review, see Stock this issue), using both museum skeletal material (Endicott et al. 2003) and modern populations (Endicott et al. 2003; Thangaraj et al. 2003), stressed an affinity with Asian rather than African populations. Subsequent high-resolution research, using complete mtDNA genomes, identified two apparently Andaman-specific mtDNA haplogroups, M31 and M32, and interpreted them as evidence for a single rapid dispersal of humans along the coast of the Indian Ocean during the late Pleistocene ~60 ka (Thangaraj et al. 2005). The bases for this claim were that these haplogroups were found only in the Andamans and that their age was effectively that of macro-haplogroup M, which is presumed to have arisen sometime soon after the exodus from Africa, timed by one version of the molecular clock at ~65 ka (Macaulay et al. 2005; Forster and Matsumura 2005). Therefore, the promulgation of a deep chronology in the molecular age has, to a great extent, been linked to the use of phylogeography and phylogenetic dating of mtDNA, attempting to trace the pioneering settlement of humans from Africa to Australia (Forster and Matsumura 2005; Macaulay et al. 2005; Thangaraj et al. 2005).

At the time of the publication of this 65 ka chronology, however, there was a complete absence of comparative data from neighboring regions of both mainland Southeast Asia and ISEA, and so a more recent settlement of the Andaman archipelago from Myanmar was an equally parsimonious explanation. The subsequent sequencing of another branch of haplogroup M31 (M31b) in mainland India was used to argue for a later date of settlement (Palanichamy et al. 2006), but the branching between the two clades could still accommodate a very early arrival in the Andamans ~50 ka (Thangaraj et al. 2006). An increased number of complete mtDNA sequences refined the phylogeography and phylogeny of M31 and identified a sister clade (M31a2) to Andaman-specific M31a1 in East India (Barik et al. 2008). A third branch (M31c) was subsequently identified among populations of Northeast India (Wang et al. 2011; Reddy et al. 2007; Fornarino et al. 2009). The discovery of M31a2 in East India was particularly important because the separation time between this clade and its sister, M31a1, at ~25 ka (Barik et al. 2008) is half that previously suggested for the age of M31 overall (Thangaraj et al. 2005). If M31a1 was a founding lineage, this provides an upper limit for the settlement of the Andaman archipelago, provided that the separation of the two M31a clades occurred on the mainland.

So far, there is no strong evidence for mtDNA haplogroup M32 in South or Southeast Asia (Chandrasekar et al. 2009; Wang et al. 2011), but there is a

possible link to a lineage found in Madagascar (Dubut et al. 2009; see Phylotree.org), which was settled by Austronesian speakers from ISEA (Hurles et al. 2005). A third minor-frequency mtDNA haplogroup (R22) found among the surviving Great Andaman population also appears to have originated in Southeast Asia (Hill et al. 2007), and the Andaman lineage appears to be specific to the archipelago (our unpublished data). The chronology of a settlement after 25 ka aligns very well with a proposed expansion of other mtDNA lineages within ISEA 30–10 ka (Gunnarsdóttir et al. 2011; Jinam et al. 2012; Guillot this issue), prior to the expansion of the Austronesian and Austroasiatic language families (~4–7 ka) (Gray et al. 2009; Dunn this issue) and, significantly, after the first archaeological evidence for human settlement of the region ~45 ka (Demeter et al. 2012).

A recent genetic study, using data from ~500,000 autosomal SNPs, investigated the various predicted ancestral components of the genomes of South Asian populations and compared them with the Onge of Little Andaman (Reich et al. 2009). The data available for the surviving population of Great Andamanese were not considered because of assumed recent admixture with South Asians. The results suggested two main ancestral components within South Asia, which differentiated along a north-south axis, similar to trends previously observed within mtDNA and Y chromosome haplogroup data (Metspalu et al. 2004; Sahoo et al. 2006). The Onge were interpreted as having exclusively ancestral South Asian ancestry (Reich et al. 2009), thereby sustaining the hypothesis of an early human migration from South Asia to the Andamans, followed by long-term isolation (Thangaraj et al. 2005).

The basis for omitting the Great Andamanese from the analysis, however, is not clear, because most of the Y chromosome haplogroups claimed as evidence for recent South Asian admixture (O2, O3) (Thangaraj et al. 2003; Reich et al. 2009) are of unambiguous East and Southeast Asian origin (Shi et al. 2005). Moreover, the decision to omit many available autosomal data (HGDP-CEPH panel) from other Southeast and East Asian populations, as well as Austroasiatic (Munda) speakers from South Asia, results in a reliance upon HapMap CHB (Han Chinese in Beijing) for comparative samples. This means that the Onge must either fall with the ancestors of the single Han Chinese population, collected from Beijing, or derive from an Asian (“Indian”) ancestral group. While it is potentially informative that the Onge cluster with the south of India rather than the north, this provides no insight into potential evolutionary relationships with Southeast Asian populations.

Since the samples used for these autosomal SNP analyses may not adequately represent the genomic diversity of the geographic regions involved, we decided to reexamine the genetic evidence for a shared history between the Andaman negritos (Onge and Great Andamanese) and the rest of South and Southeast Asia (both negrito and non-negrito groups). Given the overall absence of shared mtDNA and Y chromosomal haplogroups among different phenotypically assigned negrito groups (Thangaraj et al. 2003; Macaulay et al. 2005; Delfin et al. 2011; HPASC 2009; Heyer et al. and Jinam et al. this issue), we included data from the largest genetic survey so far undertaken within these two regions together with both

the extant Great Andamanese and Onge, to facilitate a closer examination of the ancestral relationship between different negrito populations (see Table 1 for details of populations).

Materials and Methods

The study was performed using control samples collected and genotyped for population studies (HapMap 2010; Reich et al. 2009; HPASC 2009); no genotyping was performed specifically for this study (Table 1). A check for closely related individuals was carried out within each study population by calculating average IBS (identity by state) scores for all pairs of individuals (Purcell et al. 2007). First, we sought to investigate the extent of population structure and admixture among the Indian and Southeast Asian Austroasiatic speakers embedded in their autosomal genomes. After excluding SNPs unique to either of the three platforms and SNPs from mtDNA and X and Y chromosomes, the combined data set had data for 12,622 SNPs, which were used in the subsequent analyses.

We used PLINK 1.07 (Purcell et al. 2007) to filter the combined data set to include only SNPs on the 22 autosomal chromosomes with minor allele frequency >1% and genotyping success >99%. Because background linkage disequilibrium (LD) can affect both principal component analysis (PCA) (Patterson et al. 2006) and structure-like analysis (Alexander et al. 2009), we thinned the data set by removing one SNP of any pair in strong LD ($r^2 > 0.4$) in a window of 200 SNPs (sliding the window by 25 SNPs at a time). Finally, we were left with a data matrix of 1102 individuals by 8966 SNPs. For PCA, we generated an additional data set with the same filters but excluding the African and European samples, yielding a matrix of 1042 samples by 8966 SNPs.

We carried out PCA using the smartpca program (with default settings) of the EIGENSOFT package (Patterson et al. 2006) to capture genetic variability described by the first 10 principal components (PCs). The fraction of total variation described by a PC is the ratio of its eigenvalue to the sum of all eigenvalues. We also performed PCA with the whole data set (Figure available on request).

In the final setting we ran ADMIXTURE (Alexander et al. 2009), with a random seed number generator, on the LD-pruned data set 25 times at $K = 2$ to $K = 12$. Because the top values of the resulting log-likelihood scores were stable (virtually identical) within the runs of each K from $K = 2$ to $K = 9$, we can with some confidence argue that convergence at global maximum was reached. Thus, we omitted runs at $K = 10$ to $K = 12$ from further analysis. To see the robustness of the reduced number of SNPs used for the analysis, we have filtered the top 100 SNPs in all three sets showing the highest value of population differentiation (F_{ST}) between Indian Dravidian and CHB populations. More than 75% of the top SNPs were present in the merged data set, arguing against ascertainment bias with a lesser number of SNPs.

MEGA 5.0 (Tamura et al. 2007) was used to construct the neighbor-joining tree generated on the basis of F_{ST} values calculated by an algorithm provided

Table 1. Details of the Populations Used in the Present Study

POPULATION	<i>N</i>	LANGUAGE	REGION	COUNTRY	REFERENCE
Agta	8	Austronesian	Southeast Asia	Philippines	HPASC 2009
Alorese	19	Austronesian	Southeast Asia	Indonesia	HPASC 2009
Ami	10	Austronesian	Southeast Asia	Taiwan	HPASC 2009
Atayal	10	Austronesian	Southeast Asia	Taiwan	HPASC 2009
Ati	23	Austronesian	Southeast Asia	Philippines	HPASC 2009
Ayta	8	Austronesian	Southeast Asia	Philippines	HPASC 2009
Batak Karo	17	Austronesian	Southeast Asia	Indonesia	HPASC 2009
Batak Toba	20	Austronesian	Southeast Asia	Indonesia	HPASC 2009
Bidayuh Jagoi	50	Austronesian	Southeast Asia	Malaysia	HPASC 2009
Chenchu	6	Dravidian	South Asia	India	Reich et al. 2009
European Utah	30	Indo-European	Europe	Europe	HapMap 2009
Great Andamanese	7	Andamanese	South Asia	India	Reich et al. 2009
Gujarati Houston	30	Indo-European	South Asia	India	HapMap 2009
Hallaki	7	Dravidian	South Asia	India	Reich et al. 2009
Han Beijing	30	Sino-Tibetan	East Asia	China	HapMap 2009
Hindi	53	Indo-European	South Asia	India	HPASC 2009
Hmong	26	Hmong-Mien	East Asia	China	HPASC 2009
Hmong Miao	20	Hmong-Mien	East Asia	China	HPASC 2009
Htin Mal	18	Austroasiatic	Southeast Asia	Thailand	HPASC 2009
Iraya	9	Austronesian	Southeast Asia	Philippines	HPASC 2009
Javanese	30	Austronesian	Southeast Asia	Indonesia	HPASC 2009
Kharia	6	Austroasiatic	South Asia	India	Reich et al. 2009
Kurumba	9	Dravidian	South Asia	India	Reich et al. 2009
Lawa	19	Austroasiatic	Southeast Asia	Thailand	HPASC 2009
Lembata	19	Austronesian	Southeast Asia	Indonesia	HPASC 2009
Lodi	5	Indo-European	South Asia	India	Reich et al. 2009
Madiga	4	Dravidian	South Asia	India	Reich et al. 2009
Mala	3	Dravidian	South Asia	India	Reich et al. 2009

elsewhere (Cockerham and Weir 1984). The Plink software (Purcell et al. 2007) was used to find the 100 nearest neighbors for the Onge and Great Andamanese individuals.

Results and Discussion

The relationship of the Great Andamanese and Onge with other negrito and non-negrito populations was initially investigated by calculating the average number of pairwise differences and F_{ST} statistics for each pair of populations using more than 12,000 of the markers (Figure 1). The population structure was then investigated in more detail using PCA (Patterson et al. 2006) and ADMIXTURE software (Alexander et al. 2009) on the individual samples (Figures 2 and 3). With regard to F_{ST} (genetic distance), the Onge are an outlier; together with Chenchu, Mamanwa, Ayta, Iraya, and Melanians, they display the highest values relative to other populations. The Great Andamanese, however, show a closer affinity with

POPULATION	N	LANGUAGE	REGION	COUNTRY	REFERENCE
Malay	50	Austronesian	Southeast Asia	Malaysia	HPASC 2009
Malay Singapore	30	Austronesian	Southeast Asia	Malaysia	HPASC 2009
Mamanwa	19	Austronesian	Southeast Asia	Philippines	HPASC 2009
Manggarai	36	Austronesian	Southeast Asia	Indonesia	HPASC 2009
Marathi	14	Indo-European	South Asia	India	HPASC 2009
Melanesians Nasioi	5	Austronesian	Melanesia	Melanesia	HPASC 2009
Mentawai	15	Austronesian	Southeast Asia	Indonesia	HPASC 2009
Minanubu Manobo	18	Austronesian	Southeast Asia	Philippines	HPASC 2009
Mon	19	Austroasiatic	Southeast Asia	Thailand	HPASC 2009
Naidu	4	Dravidian	South Asia	India	Reich et al. 2009
Negrito Jehai	50	Austroasiatic	Southeast Asia	Malaysia	HPASC 2009
Negrito Kensiu	30	Austroasiatic	Southeast Asia	Malaysia	HPASC 2009
Okinawan	49	Altaic	East Asia	Japan	HPASC 2009
Onge	9	Andamanese	South Asia	India	Reich et al. 2009
Paluang	18	Austroasiatic	Southeast Asia	Thailand	HPASC 2009
Sahariya	4	Indo-European	South Asia	India	Reich et al. 2009
Santhal	7	Austroasiatic	South Asia	India	Reich et al. 2009
Satnami	4	Indo-European	South Asia	India	Reich et al. 2009
Zhuang Nong	26	Tai-Kadai	East Asia	China	HPASC 2009
Telugu Kannada	24	Dravidian	South Asia	India	HPASC 2009
Temuan	30	Austronesian	Southeast Asia	Malaysia	HPASC 2009
Toraja	20	Austronesian	Southeast Asia	Indonesia	HPASC 2009
Uyghur	26	Altaic	East Asia	China	HPASC 2009
Wa	56	Austroasiatic	East Asia	China	HPASC 2009
Vaish	4	Indo-European	South Asia	India	Reich et al. 2009
Velama	4	Dravidian	South Asia	India	Reich et al. 2009
Vysya	5	Dravidian	South Asia	India	Reich et al. 2009
Yoruba Nigeria	30	Afro-Asiatic	Africa	Africa	HapMap 2009

both South and Southeast Asian populations (Figure 1). The Onge and Melanesian measures of F_{ST} display the least within-population difference, which is consistent with long-term isolation and the effects of genetic drift, characteristics that are in line with their lower levels of heterozygosity (Figures 1 and 4).

The allele-sharing distance (ASD) of the Onge with respect to other populations is high, consistent with the F_{ST} and pairwise difference values. The closest population to the Onge is the Great Andamanese, but when the latter group is excluded, the Onge are clearly closer to Malaysian negritos, with a value of 0.55911, comparing with an average value for the region of 0.65355. This finding suggests that the Great Andamanese, Onge, and Malaysian negritos may have a degree of shared ancestry but that genetic drift and admixture have caused differentiation among the populations.

A PCA using all samples clearly separates Africans from the rest of the world along both axes (figure available on request). However, our main geographic focus is more constrained, and leaving out both Africans and Europeans reveals

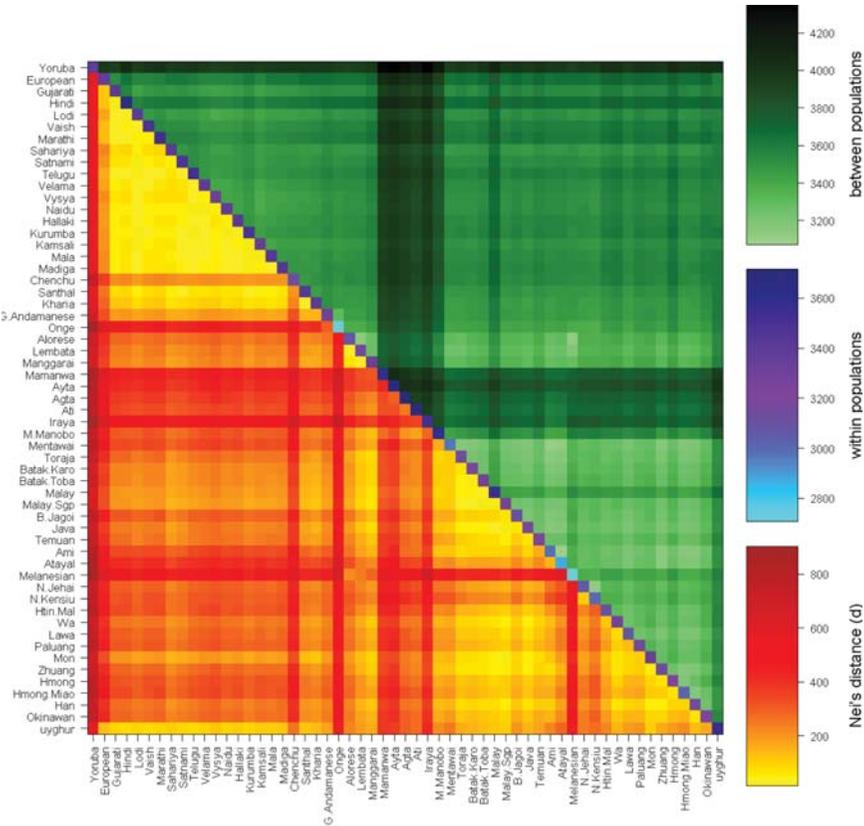
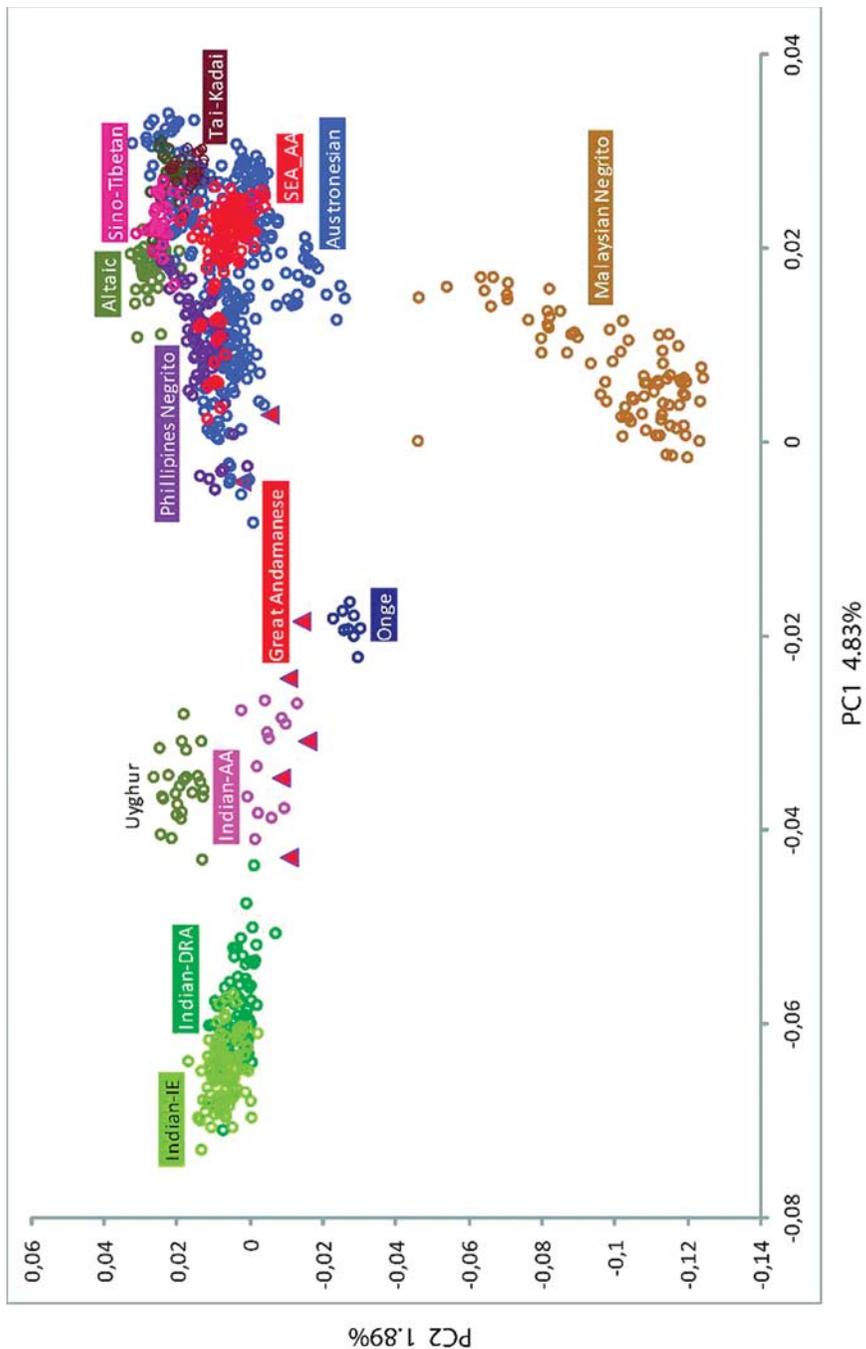


Figure 1. Heat map, constructed by using Arlequin 3.5 (Excoffier and Lischer 2010), showing average numbers of pairwise differences between populations (upper right triangular matrix), within populations (along the central axis), and for Nei's genetic distance (lower left triangular matrix). Table 1 lists the populations included.

clearly defined groupings within the regional study area, consistent with the pattern observed in the PCA of Jinam et al. (this issue). PC1 separates South Asians from both East and Southeast Asians, while PC2 differentiates both the Onge and Malaysian negritos from the other clusters (Figure 2). This grouping, together with the separation from Great Andamanese individuals, is consistent with the F_{ST}/ASD results (Figures 1 and 2). Examining the latter in greater detail, two of the Great Andamanese stretch out along the South Asian-Onge cline, three cluster loosely

Figure 2 (opposite). Plot of PC1 versus PC2 for Asian populations. Colors indicate linguistic/ethnic groupings. AA, Austroasiatic; DRA, Dravidian; IE, Indo-European; SEA, Southeast Asian.



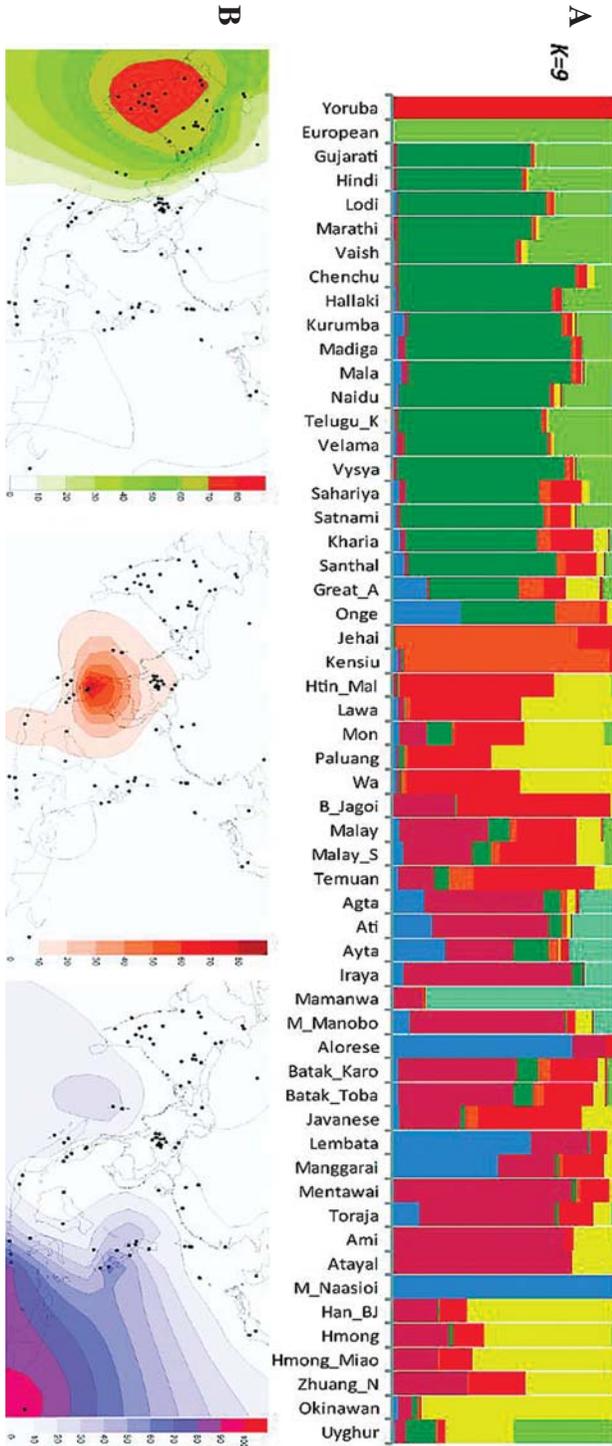


Figure 3. (a) Average individual admixture proportions estimated from >12,000 SNPs. The predefined populations are listed at the bottom of the plot. Each color refers to one of the nine putative ancestral components; the three main ancestral components among both the Great Andamanese and Onge are Melanesian (dark blue), South Asian (dark green), and Malaysian (orange). Table 1 gives a complete list of populations included. (b) Spatial geographic distribution of the three main ancestral components of the Andamanese.

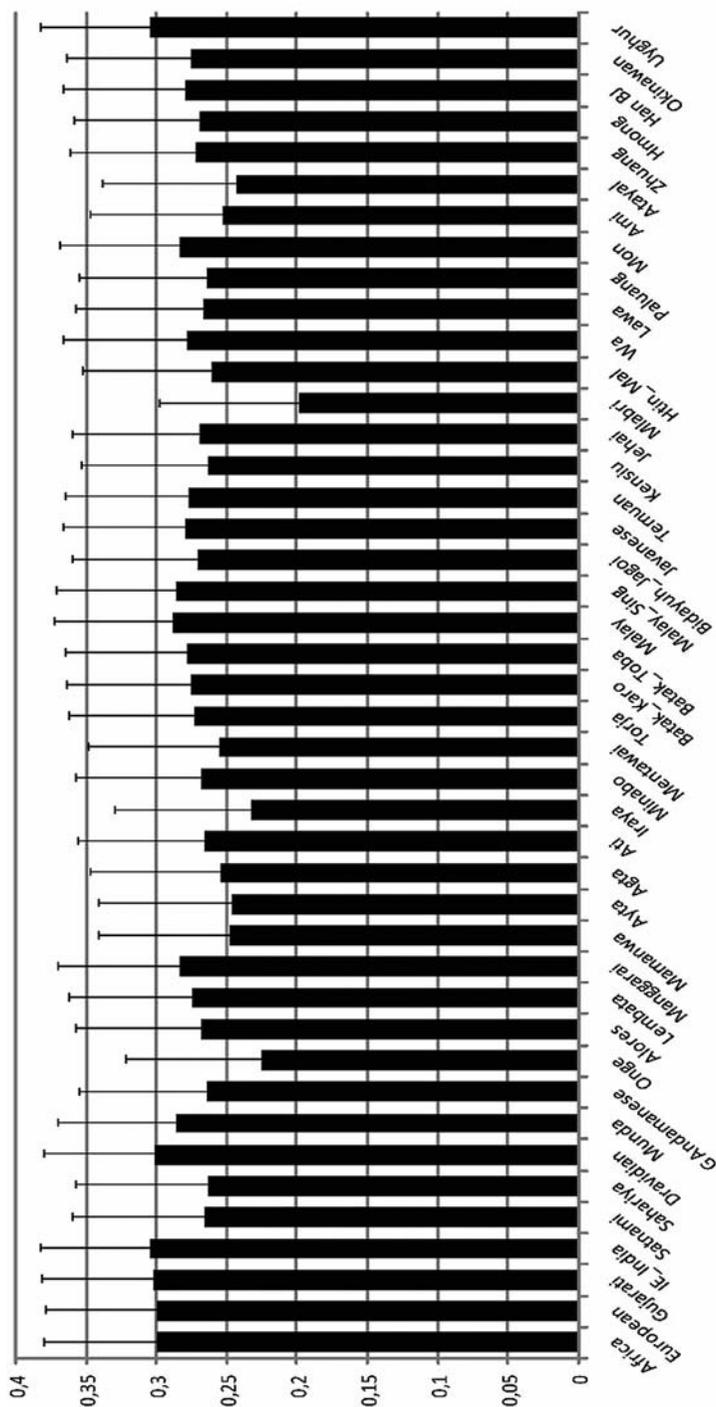


Figure 4. Heterozygosity values (y-axis) calculated on the basis of >12,000 SNPs for different population groups. IE, Indo-European.

Table 2. Summary of Individual Admixture Proportions of Andaman Negrito Populations in Onge and Great Andamanese with the Hypothetical Ancestral Populations

	MELANESIAN	INDIAN	MALAYSIAN NEGrito	PHILIPPINES NEGrito	SOUTHEAST ASIAN AUSTROROASIATIC	EAST ASIAN ASIAN	AUSTRONESIAN	NORTH INDIAN AND EUROPEAN ANCESTRY	AFRICAN
<i>Onge</i>									
	0.318427	0.4443	0.20724	0.00001	0.029983	0.00001	0.00001	0.00001	1E-05
	0.322275	0.3902	0.20462	0.00001	0.028694	0.042662	0.00001	0.00001	0.0116
	0.297889	0.4144	0.20762	0.007552	0.062419	0.010106	0.00001	0.00001	1E-05
	0.25063	0.4222	0.18471	0.00001	0.076811	0.053059	0.00001	0.00001	0.0125
	0.316066	0.4208	0.16312	0.00001	0.064605	0.035322	0.00001	0.00001	1E-05
	0.305727	0.376	0.20379	0.026784	0.06092	0.012278	0.00001	0.00001	0.0145
	0.30224	0.3988	0.18227	0.00001	0.068783	0.044905	0.00001	0.00001	0.003
	0.289123	0.3947	0.17761	0.025729	0.009012	0.09615	0.00001	0.00001	0.0076
	0.310275	0.3867	0.19229	0.00001	0.070074	0.040675	0.00001	0.00001	1E-05
<i>Great Andamanese</i>									
	0.138312	0.4934	0.12294	0.018868	0.00001	0.062605	0.00001	0.1541	0.0097
	0.102409	0.5632	0.10874	0.020726	0.053533	0.095538	0.004108	0.04877	0.003
	0.222819	0.4147	0.14342	0.016935	0.027123	0.071856	0.00001	0.08847	0.0147
	0.190077	0.4449	0.11621	0.007697	0.065582	0.121376	0.00001	0.03017	0.024
	0.227188	0.3873	0.15386	0.021138	0.044721	0.091985	0.052271	0.00001	0.0216
	0.091408	0.2836	0.04921	0.019252	0.210198	0.332886	0.001094	0.01233	1E-05
	0.14112	0.1923	0.09133	0.00001	0.279188	0.296055	0.00001	0.00001	1E-05

Boldface indicates values >10%.

with Austroasiatic populations of India, and the rest are in closer proximity with East and Southeast Asians.

While recent male gene flow into the Great Andamanese from mainland India (Reich et al. 2009) could be the reason for their clustering together with Indian Austroasiatic populations, these groups themselves have some Southeast Asian ancestry (Chaubey et al. 2011). Moreover, the Great Andamanese individuals clustering with Austronesians have East Asian Y chromosomal haplogroup O3a (Thangaraj et al. 2003), suggesting a paternal ancestry within Southeast Asia (Shi et al. 2005). So, while the results for the Onge are compatible with genetic drift and no recent admixture, their Great Andamanese neighbors appear to have received recent gene flow from populations within both the current Austroasiatic and Austronesian linguistic spheres of influence. The current Austroasiatic-speaking sphere includes Bengal, which was a source of prisoners incarcerated on Great Andaman by the nineteenth-century British administration in response to the first Indian war of independence (a.k.a. the Sepoy Mutiny). Some of these prisoners were helped to escape by the Great Andamanese during the early days of colonial occupation (Temple 1903), and these individuals are a potential source for some components of the current Great Andamanese genetic diversity.

The results from ADMIXTURE (Alexander et al. 2009), which assigns blocks of individual genomes proportionally to hypothetical ancestral populations, indicate three major components contributing to the Onge. The same three components also make up the greater part of the genomes of the Great Andamanese (Figure 3a and Table 2) and are assigned to hypothetical populations ancestral to the Melanesian, South Asian, and Malaysian (negrito) groups. Our analysis is based on a subset of SNPs used in a previous study of Asia (HPASC 2009), in order to be able to compare with data from the Onge and Great Andamanese (Reich et al. 2009). While caution should be employed when lowering the density of SNPs for any particular haplotype block in this way, our results are entirely consistent with those of the HUGO Pan-Asian SNP Consortium analysis (HPASC 2009). They are also supported by other SNP studies presented in this issue (Migliano et al. and Jinam et al. this issue) for all common populations, including the Malaysian and Philippine negritos.

Sharing of an ancestral component by different populations can imply either a shared genetic ancestry or recent admixture of sampled individuals. A closer look at the ancestry of the Onge and Great Andamanese individuals indicates a contrasting pattern, in agreement with the PCA (Figure 2 and Table 2). Individualwise component sharing among Great Andamanese individuals is highly variable compared with that of Onge individuals (Table 2). All Onge individuals have a similar proportion of components, consistent with a deep shared ancestry based on a founder effect followed by a period of genetic drift. This interpretation is supported by the heterozygosity values, F_{ST} , and pairwise differences between individuals (Figures 1 and 4, and Table 2). The Great Andamanese, while displaying the same three ancestry components, have two individuals with significant parts of their genome assigned to Southeast and East Asian sources, and three individuals displaying minor components compatible with a South Asian source. This sharing

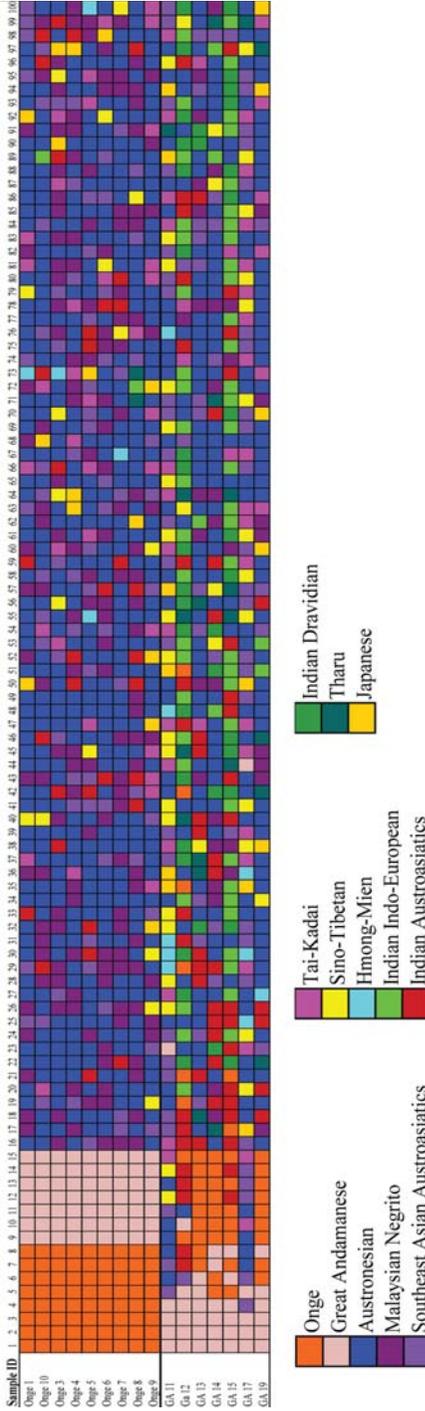


Figure 5. Plot of 100 nearest-neighbors with respect to the Ongc and Great Andamanese individuals. Different linguistic regions are indicated by different colors/patterns. This plot is also consistent with the close ancestry of Great Andamanese with regard to the Ongc and recent admixture from regionally proximal sources.

is likely responsible for the attraction of the Great Andamanese toward South Asian groups in the population-level statistical analyses.

Next, we calculated the 100 closest neighbors for the two Andaman negrito populations, on an individual basis (Figure 5). As reflected in the F_{ST} values in Figure 1, the Onge and Great Andamanese are their own closest neighbors. After excluding these comparisons, the closest neighbors of the Onge individuals were overwhelmingly Southeast Asians, rather than South Asians. The Great Andamanese individuals also have similar results in the PCA and ADMIXTURE analyses, showing some individuals inclined toward Indian Austroasiatic populations (GA12 and GA15) and others closer to Southeast Asians (GA11 and GA17). This confirms the recent admixture of some Great Andamanese, which is causing the lack of clustering in the PCA analyses, despite the fact that they are overwhelmingly genetically similar to their geographical neighbors, the Onge of Little Andaman.

Having demonstrated the joint ancestral affinities between the Onge and Great Andamanese, and the distorting effects of admixture among the latter, for the neighbor-joining tree we focus on the Onge as a proxy for the ancestral population of the Andaman archipelago. Indications from the previous analyses that the Onge represent a distinct group are confirmed in the neighbor-joining tree, where they form a deep cluster with Southeast rather than South Asian populations (Figure 6). The tree places all the negrito populations outside of a major cluster containing the majority of Austroasiatic, Hmong Mien, Tai Kadai, and Sino-Tibetan speakers. This clustering places the Onge closer to the negrito populations of Malaysia than to those of the Philippines, consistent with the ADMIXTURE analysis. The weak attraction between Philippine and Malaysian negritos concords with a recent study using a reticulated neighbor-joining tree (Jinam et al. 2012).

The dating of the split within M31a provides an upper limit for the settlement of the Andaman-specific mtDNA lineage M31a1 around 26 ka, while the ages of the diversification within M32 and M31a1 are estimated to fall within the Holocene, using whole-genome data in a Bayesian statistical setting (Barik et al. 2008). Because mtDNA divergence is anticipated to predate population divergence, collectively these estimates suggest that the Andamans were settled less than ~26 ka and that differentiation between the ancestors of the Onge and Great Andamanese commenced in the Terminal Pleistocene. Interestingly, this time frame is similar to the signal for population expansion found throughout ISEA (Guillot et al. this issue) and represents the time of topographic transition from the vast expanses of Sundaland to the submerged Southeast Asian island chains of the Holocene.

In conclusion, we find no support for the settlement of the Andaman Islands by a population descending from the initial out-of-Africa migration of humans, or their immediate descendants in South Asia. It is clear that, overall, the Onge are more closely related to Southeast Asians than they are to present-day South Asians. The similarity in proportions of the Onge genomes, attributed to the Melanesian, Malaysian (Jehai and Kensui), and South Asian ancestral components, combined with evidence for genetic drift, suggests that these constituent parts were present prior to their isolation from other parts of Southeast Asia. In turn, the Great

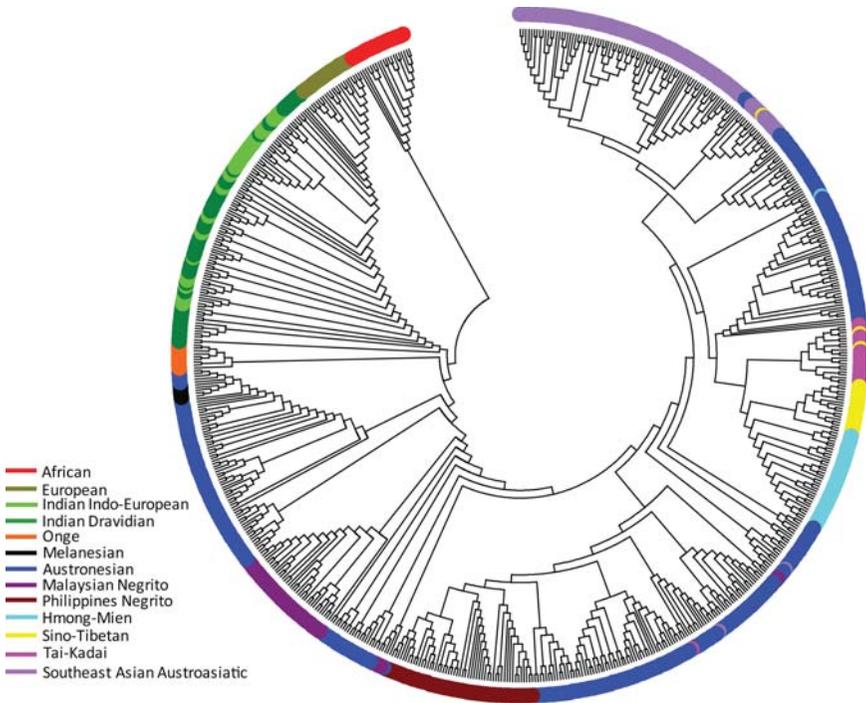


Figure 6. Neighbor-joining tree of 57 populations based on pairwise F_{ST} . Colors indicate linguistic/ethnic groups. The Great Andamanese were omitted from this tree because of their similarity to the Onge, which, combined with their recent admixture, causes them to reposition the Onge in the tree. The neighbor-joining tree is more appropriate for a species-level analysis, but with respect to the Onge it may provide an acceptable delineation from other regional populations due to elevated levels of genetic drift.

Andamanese are their closest genetic neighbors, who appear to have received a degree of relatively recent admixture from adjacent regional populations but also share a significant degree of genetic ancestry with Malaysian negrito groups. These three ancestral components—South Asian, Malaysian negrito, and Melanesian—appear in varying amounts and combinations among other negrito and non-negrito populations across the study area, including some Philippine groups. At the current level of genetic resolution, however, there is no evidence of a single ancestral population for the different groups traditionally defined as “negritos.”

Acknowledgments We thank two of the reviewers for their helpful suggestions. G.C. is supported by the European Union European Regional Development Fund through the Centre of Excellence in Genomics to Estonian Biocentre. Calculations were carried out in the High Performance Computing Center, University of Tartu, Estonia.

Received 15 November 2012; revision accepted for publication 20 March 2013.

Literature Cited

- Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1,655–1,664.
- Barik, S. S., R. Sahani, B. V. Prasad et al. 2008. Detailed mtDNA genotypes permit a reassessment of the settlement and population structure of the Andaman Islands. *Am. J. Phys. Anthropol.* 136:19–27.
- Blust, R. 2013. Terror from the sky: Unconventional linguistic clues to the negrito past. *Hum. Biol.* 85:401–416.
- Bulbeck, D. 2013. Craniodental affinities of Southeast Asia's "negritos" and the concordance with their genetic affinities. *Hum. Biol.* 85:95–134.
- Chandrasekar, A., S. Kumar, J. Sreenath et al. 2009. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: Dispersal of modern human in South Asian corridor. *PLoS ONE* 4:e7447.
- Chaubey, G., M. Metspalu, Y. Choi et al. 2011. Population genetic structure in Indian Austroasiatic speakers: The role of landscape barriers and sex-specific admixture. *Mol. Biol. Evol.* 28:1,013–1,024.
- Cockerham, C. C., and B. S. Weir. 1984. Covariances of relatives stemming from a population undergoing mixed self and random mating. *Biometrics* 40:157–164.
- Cooper, Z. 1989. Analysis of the nature of contacts with the Andaman Islands during the last two millennia. *South Asian Studies* 5:133–147.
- Cooper, Z. 2002. *Archaeology and History: Early Settlements in the Andaman Islands*. New Delhi and Oxford: Oxford University Press.
- Delfin, F., J. M. Salvador, G. C. Calacal et al. 2011. The Y-chromosome landscape of the Philippines: Extensive heterogeneity and varying genetic affinities of negrito and non-negrito groups. *Eur. J. Hum. Genet.* 19:224–230.
- Demeter, F., L. L. Shackelford, A. M. Bacon et al. 2012. Anatomically modern human in Southeast Asia (Laos) by 46 ka. *Proc. Natl. Acad. Sci. USA* 109:14375–14380.
- Dubut, V., F. Cartault, C. Payet et al. 2009. Complete mitochondrial sequences for haplogroups M23 and M46: Insights into the Asian ancestry of the Malagasy population. *Hum. Biol.* 81:495–500.
- Dunn, M., N. Kruspe, and N. Burenhult. 2013. Time and place in the prehistory of the Aslian languages. *Hum. Biol.* 85:383–400.
- Endicott, P., M. T. P. Gilbert, C. Stringer et al. 2003. The genetic origins of the Andaman Islanders. *Am. J. Hum. Genet.* 72:178–184.
- Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10:564–567.
- Fornarino, S., M. Pala, V. Battaglia et al. 2009. Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): A reservoir of genetic variation. *BMC Evol. Biol.* 9:154.
- Forster, P., and S. Matsumura. 2005. Evolution. Did early humans go north or south? *Science* 308:965–966.
- Gray, R. D., A. J. Drummond, and S. J. Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323:479–483.
- Guillot, E. G., M. K. Tumonggor, J. S. Lansing et al. 2013. Climate change influenced female population sizes through time across the Indonesian archipelago. *Hum. Biol.* 85:135–152.
- Gunnarsdóttir, E. D., M. Li, M. Bauchet et al. 2011. High-throughput sequencing of complete human mtDNA genomes from the Philippines. *Genome Res.* 21:1–11.
- HapMap (International HapMap 3 Consortium), D. M. Altshuler, R. A. Gibbs et al. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature* 467:52–58.

- Heyer, E., M. Georges, M. Pachner, and P. Endicott. 2013. Genetic diversity of four Filipino negrito populations from Luzon: Comparison of male and female effective population sizes and differential integration of immigrants in Aeta and Agta communities. *Hum. Biol.* 85:189–208.
- Higham, C. 2013. Hunter-gatherers in Southeast Asia: From prehistory to the present. *Hum. Biol.* 85:21–44.
- Hill, C., P. Soares, M. Mormina et al. 2007. A mitochondrial stratigraphy for island Southeast Asia. *Am. J. Hum. Genet.* 80:29–43.
- HPASC (HUGO Pan-Asian SNP Consortium). 2009. Mapping human genetic diversity in Asia. *Science* 326:1,541–1,545.
- Hurles, M. E., B. C. Sykes, M. A. Jobling et al. 2005. The dual origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from maternal and paternal lineages. *Am. J. Hum. Genet.* 76:894–901.
- Jinam, T. A., L. C. Hong, M. E. Phipps et al. 2012. Evolutionary history of continental Southeast Asians: “Early train” hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. *Mol. Biol. Evol.* 29:3,513–3,527.
- Jinam, T. A., M. E. Phipps, N. Saitou, and The Hugo Pan-Asian SNP Consortium. 2013. Admixture patterns and genetic differentiation in negrito groups from West Malaysia estimated from genome-wide SNP data. *Hum. Biol.* 85:173–188.
- Macaulay, V., C. Hill, A. Achilli et al. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1,034–1,036.
- Metspalu, M., T. Kivisild, E. Metspalu 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.
- Metspalu, M., I. G. Romero, B. Yunusbayev et al. 2011. Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *Am. J. Hum. Genet.* 89:731–744.
- Migliano, A. B., I. G. Romero, M. Mespalu et al. 2013. Evolution of the pygmy phenotype: Evidence of positive selection from genome-wide scans in African, Asian, and Melanesian pygmies. *Hum. Biol.* 85:251–284.
- Morrison, K. 2007. Foragers and forager-traders in South Asian worlds: Some thoughts from the last 10,000 years. In *The Evolution and History of Human Populations in South Asia: Interdisciplinary Studies in Archaeology, Biological Anthropology, Linguistics and Genetics*, M. Petraglia and B. Allchin, eds. Dordrecht: Springer/Kluwer Academic, 321–328.
- Palanichamy, M. G., S. Agrawal, Y. G. Yao et al. 2006. Comment on “Reconstructing the origin of Andaman islanders.” *Science* 311:470.
- Patterson, N., A. L. Price, and D. Reich 2006. Population structure and eigenanalysis. *PLoS Genet.* 2:e190.
- Portman, M. 1884. *The Andaman Islanders*. Calcutta: Office of the Superintendent of Government Printing.
- Purcell, S., B. Neale, K. Todd-Brown et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575.
- Quatrefages, A. de. 1895. *The Pygmies*. New York: Appleton.
- Radcliffe-Brown, A. R. 1922. *The Andaman Islanders*. New York: Free Press of Glencoe.
- Reddy, B. M., B. T. Langstieh, V. Kumar et al. 2007. Austro-Asiatic tribes of Northeast India provide hitherto missing genetic link between South and Southeast Asia. *PLoS ONE* 2:e1141.
- Reich, D., K. Thangaraj, N. Patterson et al. 2009. Reconstructing Indian population history. *Nature* 461:489–494.
- Reid, L. A. 2013. Who are the Philippine negritos? Evidence from language. *Hum. Biol.* 85:329–358.
- Sahoo, S., A. Singh, G. Himabindu et al. 2006. A prehistory of Indian Y chromosomes: Evaluating demic diffusion scenarios. *Proc. Natl. Acad. Sci. USA* 103:843–848.
- Shi, H., Y. L. Dong, B. Wen et al. 2005. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *Am. J. Hum. Genet.* 77:408–419.

- Stock, J. T. 2013. The skeletal phenotype of “negritos” from the Andaman Islands and the Philippines relative to global variation among hunter-gatherers. *Hum. Biol.* 85:67–94.
- Tamura, K., J. Dudley, M. Nei et al. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 10.1093/molbev/msm092. *Mol. Biol. Evol.* 24:1,596–1,599.
- Temple, R. C. 1903. *Andaman and Nicobar Islands, in Census of India, 1901*. Calcutta: Government of India.
- Thangaraj, K., G. Chaubey, T. Kivisild et al. 2005. Reconstructing the origin of Andaman Islanders. *Science* 308:996.
- Thangaraj, K., G. Chaubey, T. Kivisild et al. 2006. Response to comment on “Reconstructing the origin of Andaman Islanders.” *Science* 27:470.
- Thangaraj, K., L. Singh, A. G. Reddy et al. 2003. Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr. Biol.* 13:86–93.
- Wang, H. W., B. Mitra, T. K. Chaudhuri et al. 2011. Mitochondrial DNA evidence supports north-east Indian origin of the aboriginal Andamanese in the Late Paleolithic. *J. Genet. Genomics* 38:117–122.