

***THE SHOMPEN OF GREAT NICOBAR  
ISLAND: NEW LINGUISTIC AND GENETIC  
DATA, AND THE AUSTROASIATIC  
HOMELAND REVISITED***

George van Driem  
*Leiden University, Leiden*

ABSTRACT

The Shompen are indigenous foragers living on Great Nicobar Island. The Danish scholar F.A. de Röepstorff was the first to record ethnographic and linguistic data on Nicobarese people and their languages in 1876. Yet data on the Shompen have always been rare. In a small book which appeared in 2003, Subhash Chandra Chattopadhyay and Asok Kumar Mukhopadhyay made a considerable body of Shompen data available for the first time. All these data were studied in Amsterdam this year by Roger Blench of Cambridge. His comparison of the Shompen data with Nicobarese and Austroasiatic lexical resources is slated for publication in *Mother Tongue*. The new data and comparative study have changed our view of Shompen. How can we assess the Shompen data and, in particular, various claims that have been made about Shompen? Is Shompen a Nicobarese language or, indeed, even Austroasiatic? Is Shompen a language isolate? What are the possible implications of the Shompen data for ethnolinguistic prehistory? What new questions do these data compel us to address?

***Introduction***

Recently, Roger Blench rendered the valuable service of making a newly available Shompen data set more widely

accessible. On the basis of those new data, Blench put forward the new and interesting idea that Shompen might represent a language isolate. Here a modicum of other newly available Shompen data collected by the late Elangaiyan is made more widely accessible. The earlier conjecture concerning the independent phylogenetic status of Shompen, however, is called into question. The view presented here is that Shompen is still just likely to be another language of the Nicobarese subgroup within the Nico-Monic branch of Austroasiatic.

### *The Nicobars and Austroasiatic*

The Nicobars form an archipelago between the Bay of Bengal and the Andaman Sea, located to the south-southeast of the Andaman Islands and just north-northwest of the northern tip of Sumatra. Whereas the languages of the Andamans have no known linguistic relatives anywhere else in the world, the Nicobarese languages constitute a sub-branch within the Nico-Monic or Southern Mon-Khmer branch of the Austroasiatic language family, as shown in Diagram 1. The Mon-Khmer-Kolarian language family was first recognised in the middle of the 19th century by Francis Mason (1854, 1860) and renamed Austroasiatic at the beginning of the 20th century by the Austrian Jesuit priest Wilhelm Schmidt (1904, 1906).

The languages of the Nicobarese subfamily are spoken by a little over 20,000 people on the Nicobar Islands. The specialist literature contains Nicobarese language names that generally resemble the names provided by Heinz-Jürgen Pinnow (1959). Recently, a research group led by V.R. Rajasingh conducted a pilot study in 2002 which identified new language names and has grouped together as ‘dialects’ related speech varieties.<sup>1</sup> In the northern portion of the archipelago, Pu: or Pu is spoken on Car Nicobar Island, and Tətət or Saněnyö is spoken on Chowra Island. Təihləŋ or Lurö is spoken on Teressa Island, and the closely related Po:əhət or Poahat is spoken on Bompoka Island. The 2002 study considers Po:əhət to be a dialect of Lurö.

The four speech forms spoken in the central portion of the archipelago, on the islands of Nancowry, Camorta, Trinkut and Katchall, are identified by the new survey as representing four dialects of a single language. Rajasingh refers to this language as Muöt, with Muöt proper being spoken on Nancowry Island. Pinnow refers to the language spoken on the islands of Nancowry and Camorta as Nancowry or Na:ŋkəuri, whilst the new survey assigns a distinct dialect name, viz. Kinlaka, to the Camorta island dialect. La:fu:l or Laful is spoken on Trinkut Island, and Te:hju or Tehnyu is spoken on Katchall Island.

In the south of the Nicobar archipelago, Lo'ɔŋ or Takahanylāng is spoken along the coast of Great Nicobar Island. The 2002 survey groups together the forms of speech on the islands of Milo, Condul and Little Nicobar as dialects of a single language called Lamongse, with Lamongse proper being spoken on Little Nicobar and Condul. Pinnow, however, distinguished under the name ɔŋ the distinct variety spoken on Little Nicobar Island, and reserved the term La:mɔŋje for the language of Condul. Miloh or Pihouny is spoken on Milo. Distinct from all other Nicobarese languages is ʃompē or Shompen, spoken in the hinterland of Great Nicobar Island.

The 1901 census counted 3,451 Car Nicobarese, 522 natives of Chowra, 702 Nicobarese on Teressa Island, a total of 1,095 natives on the central portion of the archipelago, with just 192 Nicobarese in the southern portion of the archipelago, in addition to 348 Shompen in the interior of Great Nicobar Island, giving a total native Nicobarese population of 6,310, excluding the 201 foreign traders then registered on the islands (Temple 1903, III: 142). Eighty years later, the 1981 census enumerated a total of 20,940 native Nicobarese plus 223 members of the Shompen tribe (Singh 1988: 60). Of these 223 Shompen, 46 were registered as 'workers', and 44 were recorded as being engaged in hunting and fishing. There were reportedly four literate Shompen men and two literate women. Recently, Singh reported that the major concentration of

Shompen was currently located ‘at a distance of 27 kilometres from Campbell Bay on East West Road’ (1994a: 1076). The Boxing Day Tsunami of 2004 disastrously affected the demography of all Nicobarese language communities.

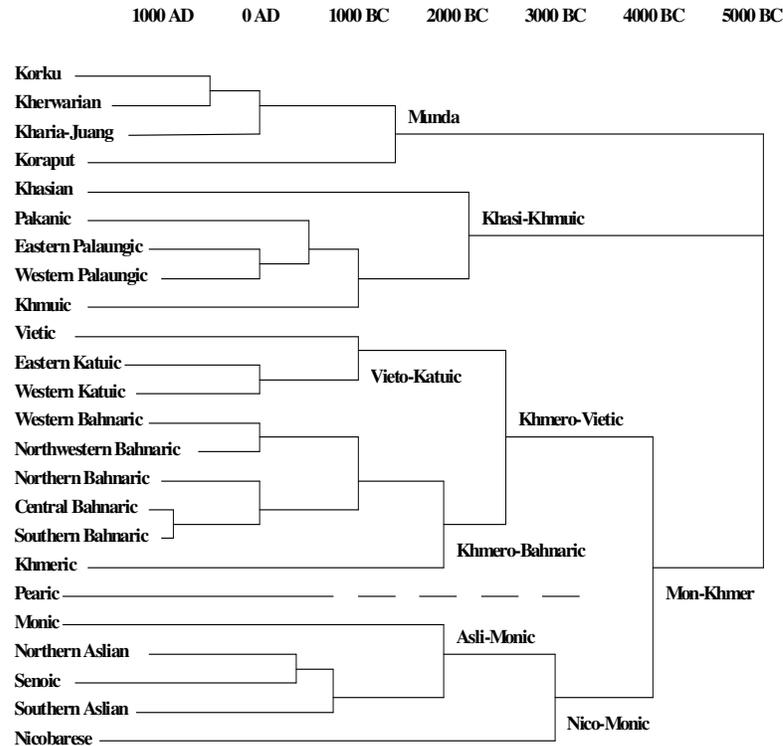


Diagram 1: Diffloth’s (2001, 2005) Austroasiatic languages family tree with his tentative calibration of time depths

***Early and Recent Glimpses of the Shompen Language***

Early Nancowry dictionaries and word lists of other Nicobarese languages were first compiled by two men of markedly different backgrounds, i.e. the Danish scholar Frederik Adolph de Roepstorff (1870, 1875 and posthumously

1884) and the Englishman Edward Horace Man (1872, 1888, 1889b). Both men recorded data on the Shompen or Shom Pen language. The Shompen are indigenous foragers who reside in the hinterland of Great Nicobar Island, and their language has always appeared to differ considerably from the other languages spoken on the Nicobars.

Frederik Adolph de Roepstorff<sup>2</sup> was born on the 25th of March 1842 at sea on a British vessel sailing from Madras to Europe, a circumstance which entitled him to British citizenship. He was christened at Cape Town and raised in Denmark. After his schooling, he returned to India in 1867, whereby he made use of his right to be recognised as a British citizen to become extra assistant superintendent on the Andamans in 1868, and later assistant superintendent of the Nicobars in 1877. On the 11th of January 1872, during home leave in Denmark, he married Hedevig Christiane Willemoës (born 30 November 1843, died 21 August 1896 at Copenhagen). He was murdered on the 24th of October 1883 by the bullet of a captive sepoy on Camorta (Bricka 1900, XIV: 519-520).<sup>3</sup> His grave lies in ‘the little Camorta graveyard, where the bluff near the English settlement overlooks the beautiful Nancowry harbour, and the nestling huts of the natives whom he loved so well’ (Chard 1884: i).

Edward Horace Man was born in Singapore on the 13th of September 1846 and educated in England. He first arrived at Port Blair in the Andamans in 1871 in order to take up employment as an assistant superintendent under his father Henry Stuart Man. Edward’s elder brother A.C. Man had preceded him in 1869 and had already compiled a first Andamanese word list, although this elder brother would later be killed in Burma. During his many years in the Andaman and Nicobar archipelagos, Edward Horace Man authored numerous Andamanese and Nicobarese linguistic studies. After his long service in the Nicobars and Andamans, he

enjoyed three decades of retirement in Brighton before dying of an illness on the 29th of September 1929.

Before Frederik de Roepstorff and Edward Horace Man, data on Nicobarese languages were collected sporadically. As early as 1778, Fontana (1792) recorded the very first short Nicobarese word list, and David Rosen (1839), a Danish pastor, published 63 Nancowry words and the Nancowry numerals. Frederik de Roepstorff provides a good account of much earlier and contemporaneous fieldwork on the Nicobars, but de Roepstorff remains the first scholar ever to have collected Shompen data. He held the Shompen or 'Shobængs' to be 'the aborigines of the Nicobars'. He reported that 'The Shobængs at Great Nicobar are hostile to the Nancowry people who reside along the coast, and not long ago a coastman was killed by them. This happened in December 1872' (1875: 2-3).

In contrasting his impressions of the Shompen as opposed to the coastal Great Nicobarese, Edward Horace Man seconded de Roepstorff's opinion that the Shompen represented the true aboriginal population of the Nicobars.

The Shom Pen have been — and I believe with good reason — accepted as the pristine indigenes, and their remote origin and purity of breed is apparently beyond question, while the various sections of the coast tribe, although differing from each other according to external influences and other circumstances, are without doubt descended from a mongrel Malay stock, the crosses being probably in the majority of cases with Burmese, and occasionally with natives of the opposite coast of Siam, and perchance also in remote times with such of the Shom Pen as may have settled in their midst; the fact that the Shom Pen present Mongolian affinities would thus to some extent account for the frequent occurrence of the oblique eye in a more or less marked degree throughout the group. (1889a: 365-366)

Frederik de Roepstorff described how he had been ‘fortunate enough to see one of these Shobængs. He was a big, strong youth, nearly as well built as those of Nancowry’. Based on his observation of the phenotypes, he developed a theory that the modern Nicobarese or ‘Nancowry race’, who ‘inhabit Trinkut, Nancowry, Camorta, Katchall, Car Nicobar and the coasts of Little and Great Nicobar’, had largely replaced the original inhabitants of the Nicobars, who had been ‘attacked and driven away from the best places, and a remnant of them is now found in the interior of Great Nicobar and on the little isolated island of Schowra’ [i.e. Chowra, just north-northwest of Teressa island] (1875: 3-4). Roepstorff managed to collect only ‘a few words’, he reported, ‘as it was not easy matter to obtain them from my Shobæng acquaintance’.

In fact, de Roepstorff recorded 329 words or expressions in the language of the ‘Shobængs’ or ‘inland race’ in addition to the Shompen numerals from one through ten. His comparative Nicobarese list contains many more items from the languages of Nancowry, Car Nicobar and Teressa Island and the Great Nicobar coastal dialect spoken by a language community of the ‘Nancowry race’. Later, Edward Horace Man, in his 1889 Nancowry dictionary, included 237 Shompen words, expressions and the numerals in an appendix entitled ‘Comparative List of Words in Common Use in the Six Dialects of the Nicobar Group’. At the time, Man estimated the population of the Shompen to be ‘say 750-1000’.

After the pioneering work of de Roepstorff and Man, no new linguistic data were seen from Great Nicobar Island for over a century.<sup>4</sup> Then in a small book which appeared in 2003, two Bengali linguists Subhash Chandra Chattopadhyay and Asok Kumar Mukhopadhyay made a considerable body of new Shompen data available. The new field research yielded a harvest of 723 Shompen words, 18 phrases and 23 sentences. A copy of this rare publication was brought to Europe in the spring of 2007 by my colleague and old friend Suhnu Ram

Sharma, who lent it to Laurie Reid, likewise a visiting scholar at Leiden, and through Laurie also to Roger Blench of Cambridge. The new Shompen data were studied in Amsterdam by Roger Blench, and his comparison of the Shompen data with Nicobarese and Austroasiatic lexical resources has now appeared in print, viz. Blench (2007). The new Shompen data were also made available to Gérard Diffloth, who assessed them against the earlier Shompen data and his own comparative Austroasiatic database.

In addition to the new data published by Chattopadhyay and Mukhopadhyay, unpublished material was collected by the late Rathinasabapathy Elangaiyan, who passed away on 18 January 2008. Elangaiyan undertook some eight to nine trips to the Nicobars since 1983 until just before the tsunami in 2004, staying for sojourns which varied in duration from two to four months. His main focus was the Pu language of Car Nicobar Island, but he also undertook to investigate the Shompen language in the interior of Great Nicobar Island. Elangaiyan visited the Shompen twice. Elangaiyan stayed at the Shompen Hut Complex, a collection of a few huts set up by the government to serve as the site for a health post and food distribution centre. There has never been a physician or any health workers permanently on duty at the hut complex, however.

On his first visit, Elangaiyan arrived at the hut complex with the assistance of porters which he had hired. Elangaiyan camped at the Shompen Hut Complex alone. Heavy rains ensued, and later he was stricken with *Plasmodium vivax* malaria. His condition and the water-logged terrain prevented him from leaving the site. During his illness and convalescence, the Shompen regularly visited him, and Elangaiyan conducted his first fieldwork whilst being tended and looked after by the helpful and friendly Shompen. After more than one and a half months at the hut complex, a small number of naval people came to the site for a picnic and stumbled upon Elangaiyan.

They sent back a message to the township and evacuated the much weakened Elangaiyan.

On his second visit, Elangaiyan again stayed at the township for a period of two and a half months. Elangaiyan's corpus of reliable data is scanty, he told me, because a monolingual approach without any contact language severely limits a linguist's ability of ascertaining the precise meaning of target language forms. The fieldwork was consequently beset with difficulties in ascertaining a precise description of the meanings. The fact that the Shompen at the hut complex are monolinguals also appears to have adversely affected the quality of the new data set provided by Chattopadhyay and Mukhopadhyay, whose fieldwork was subject to the same limitation. Elangaiyan reported that his knowledge of Pu, the language of Car Nicobar, was only somewhat helpful to him in dealing with the Shompen.

Elangaiyan prepared the native language primers for Pu, i.e. Car Nicobarese, used in mother tongue instruction. These are sound pedagogical textbooks. Likewise, the Shompen language primer is based mainly on Elangaiyan's fieldwork, and he is mentioned as a co-author in the produced primer. However, Elangaiyan was not at all pleased with the quality of the Shompen primer. He had strong reservations about the Shompen language primer even before its publication because his fieldwork data, though valuable, were intended for scholarly consumption by linguists only, with qualifications about specific uncertainties regarding certain forms and especially meanings. Nonetheless, administrative exigencies compelled the hasty publication of the Shompen primer. The Pu primers, entitled *Āô Tarik 1* and *Āô Tarik 2*, appeared in 1985 and 1987 respectively, published in Devanāgarī script by the Central Institute of Indian Languages at Mysore. The level 1 primer, entitled *Shompen-Hindi Bilingual Primer Śompen Bhāratī 1*, written in Devanāgarī script, appeared in 1995, jointly published by the Central Institute of Indian Languages

at Mysore and the Tribal Welfare Department of the Andaman and Nicobar Administration at Port Blair. The Shompen primer opens with the following words, authored by V. Gnanasundaram and M.R. Ranganatha of the Central Institute of Indian Languages at Mysore:

The Shompens are still a shy people who feel uncomfortable in the company of outsiders and at the first opportunity escape into the jungle. They never allow outsiders to know where they live. Their villages and homes are beyond the reach of outsiders.

Gérard Diffloth and I looked at his copies of these Nicobarese primers. The Shompen primer data consist of the following 70 items: *la:ʔə* ‘fish’, *ka:ʔav* ‘rain’, *ə:kʔa:t* ‘girl’, *kə:v* ‘dog’, *kagəy* ‘stars’, *kayəyəy* ‘parrot’, *pa:ʔa* ‘breadfruit’, *hnaʔu* ‘pig’, *mi:ʔi* ‘owl’, *əʔəʔi:* ‘black’, *əʔiyu* ‘red’, *ni:yi* ‘mouse’, *hiv ~ hi:v* ‘sun’, *giya:v* ‘scorpion’, *ηəηvo* ‘bamboo’, *phəηə:* ‘beehive’, *hməηəy* ‘snake’, *ηaɖuvi* ‘hoe’, *jəva:k* ‘spider’, *lovvu* ‘necklace or bracelet’, *thlovvu* ‘stone, rock’, *ə:ʔuvi ~ ə:ʔuvi* ‘old man’, *ʔə:yc* ‘cockroach’, *ɖə:ɖiyav* ‘woman’, *ba:pa:y* ‘papaya’, *o:kʔa:y* ‘infant’, *cyo:y* ‘macaque’, *ʔyo:y* ‘bread, taro, potato’, *ɖo:ʔo:* ‘hill’, *ā:ə* ‘mosquito’, *vahā:* ‘branch’, *ehā:* ‘root’, *mā:ə:v* ‘butterfly’, *okhʔam* ‘man pointing with both index fingers to the sides of his head’, *niyo* ‘house’, *thla:tayo* ‘housefly’, *əyə* ‘bat’, *əyʔev ~ əyayov* ‘centipede’, *ηa:niyō:* ‘log’, *ηa:tiyyū* ‘tree’, *agayñhyā:* ‘cloud’, *əhyu:* ‘pigeon’, *ñu:yi* ‘squirrel’, *pmāʔā:v* ‘frog’, *tyōvgo:* ‘beach, sand’, *təghəyə* ‘mango’, *bəvvu* ‘sprout’, *təmhəya:v* ‘coconut’, *ηə:ʔa* ‘eagle’, *ləv* ‘thigh’, *miyəv* ‘cheek’, *to:y* ‘lip’, *naη* ‘ear’, *ηə:yahi* ‘chin’, *hma:ñ* ‘eyes’, *hiyəhə* ‘anklebone’, *nuva:ñ* ‘neck’, *kuma:ñ* ‘forehead’, *həgʔa:y* ‘waist’, *ugiyəv* ‘fingernail’, *iya:i* ‘tongue’, *ləgə:v* ‘crab’, *ηahā:* ‘leaf’, *mə:ʔəy* ‘banana’, *omiyo:* ‘cat’, *ʔigʔa:k* ‘gaviyal’, *opʔa:k* ‘lead’, *phayayov* ‘red ant’, *həgvo:* ‘sea’, *kʔa:y* ‘moon’.

The romanisation here is a transliteration of the Devanāgarī orthography specifically developed for the Shompen primer

and is based on the phonetic explanations provided on two unnumbered pages in the introduction. We have made a number of transcriptional decisions. For example, the phonetic symbols [æ] and [ɛ] have been introduced to transliterate newly devised Devanāgarī vowel signs, and a vowel that might in fact be some central vowel has been transliterated here from the original Devanāgarī orthography as [ø], in strict adherence with the description provided in the front of the primer. The primer gives the Shompen words for ‘sun’, ‘centipede’ and ‘old man’ in two different Devanāgarī spellings. The meaning of some words was difficult to ascertain on the basis of the accompanying illustration alone. Although Elangaiyan stressed the unreliability of the data in this primer and the possibility of intra-Nicobarese loans in the data, Gérard Diffloth observed that it is nonetheless easy, even upon casual observation, to spot several well-known Nicobarese and Mon-Khmer etyma reflected in the data culled from this Shompen primer, e.g. *naŋ* ‘ear’, *løv* ‘thigh’, *ñiyo* ‘house’, *tɔmhøya:v* ‘coconut’.

#### ***Observations Regarding the Shompen Material***

Other than the Shompen primer and Elangaiyan’s unpublished field notes, the Shompen material comprises three distinct data sets. The early material consists of the 339 ‘Shobæng’ words or expressions, including the numerals from one to ten, that were published by de Roepstorff in 1875 and the 237 ‘Shom Peñ’ words, expressions and numerals published by Man in 1889. Man reported that the name ‘Shom Peñ’ was the coastal Great Nicobarese term for the inland people, consisting of the element *shom*, signifying ‘people’ or ‘natives’, and *peñ*, the proper name of a tribe, pronounced like French *pain*. The Shompen themselves, according to Man, referred to themselves as *Shab Daw’a* (1886: 432). The third data set, presented in 2003 by the two Bengali linguists Subhash Chandra Chattopadhyay and Asok Kumar Mukhopadhyay, comprises 723 Shompen words, 18 phrases and 23 sentences.

Impressions of Shompen phonology can be gleaned from the available material. Frederik de Roepstorff's notation distinguished  $a \sim \bar{a}$ , and perhaps this orthographic distinction denoted two distinct vowels, viz. /ə/ vs. /a/, in accordance with Indological convention. His notation also differentiated  $e \sim \acute{e}$  and  $o \sim \bar{o}$ . These distinctions suggest a possible length contrast or tense vs. lax opposition. Similarly, Man's notation differentiated the Shompen vowels  $a \sim \grave{a} \sim \hat{a}$  and also made the distinctions  $e \sim \bar{e}$ ,  $i \sim \bar{i}$ ,  $o \sim \hat{o} \sim \bar{o}$  and  $u \sim \bar{u}$ . Chattopadhyay and Mukhopadhyay describe Shompen as having seven or eight vowels /i, e, ε, a, ā, ɔ, o, u/, depending on what we are inclined to think about the contrast represented as  $a \sim \bar{a}$ . All eight of these vowels can reportedly be nasalised. Due to font difficulties, Chattopadhyay and Mukhopadhyay use capital E for Shompen /ε/ and capital O for the vowel /ɔ/. Blench takes Chattopadhyay and Mukhopadhyay's account at face value and accepts that their orthographic distinction  $a \sim \bar{a}$  as representing a length contrast, whilst I am inclined not to exclude the possibility that what the two authors mean by 'phonemic length', restricted to just this one Shompen vowel, might very well just represent two vowels of an altogether different timbre.

The Shompen consonant phoneme inventory according to Chattopadhyay and Mukhopadhyay comprises the phonemes /ʔ, k, kh, g, gh, ŋ, c, j, ɲ, t, th, d, n, p, ph, b, bh, m, y, ɣ, l, w, φ, x, h/. Shompen purportedly lacks a phoneme /dh/, analogous to Shompen /gh/ and /bh/. Shompen has no sibilants, but has the fricatives /φ/ and /x/. Shompen has a phonemic glottal stop. In the notation used by Blench, Chattopadhyay and Mukhopadhyay's symbols ʔ, ñ and ñ̄ have been replaced by the more current phonetic symbols ʔ, ŋ and ɲ respectively.

In evaluating the Shompen lexical material, the differences between the three data sets is the first observation to which any close scrutiny will lead. Chattopadhyay and Mukhopadhyay's (2003) data set resembles that of Man (1889b), but neither Chattopadhyay and Mukhopadhyay nor Man very closely re-

semble de Roepstorff's (1875) data set. At the same time, the selection of lexical items reflected in the material collected by Chattopadhyay and Mukhopadhyay appears to be somewhat imbalanced. There are two likely causes to which these discrepancies might be attributed.

First, Man observed that Shompen is not so much a single language as an internally diverse group of inland dialects, with each community possessing 'a dialect more or less distinct, but this is what might reasonably be expected when we consider the isolation of the several encampments, and the difficulties of intercommunication, apart even from the hostile relations in which they stand towards one another' (1886: 449). Man remarked in particular that the *dakan-kat* dialect<sup>5</sup> of Shompen spoken near Kashindōn on the west coast exhibited a high degree of lexical divergence from the Shompen spoken at Lafal and Ganges Harbour (1886: 448).

Over a century later, Chattopadhyay and Mukhopadhyay too reported two groups of Shompen. One Shompen population is a semi-nomadic hunter-gatherer group 'living in deep forests in the northern and the central parts of the island around the Galathia and the Alexandria rivers'. They barter jungle produce for food and also receive food and medical care through a government welfare programme. They hunt with spear and are reportedly unfamiliar with bow and arrow. The other Shompen group lives on the east coast of Great Nicobar, where they 'are in better contact, especially with the local Nicobarese tribe'. The eastern coastal group speak some Lo'ōŋ, i.e. coastal Great Nicobarese, and some of these Shompen also understand Hindi and frequent the government offices at Campbell Bay. Chattopadhyay and Mukhopadhyay reportedly collected their data 'from the last week of December 2000 up to the 1st week of February 2001' from the semi-nomadic deep forest group at the Shompen Hut Complex, located 27 km from Campbell Bay on the East-West Road. The authors assert that

these deep forest Shompen never go to Campbell Bay (2003: 1-3).

Secondly, an impression which Gérard Diffloth and I shared when studying the 2003 data set is that another cause for the discrepancy between the three available data sets might be a fieldwork problem especially affecting the most recent study. It is unclear which contact language the researchers used with the reportedly monolingual and shy Shompen and what consequences this difficult fieldwork situation may have had on the quality of the data elicited. Chattopadhyay and Mukhopadhyay record the Shompen pronominal forms *iṣ* ~ *ihṣ* 'I', *ca* 'my', *emāu* 'we' (dual exclusive), *eo* 'we' (dual inclusive), *eō* 'he', *onā* 'his'. Yet the data set contains no words for 'we' in the plural (vs. the dual), nor does the glossary contain any second person pronominal form. However, the authors record three utterly different words for 'vagina', i.e. *ipudāo*, *ugāu*, *totoghāb*. Also, Shompen purportedly has a lexicalised expression *γiāi igoki*, glossed by Chattopadhyay and Mukhopadhyay as 'dismatting' (2003: 37), an unfamiliar, possibly administrative term which can also be found on a few Keralan and Bengali websites.

The new data set by Chattopadhyay and Mukhopadhyay provides the Shompen form *koceon* for 'cat', a Malay loan word found throughout the Nicobars, but Frederik de Roepstorff recorded an abbreviated form *tjīng* for Shompen 'cat'. It is conceivable that the truncated form was the earlier loan which Shompen acquired from Lo'ṅ or Coastal Great Nicobarese, and that the word was subsequently loaned again. Nandan (1993: xx) records the Coastal Great Nicobarese form *kuching* 'cat'. Finally, Chattopadhyay and Mukhopadhyay report that syntactically the basic syntactic element order of Shompen is verb-subject-object (VSO).

Chattopadhyay and Mukhopadhyay's data set is therefore problematic, and a comparative study based on the 2003 data

set led Roger Blench to conclude that Shompen has ‘no obvious relationship with other Nicobarese languages or other Mon-Khmer languages’. Blench goes on to speculate that: ‘As with the Andamans, the possibility that the Shom Pen represent a relic of early human expansion around the rim of the Indian Ocean should be seriously considered’. Is Shompen then not Austroasiatic at all and therefore perhaps a language isolate of South Asia like Nahali, Vedda, Kusunda or Burushaski? Have the new data changed our view of Shompen? What are the possible implications of the new Shompen data for ethnolinguistic prehistory?<sup>6</sup>

Only a thorough holistic description of the language can resolve such uncertainties. New work on Shompen urgently needs to be undertaken by a gifted and dedicated field linguist willing to brave the dangers of malaria and the discomforts of conducting fieldwork at the Shompen Hut Settlement. There a linguist could take up the challenge of conducting arduous work with monolingual Shompen speakers. Also, new comparative tools such as Stampe’s Munda database and Shorto’s (2006) comparative Mon-Khmer dictionary are now available. Diffloth (2008) should be carefully consulted, however, before considering using Shorto (2006) as a reference.<sup>7</sup> At the same time, new data on Nicobarese languages have been provided in several studies, e.g. Whitehead (1925), Radhakrishnan (1981).

Meanwhile, we can best trust Gérard Diffloth’s assessment of the more reliable earlier Shompen data collected by Frederik de Roepstorff and Edward Horace Man in light of his comparative Austroasiatic database. Diffloth assesses that ‘out of 222 Shompen lexemes, 109 have cognates with other Nicobarese languages’, whereas ‘102 have no identifiable cognates’, and ‘7 have *South Mon-Khmer cognates not found* in other Nicobarese languages’. Two of the 222 lexical items can be identified as borrowings from Malay. Out of the 109 shared Nicobarese etyma in Shompen, 57 also have good

Southern Mon-Khmer cognates. The seven Shompen lexical items that have no Nicobarese cognates but are shared with other South Mon-Khmer or Nico-Monic languages are *toak* ‘afraid’, *hohom* ‘bathe’, *aløv* ‘pig’, *chuk* ‘foot’, *kateap* ‘egg’, *kakoay* ‘sit’ and *kamyak* ‘husband’. Gérard also points out that Shompen has undergone a regular sound change, whereby Austroasiatic final nasals, retained as final nasals in Nicobarese and most mainland Mon-Khmer languages, are reflected as devoiced stops. This fact indicates that such good Austroasiatic roots cannot have been borrowed from mainland Mon-Khmer languages, and that Shompen is a language belonging to the Nicobarese branch, not a language isolate (Diffloth 2007).

#### ***The Physical Anthropology of the Shompen***

Even in the old physical anthropology of frizzy hair and phenotypes, the somatological affinities of the Shompen were a heated topic from the start. The proximity of the negrito populations of the Andamans in conjunction with the idea that the inland Shompen represented some aboriginal remnant group suggested to the minds of many that the Shompen too were a negrito people. Frederik de Roepstorff was the first to assail the then widely held view that the Shompen were a negrito population. He maintained that the Shompen were of ‘Mongoloid’ stock. Some resisted this idea, preferring to entertain the view that the Shompen were of ‘Negrito stock, allied to the Andamanese or the Semangs of the Malay peninsula’ (Distant 1879: 336).<sup>8</sup>

A detailed old-fashioned physical anthropology of the Nicobarese peoples is provided by Man, who noted that the ‘characteristic tint’ of the Shompen was ‘a dull brown’ lacking ‘the healthy appearance which distinguishes the coast people’ (1889a: 390). The ossuary practices on the islands of Bompoka and Teresa suggested to Bonington early cultural contacts with Melanesians or, in his own words, ‘the existence of a

strong Melanesian element in the Nicobars in spite of their Mon language' (1926: 106). Studies such as Ball (1881), Man (1889a), Boden Kloss (1903) and Meerwarth (1919) contain interesting descriptions and valuable photographic documentation of the Nicobarese people and their architecture. Recent accounts of the Nicobarese in their current circumstances, sometimes including pictorial documentation, are provided by Agarwal (1967), Dagar and Dagar (1999), Krishan (1986), Lal (1977), Justin (1990), Nandan (1993) and Rizvi (1990).

The new physical anthropology focuses on molecular polymorphisms in the double helices of the chromosomes and on the mitochondrial DNA. Recently some molecular genetic work has been done on the Shompen. Twelve Shompen males were sampled in a study, and all were found to bear the O2a (M95) haplogroup on their Y chromosome (Trivedi *et al.* 2006).<sup>9</sup> This single nucleotide polymorphism has been identified as a possible marker for a paternal lineage reflecting an ancient male-driven spread of the Austroasiatic language family (van Driem 2007).<sup>10</sup> In fact, the correlation of linguistic and population genetic findings has suggested that many language communities speak father tongues rather than mother tongues. Languages and entire language families appear often to have been disseminated by male speakers.

The widespread nature of the correlation of language with a few predominant Y haplogroups suggests that it must have been a recurrent motif in ethnolinguistic history that mothers at one point in time were compelled to raise their children in the language of the fathers. Based on the work of Estella Poloni and her teammates (1997, 2000), this phenomenon, which I called the 'Father Tongue hypothesis' in Taipei in 2002, has consequences for the way historical linguists will in future have to think about language change. This phenomenon also opens up the question of whether the sexual dimorphism in our species with respect to linguistic abilities and language

sensibility could have its evolutionary origins in the dynamics of warfare, competition and linguistic assimilation between rival language communities in an ancestral age.

Trivedi *et al.* (2006) do not specify other single nucleotide polymorphisms (SNPs) which they may have typed that might have distinguished different lineages within the clade. This would have been helpful, for we have more recently come to know that the O2a (M95) haplogroup can be subdivided into O2a\*, bearing only the M95 mutation, and O2a1a (PK4) and O2a1\* (M88, M111). In their study, the short tandem repeats (STR) within the O2a haplogroup suggested a greater affinity between the Shompen and the Munda than with other Nicobarese, and the greatest distance to Austroasiatic language communities of Southeast Asia. However, short tandem repeats are highly variable and especially useful as forensic markers. Therefore, whilst the STR profile provided by Trivedi *et al.* (2006) is suggestive, the short tandem repeats provide no clear-cut picture of affinities and lack monophyletic resolution. Trivedi *et al.* (2006) claim that the Shompen represent the ‘descendants of Mesolithic hunter-gatherers’. Although their data provide no support for this assertion, it may of course be true that most people on earth today happen to descend from Mesolithic hunter-gatherers at some time and place.

The mitochondrial DNA of the Shompen is reportedly characterised by the two clades B5a and R12. The B5a configuration represents a newly identified clade with a coalescence age of 17,000 years and geographical distribution mainly in insular and littoral Southeast Asia. The ‘R12’ clade, which will probably be relabelled ‘R22’ in the newly emergent conventional mtDNA nomenclature, is common amongst other populations native to the Nicobars and represents a lineage which is also seen in Vietnam, Indonesia, the Philippines and on Taiwan. In short, the population genetic data can be seen as corroborating to some extent the linguistic view that we have

of Nicobarese as a branch of Austroasiatic, though, of course, population genetic data should not necessarily be expected to do so. The newly developed autosomal markers have yet to be tested on the Shompen, other Nicobarese peoples and Austroasiatic language communities.

### ***Linguistic Palaeontology and the Austro-asiatic Homeland***

In addressing the question of the precise whereabouts of the Austroasiatic ancestral homeland from a purely linguistic point of view, the two foremost criteria in our deliberations are the findings of linguistic palaeontology and the geographical centre of gravity of the language family based on the distribution of modern Austroasiatic language communities and deep phylogenetic divisions in the family. Then these inferences can be critically assessed in view of relevant information from other fields such as archaeology and population genetics. The distribution of the modern language communities and the geography of the deepest historical divisions in the family's linguistic phylogeny would put the geographical centre of the family somewhere between South Asia and Southeast Asia, in the area around the northern coast of the Bay of Bengal.

Gérard Diffloth pointed out in his keynote address on 'Considerations of the homeland of Austroasiatic', with which he inaugurated the 3rd International Conference on Austroasiatic Linguistics (ICAAL 3) at Deccan College on 26 November 2007, that nobody knows the higher-level nodes of Austroasiatic for sure, which leaves the question of the earliest branchings undetermined. If the deepest division in the family lies between Munda and the rest, as an older generation of scholars used to suspect, then the geography of deep historical divisions in linguistic phylogeny would compel us to look for a homeland on either side of the Ganges delta, although we would be unable to say precisely whether this homeland would have to have lain to the east or to the west of the delta. If we

assume the veracity of Diffloth's new tripartite division, shown in Diagram 1, the geography of the deepest phylogenetic divisions within Austroasiatic would likewise suggest a homeland in this region.

Linguistic palaeontology, a term introduced by Adolphe Pictet in 1859, is an attempt to understand the ancient material culture of a language family on the basis of the lexical items which can be reliably reconstructed for the common ancestral language. The linguistic palaeontology of Austroasiatic strongly qualifies the ancient Austroasiatics as the most likely candidates for the first cultivators of rice. At the same time, Diffloth has shown that the reconstructible Austroasiatic lexicon paints the picture of a fauna, flora and ecology of a tropical humid homeland environment.

Diffloth (2005: 78) has shown that three salient isoglosses diagnostic for the faunal ecology of the Proto-Austroasiatic homeland can be reconstructed all the way to the Austroasiatic level and are reflected in all branches, including Munda, i.e. \*mra:k 'peacock *Pavo muticus*', \*tərkuət 'tree monitor lizard *Varanus nebulosus* or *bengalensis*' and \*tənyu:ʔ 'binturong' or the 'bear cat *Arctitis binturong*', a black tropical mammal that is the largest of the civet cats. All of these species are not native to areas that currently lie within China, and, to our present knowledge, these species never were native to the area that is today China. More reconstructible Proto-Austroasiatic roots indicative of a tropical or subtropical climate are adduced by Diffloth (2005: 78), i.e. \*(bən)jo:l ~ \*j(ərm)o:l 'ant eater, *Manis javanica*', \*dəkan 'bamboo rat, *Rhizomys sumatrensis*' (an Austroasiatic root which has found its way into Malay as a loan), \*kaciaŋ 'the Asian elephant, *Elephas maximus*', \*kiae 'mountain goat, *Capricornis sumatrensis*', \*rəma:s 'rhinoceros, *Dicerorhinus sumatrensis*' and \*tənriak 'buffalo, *Bubalus bubalus*'.

Finally, Diffloth (2005: 78) points out a fact long noted by scholars of Austroasiatic linguistics, e.g. Osada (1995), namely that a rich repertoire of reconstructible roots representing ancient rice agriculture is robustly reflected in all branches of Austroasiatic, viz. *\*(kə)ba:ʔ* ‘rice plant’, *\*rəŋko:ʔ* ‘rice grain’, *\*cəŋka:m* ‘rice outer husk’, *\*kəndək* ‘rice outer husk’, *\*phe:ʔ* ‘rice bran’, *\*təmpal* ‘mortar’, *\*jənrəʔ* ‘pestle’, *\*jəmpiər* ‘winnowing tray’, *\*gu:m* ‘to winnow’, *\*jərmuəl* ‘dibbling stick’ and *\*kəntu:ʔ* ‘rice complement’, i.e. accompanying cooked food other than rice.

Nicole Revel (1988) contributed one of the most elaborate ethnobotanical studies on rice, rice cultivation practices and rice terminology in various Asian language communities. The other main candidate for early cultivators of rice are the ancestral Hmong-Mien. Great strides have been made in our understanding of Hmong-Mien historical phonology (Haudricourt 1954, Purnell 1970, Wáng and Máo 1995, Niederer 1998), although the reconstructible lexicon specific to rice cultivation is less impressive than the Austroasiatic repertoire. The three Hmong-Mien etyma relating to rice cultivation that appear to be original to the linguistic phylum are *\*ntsə:i* ‘husked rice’, *\*ja:ŋ* ‘cooked rice’ and *\*ŋjeŋ* ‘rice head, head of grain’, whereas the Hmong-Mien terms for glutinous (rice), (paddy) field, sickle, rice cake and (rice) seedling ‘are likely to have had a Chinese origin’ (Ratliff 2004: 158-159).

The rice story is complex, and the plot of the story has changed more than once in recent decades. Whereas the origin of rice cultivation was once held ‘incontestably’ to have lain in the Indian subcontinent (Haudricourt and Hédin 1987: 159-161, 176), subsequent scholarship moved the homeland of rice agriculture from the Ganges to the Yangtze. For years of conventional wisdom in archaeological circles dictated that rice was domesticated in the Middle Yangtze, perhaps as early as the sixth millennium BC.

More recently, scholars have increasingly begun to take note of findings that would move the original homeland of rice cultivation back to the Indian subcontinent. Against the background of older datings of domesticated rice and ceramic culture from Gangetic basin and Doab sites such as Koldihawa and Mahagarha, reportedly dating from the seventh millennium BC (Sharma *et al.* 1980, Pal 1990, Agrawal, 2002), there are now newer sites with more reliable dates at Lahuradewa (Lahurādevā), Tokuvā and Sarāī Nahar Rāī.

At the Lahuradewa site (26°46' N, 82°57' E), the early farming phase, corresponding to period 1A in the site's clear-cut stratigraphy, has radiocarbon dates ranging from ca. 5300 to 4300 BC. Carbonised material from period 1A was collected by the flotation method, yielding *Setaria glauca* and *Oryza rufipogon* as well as a morphologically distinct, fully domesticated form of rice 'comparable to cultivated *Oryza sativa*' (Tewari *et al.* 2002). More recently, accelerator mass spectroscopy dates were obtained on the rice grains themselves, corroborating the antiquity of rice agriculture at the site.

Most recently, new radiocarbon dates for rice agriculture have been coming from the Ganges basin, with the Tokuvā site near Allahabad now yielding similar dates (Vasant Shinde [Vasant Śivarām Śinde], personal communication 27 November 2007), and exciting new dates for ancient rice agriculture are also emerging from Sarāī Nahar Rāī (Manjil Hazarika, personal communication 7 March 2008). Of course, we are living at a time when a more reliable calibration of radiocarbon dates in general has become a matter of great urgency. At the same time, as Prof. Rām Dayāl Muṇḍā of Ranchi University pointed out in his inaugural address at the opening session of the 3rd International Conference on Austroasiatic Linguistics (ICAAL 3), the bulldozer effect of globalisation in present and former Munda areas is effacing the traces of ancient Austroasiatic archaeology and palaeobotany.

Further east, at least five species of wild rice are native to northeastern India, viz. *Oryza nivara*, *Oryza officianalis* (*O. latifolia*), *Oryza perennis* (*O. longistaminata*), *Oryza meyeriana* (*O. granulata*) and *Oryza rufipogon*, and reportedly over a thousand varieties of domesticated rice are currently in use in the region (Hazarika 2005, 2006a). The different varieties of rice in northeastern India are cultivated in three periods by distinct cultivation processes. In the process of *āhu kheti*, the rice is sown in the months of *Phāgun* and *Sot*, i.e. mid February to early April. The seedlings are not transplanted but ripen in just four months in fields which must be constantly weeded. In *bāu kheti*, the rice seedlings are sown from mid March to mid April in ploughed wet fields and likewise do not need to be transplanted. In *śāli kheti*, the rice is sown from mid May to mid June, and the seedlings are transplanted. *Śāli kheti* rice varieties are suspected to derive from the wild *officianalis* rice still widely found in swampy village areas. The wild *rufipogon* rice cannot be used for human consumption because the plants shed their seeds before they ripen, so that *rufipogon* rice is used in Assam and other parts of northeastern India as cattle feed (Hazarika 2006b).

Whilst claims have been published of rice cultivation in East Asia as long as around 10,000 BC, the currently available evidence indicates that immature morphologically wild rice may have been used by foragers before actual domestication of the crop, e.g. at the 八十擋 Bāshídàng site (7000-6000 BC) belonging to the 彭頭山 Péngtóushān culture in the Middle Yangtze and at sites in the Yangtze delta area such as 跨湖橋 Kuàhúqiáo, 馬家浜 Mǎjiābāng 河姆渡 (5000-3000 BC) and Hémǔdù (5000-4500 BC). However, only ca. 5000 BC was the actual cultivation of rice probably first undertaken by people in the Lower Yangtze, who at the time relied far more heavily on the collecting of acorns and water chestnuts (Yasuda 2002,

Fuller 2005a, 2005b, 2005c, 2006a, 2006b, 2006c, 2007a, 2007b, Fuller *et al.* 2007, Zong *et al.* 2007). There is also currently no evidence for the co-cultivation of rice and foxtail millet along the middle Yangtze until around 3800 BC (Nasu *et al.* 2006).

Today, our understanding of the palaeoethnobotanical picture is more complex. The two main domesticated varieties of rice, *Oryza indica* and *Oryza japonica*, are phylogenetically distinct and would appear to have been domesticated separately. *Oryza indica* derives from the wild progenitor *Oryza nivara* and was first cultivated in South Asia or western Southeast Asia, perhaps in two separate domestication events. On the semi-arid Gangetic plain at the end of the mid-Holocene wet period, habitats for wild rices increasingly shifted to oxbows as palaeochannels dried up and turned into oxbow ponds. This shift favoured monsoonal rather than marshland rice species, including *Oryza nivara*, the wild progenitor of *Oryza indica* (Fuller 2006a).

*Oryza japonica* derives from the wild progenitor *Oryza rufipogon*, and it is currently believed that the *rufipogon* variety was first cultivated to yield early *Oryza japonica* along the Middle Yangtze. Harvey *et al.* (2006) have critically reassessed the morphometrics of rice finds associated with various Neolithic sites throughout the Yangtze basin in light of recent genetic findings. It appears that the wild progenitor *Oryza rufipogon* was not fully domesticated in the Lower Yangtze to yield early *Oryza japonica* until ca. 4000 BC. Generally, the archaeological record shows a delay of one to two millennia between the beginning of cultivation and the first clear evidence of domestication *sensu stricto*, i.e. genetic modification by selective breeding.

Twelve wild forest-margin rice species are known, found mostly in Southeast Asia as well as at old sites of human habitation, e.g. Jiǎhú in the seventh millennium BC or Hémǔdù

in the first half of the fifth millennium BC. Extinct wild varieties of rice also appear to be preserved in the modern *japonica* genome. Based on the genetics of the *officinalis* variety, the seasonally wet, puddle-adapted *Oryza nivara*, and the always wet perennial *Oryza rufipogon*, there may be evidence for multiple rice domestications in South, Southeast and East Asia. So, maybe the domesticators of *Oryza nivara* were ancient Austroasiatics, and maybe the domesticators of ancient *Oryza rufipogon* were ancient Hmong-Mien.

O'Connor (1995) and Blench (2001) have argued that irrigated rice agriculture enabled people to seize control of lowlands and flood plains. People were able to move down from upland areas that had hitherto been more favourable habitats after wet cultivation had transformed lowlands from epidemiologically undesirable places into bountiful habitats. But what if the first cultivators and domesticators of rice already inhabited lowland river basins and flood plains, such as the Ganges or Yangtze basins or even the Brahmaputran flood plains?

Turning to northeastern India and the Indo-Burmese borderlands, we must recognise that, notwithstanding the excellent archaeological work conducted in the Ganges and Yangtze river basins, much of the archaeology of ancient rice agriculture is simply not known because no substantive archaeological work has been done on the Neolithic in the most relevant areas, e.g. northeastern India, Bangladesh and Burma. The sheer dearth of archaeological research in these areas leaves entirely open the possibility that rice cultivation may have originated in this region. We might expect to find traces of ancient farming communities better preserved in the hill tracts surrounding the Brahmaputran flood plains than on the fertile fields themselves, although the earliest rice-based cultures may first have developed on those very flood plains. Perhaps the remains of the first rice cultivating cultural assemblages lie buried forever in the silty sediments of the

sinuous lower Brahmaputran basin or were washed out by the Brahmaputra long ago into the depths of the Bay of Bengal.

#### NOTES

1. Unless stated otherwise, I first provide the language name given by Pinnow (1959) and then the recently introduced language name identified in the 2002 pilot survey. I thank V.R. Rajasingh for kindly providing me with these newer names from their yet unpublished pilot survey report.
2. The surname has sometimes appeared in print in the orthography 'de Röepstorff'.
3. In a study published in the formerly Danish city of Lund, Simron Jit Singh (2003) provides a valuable historical account of European dealings in the Nicobars, with special emphasis on the Danes, yet somehow he manages to entirely overlook Frederik Adolph de Roepstorff.
4. In 1993, Nandan included a glossary of 137 words and expressions from Great Nicobar, including several obvious Indo-Aryan loans like 'chāpāti', 'dāl', 'ātā' and 'ghee'. Judging from the items, the language documented is Lo'ṅṅ, the coastal dialect of Great Nicobar, not Shompen, e.g. Nandan's *nang* 'ear' vs. Shompen *gña*, Nandan's *pukoi* 'pig' (cf. de Roepstorff's *bakoi*) vs. Shompen *noṅg*, Nandan's *em* 'dog' vs. Shompen *küip*.
5. The term *dakan-kat* would appear to denote the 'ill-adjusted loin-cloth' worn by this group of unkempt Shompen 'which they evidently wear in imitation of the *neng* of the coast men' (Man 1886: 447).
6. Chattopadhyay and Mukhopadhyay venture an attempt to relate Shompen to Tibeto-Burman, Kra-Dai (Daic), Austroasiatic and Austronesian. To this end, the only evidence adduced consists of three Shompen, Fijian and Samoan lexical items glossed as 'canoe', 'pandanus' and 'coconut'.
7. In fact, it may not be too late to follow up on Diffloth's suggestion of publishing a photo-facsimile edition of Shorto's original manuscript and notes, just as the Soviet Academy of Sciences did belatedly in 1960 with the valuable polyglot notes

- of the murdered Tangut scholar Nikolaj Aleksandrovič Nevksij (cf. van Driem and Kepping 1991, van Driem 1993).
8. In his recounting of the tale, Roger Blench writes that ‘the fact that the Shom Pen have straight hair, like the Nicobarese, brought an untimely end to such speculation’, i.e. the conjecture of early ethnographers that the Shompen might represent a missing link between the Andamanese and the indigenous negrito population groups of the Malayan peninsula. This statement is placed underneath a photograph showing at least two Shompen men with unmistakably frizzy hair, one of whom could even be said to be sporting the coiffure once popularly referred to as an ‘afro’. Blench hastens to observe, however, that ‘the issue of straight hair has been questioned, with some populations apparently having wavy hair’.
  9. Some Nicobarese population genetic data were also included in recent Andamanese studies, i.e. Thangaraj *et al.* (2003), Thangaraj *et al.* (2005), Palanichamy *et al.* (2006).
  10. Kumar *et al.* (2007) essentially corroborate my interpretation of the earlier work on the O2a haplogroup and conclude on the basis of M95 ‘that the Mundari populations are one of the earliest settlers in the Indian Subcontinent’. The study by Kumar *et al.* (2007) is informative for the Munda groups, though the dating is wrong. Their article argues in favour of a hypothesis about Austroasiatic origins which is entirely untestable on the basis of their sampling, including their speculation that ‘these populations have come from Central Asia through the Western Indian corridor and subsequently colonized Southeast Asia’.

### **References**

- Agrawal, D.P. 2002. ‘The earliest pottery and agriculture in South Asia’, pp. 81-88 in Yoshinori Yasuda, (ed.). *The Origins of Pottery and Agriculture*. New Delhi: Lusre Press and Roli Books, for the International Center for Japanese Studies.
- Agarwal, H.N. 1967. ‘Physical characteristics of the Shompen of Great Nicobar Island’. *Bulletin of the Anthropological Survey of India, Calcutta* 14: 83-97.

- Ball, V. 1881. 'On Nicobarese ideographs'. *Journal of the Anthropological Institute of Great Britain and Ireland* 10: 103-108.
- Blench, Roger Marsh. 2001. 'From the mountains to the valleys: Understanding ethnolinguistic geography in Southeast Asia', pp. 31-50; In: Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas, (eds.). *The Peopling of East Asia: Putting Together the Archaeology, Linguistics and Genetics*. London: Routledge.
- . 2007. 'The language of the Shompen: A language isolate in the Nicobar Islands'. *Mother Tongue* XII: 179-202.
- Boden Kloss, C. 1903. *In the Andamans and Nicobars: The Narrative of a Cruise in the Schooner 'Terrapin', with Notices of the Islands, Their Fauna, Ethnology, Etc.* London: John Murray.
- Bonington, C.J. 1926. 'Ossuary practices in the Nicobars, with particular reference to the practice of keeping the skull of an ancestor on or in a life-size wooden body on the islands of Teresa and Bornpoka', *Man*, XXXII (133-157): 105-106.
- Bricka, Carl Frederik, ed. 1887-1905. *Dansk Biografisk Lexikon* (19 vols.). Copenhagen: Gyldendalske Boghandels Forlag.
- Chard, C.H. 1884. 'Preface' and 'Introduction' in Frederick Adolph de Roepstorff (1884).
- O'Connor, Richard A. 1995. 'Agricultural change and ethnic succession in Southeast Asian states: A case for regional anthropology.' *Journal of Asian Studies* 54.4: 968-996.
- Dagar, J.C., and Dagar, H.S. 1999. *Ethnobotany of Aborigines of Andaman-Nicobar Islands*. Dehra Dun: Surya International Publications.
- Diffloth, Gérard. 2001. 'Tentative calibration of time depths in Austroasiatic branches.' Paper presented at the Colloque «Perspectives sur la Phylogénie des Langues d'Asie Orientales» at Périgueux, 30 August 2001.
- . 2005. 'The contribution of linguistic palaeontology to the homeland of Austroasiatic', pp. 77-80. In: Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas (eds.). *The Peopling of*

- East Asia: Putting Together the Archaeology, Linguistics and Genetics*. London: Routledge Curzon.
- . 2007. 'A report on Shompen to Our Excellency George van Driem'. Deccan College at Pune: unpublished memorandum dated 28 November 2007.
- . 2008. Review of Shorto (2006), *Diachronica*, XXV (1): 137-142.
- Distant, W.L. 1879. 'The people inhabiting the interior of the Great Nicobar Island'. *The Journal of the Anthropological Institute of Great Britain and Ireland* 8: 336.
- van Driem, George, and Ksenia Borisovna von Kepping. 1991. 'The Tibetan transcriptions of Tangut (Hsi-hsia) Ideograms', *Linguistics of the Tibeto-Burman Area* 14.1: 117-128.
- van Driem, George. 1993. 'Ancient Tangut manuscripts rediscovered'. *Linguistics of the Tibeto-Burman Area* 16.1: 137-156.
- . 2007. 'Austroasiatic phylogeny and the Austroasiatic homeland in light of recent population genetic studies'. *Mon-Khmer Studies* 37. 1-14.
- Elangaiyan, Rathinasabapathy, et al. 1985. *Ādivāsī Kalyāṇa Vibhāg Aṇḍamān tathā Nikobār Praśāsan*, and Central Institute of Indian Languages [ix + 155 pp.].
- . 1987. *Ādivāsī Kalyāṇa Vibhāg Aṇḍamān tathā Nikobār Praśāsan*, and Central Institute of Indian Languages [x + 161 pp.].
- . 1995. *Shompen-Hindi Bilingual Primer Śompen Bhāratī 1*. Port Blair and Mysore: Tribal Welfare Department of the Andaman and Nicobar Administration, Ādivāsī Kalyāṇa Vibhāg Aṇḍamān tathā Nikobār Praśāsan, and Central Institute of Indian Languages [23 unnumbered and 192 numbered pages].
- Fontana, N. 1792. 'On the Nicobar isles and the fruit of the melli'. *Asiatick Researches* III. 149-163.
- Fuller, Dorian Q. 2005a. 'Ceramics, seeds and culinary change in prehistoric India'. *Antiquity* 79. 761-777.

- . 2005b. 'Formation processes and palaeoethnobotanical interpretation in South Asia'. *Journal of Interdisciplinary Studies* 2.1: 93-115.
- . 2005c. 'The Ganges on the world Neolithic map: The significance of recent research on agricultural origins in northern India'. *Prāgdhārā* 16. 187-206.
- . 2006a. 'Agricultural origins and frontiers in South Asia: A working synthesis'. *Journal of World Prehistory* 20. 1-86.
- . 2006b. 'Dung mounds and domesticators: Early cultivation and pastoralism in Karnataka', pp. 117-127 in Catherine Jarrige and Vincent Lefèvre, eds., *South Asian Archaeology 2001: Proceedings of the Sixteenth International Conference on South Asian Archaeology, European Association of South Asian Archaeologists, Paris, 2-6 July 2001*. Paris: Éditions Recherche sur les Civilisations.
- . 2006c. 'Silence before sedentism and the advent of cash crops: A status report on early agriculture in South Asia from plant domestication to the development of political economies (with an excursus on the problem of semantic shift among millets and rice)', pp. 175-213 in Osada Toshiki, ed., *Proceedings of the pre-Symposium of the Research Institute for Humanity and Nature and 7th Ethnogenesis of South and Central Asia Round Table Harvard-Kyoto Roundtable*. Kyoto: Research Institute for Humanity and Nature.
- . 2007a. 'Contrasting patterns in crop domestication and domestication rates: Recent archaeobotanical insights from the Old World'. *Annals of Botany* 100. 1-22.
- . 2007b. 'Non-human genetics, agricultural origins and historical linguistics in South Asia', pp. 393-443 in Michael D. Petraglia and Bridget Allchin, eds., *The Evolution and History of Human Populations in South Asia: Interdisciplinary Studies in Archaeology, Biological Anthropology, Linguistics and Genetics*. Dordrecht: Springer.
- Fuller, Dorian Q.; Harvey, Emma; and Qin, Ling. 2007. 'Presumed domestication? Evidence for wild rice cultivation and

- domestication in the fifth millennium BC of the Lower Yangtze region'. *Antiquity* 81. 316-331.
- Harvey, Emma L.; Fuller, Dorian Q.; Mohanty, R.K.; and Mohanta, Basanta. 2006. 'Early agriculture in Orissa: Some archaeobotanical results and field observations on the Neolithic'. *Man and Environment XXXI* (2). 21-32.
- Haudricourt, Andre-Georges. 1954. 'Introduction à la phonologie historique des langues miao-yao', *Bulletin de l'École Française d'Extrême-Orient* 44. 555-574.
- Haudricourt, André-Georges, and Hédin, Louis. 1987. *L'homme et les plantes cultivées*. Paris: Éditions A.-M. Métailié.
- Hazarika, Manjil. 2005. *Neolithic culture of Northeast India with Special Reference to the Origins of Agriculture and Pottery*. Pune: Unpublished Master's thesis, Deccan College.
- . 2006a. 'Neolithic culture of northeast India: A recent perspective on the origins of pottery and agriculture'. *Ancient Asia* 1. 25-43.
- . 2006b. 'Understanding the process of plant and animal domestication in northeast India: A hypothetical approach'. *Asian Agri-History* 10.3 : 203-212.
- Justin, Anstice. 1990. *The Nicobarese* (The Anthropological Survey of India Andaman and Nicobar Tribe Series). Calcutta: Seagull Books.
- Krishan, G. 1986. 'A note on the anthropometry of the Shompen of Great Nicobar'. *Human Science* 35. 232-236.
- Kumar, Vikrant, Arimanda; Reddy, N.S.; Babu, Jagadeesh P.; Rao, Tipiriseti N.; Langstieh, Banrida T.; Thangaraj, Kumarasamy; Reddy, Alla G.; Singh, Lalji; and Reddy, Battini M. 2007. 'Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations', *BMC Evolutionary Biology* 7:47 doi:10.1186/1471-2148-7-47.
- Lal, Parmanand. 1977. *Great Nicobar Island: A Study in Human Ecology*. Calcutta: Anthropological Survey of India.

- Man, Edward Horace. 1872. 'List of words of the Nicobar language as spoken at Camorta, Nancowry, Trinkutt and Katschal', *Journal of the Asiatic Society of Bengal*, XLI (I): 1-7.
- . 1878. 'On the arts of the Andamanese and Nicobarese'. *The Journal of the Anthropological Institute of Great Britain and Ireland* 7. 451-469.
- . 1886. 'A brief account of the Nicobar Islanders, with special reference to the inland tribe of Great Nicobar'. *Journal of the Anthropological Institute of Great Britain and Ireland*, XV. 428-451.
- . 1889a. 'The Nicobar Islanders. Part I'. *The Journal of the Anthropological Institute of Great Britain and Ireland* 18. 354-394.
- . 1889b. *A Dictionary of the Central Nicobarese Language (English-Nicobarese and Nicobarese-English), with Appendices Containing a Comparison of Synonymous Words in the Remaining Nicobarese Forms and Other Matters, Preceded by Notes on the Grammar of the Central Form*. London: W.H. Allen.
- . 1894. 'Nicobar pottery'. *The Journal of the Anthropological Institute of Great Britain and Ireland* 23. 21-27.
- . 1923b [posthumous]. *The Nicobar Islands and their People*. Guildford: Royal Anthropological Institute of Great Britain and Ireland.
- Mason, Francis. 1854. 'The Talaing language.' *Journal of the American Oriental Society*, IV. 277-289.
- . 1860. *Burmah, Its people and Natural Productions, or Notes on the Nations, Fauna, Flora and Minerals of Tenasserim, Pegu and Burmah*. Rangoon: Thomas Stowe Ranney.
- Meerwarth, A.M. 1919. *The Andamanese, Nicobarese and Hill Tribes of Assam*. Calcutta: Superintendent of Government Printing, India.
- Nandan, Anshu Prokash. 1993. *The Nicobarese of Great Nicobar: An Ethnography*. New Delhi: Gyan Publishing House.

- Nasu Hirō, Arata Momohara, Yoshinori Yasuda and Jiejun He. 2006. 'The occurrence and identification of *Setaria italica* (L.) P. Beauv. (foxtail millet) grains from the Chengtoushan site (ca. 5800 cal B.P.) in central China, with reference to the domestication centre in Asia', *Vegetation History and Archaeobotany* 16 (6): 481-494.
- Nevskij, Nikolaj Aleksandrovič. 1960 [posthumous]. *Tangutskaja filologija: Issledovanija i slovar'* (2 vols.). Moscow: Izdatel'stvo Vostočnoj Literatury.
- Niederer, Barbara. 1998. *Les langues hmong-mjen (miao-yao): Phonologie historique*. München: Lincom Europa.
- Osada, Toshiki. 1995. *Mundajin no Nōkō Bunka to Shokuji Bunka: Minzoku Gengo gaku teki Kōsatsu* ['The rice and food culture of Munda in eastern India: An ethnolinguistic study']. Kyoto: Kokusai Nihon Bunka Kenkyū Sentā.
- Pal, J.N. 1990. 'The early farming culture of northern India'. *Bulletin of Deccan College Post-Graduate and Research Institute* 49. 297-304.
- Palanichamy, Malliya Gounder, Suraksha Agrawal, Yong-Gang Yao, Qing-Peng Kong, Chang Sun, Faisal Khan, Tapas Kumar Chaudhuri and Ya-Ping Zhang. 2006. 'Comment on "Reconstructing the Origin of Andaman Islanders"'. *Science* 311. 470a.
- Pinnow, Heinz-Jürgen. 1959. *Versuch einer historischen Lautlehre der Kharia-Sprache*. Wiesbaden: Otto Harrassowitz.
- Poloni, Estella Simone, et al. 1997. 'Human genetic affinities for Y chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics', *American Journal of Human Genetics* 61: 1015-1035 (cf. the erratum published in 1998 in *American Journal of Human Genetics* 62. 1267).
- Poloni, Estella Simone, et al. 2000. 'Languages and genes: Modes of transmission observed through the analysis of male-specific and female-specific genes', pp. 185-186. In: Jean-Louis Dessalles and Laleh Ghadakpour (eds.), *Proceedings: Evolution of*

- Language, 3rd International Conference 3-6 April 2000*. Paris: École Nationale Supérieure des Télécommunications.
- Purnell, Herbert C. Jr. 1970. *Toward a Reconstruction of Proto Miao-Yao*. Ithaca: Cornell University Ph.D. dissertation.
- Radhakrishnan, R. 1981. *The Nancowry Word, Phonology, Affixal Morphology and Roots of a Nicobarese Language* (Current Inquiry into Language and Linguistics 37). Carbondale and Edmonton: Linguistic Research.
- Ratliff, Martha. 2004. 'Vocabulary of environment and subsistence and in the Hmong-Mien protolanguage', pp. 147-165. In: Nicholas Tapp, Jean Michaud, Christian Culas and Gary Yia Lee (eds.), *Hmong/Miao in Asia*. Chiang Mai: Silkwoon Books.
- Revel, Nicole. 1988. *Le riz en Asie du sud-est* (3 vols.). Paris: Éditions de l'École des Hautes Études en Sciences Sociales.
- Rizvi, S.N.H. 1990. *The Shompen: A Vanishing Tribe of the Great Nicobar Island* (The Anthropological Survey of India Andaman and Nicobar Tribe Series). Calcutta: Seagull Books.
- de Roepstorff, Frederik Adolph. 1870. 'A short list of Andamanese test words'. *Proceedings of the Asiatic Society of Bengal* (June 1870): 178-180.
- . 1875. *Vocabulary of Dialects spoken in the Nicobar and Andaman Isles, with a short account of the natives, their customs and habits, and of previous attempts at colonisation*. Calcutta: Asiatic Society of Bengal. [114 pp. with fold-out map].
- . 1884 [posthumous]. *A Dictionary of the Nancowry Dialect of the Nicobarese Language* (edited by the author's widow Hedevig Christiane Willemoës de Roepstorff). Calcutta: Home Department Press.
- Rosen, David. 1839. *Erindringer fra mit Ophold paa de Nicobarske Øer med en kort Skildring af Øernes naturlige Beskaffenhed, og deres Indbygges Ejendommelighed*. Copenhagen.
- Schmidt, Wilhelm. 1904. *Grundzüge einer Lautlehre der Khasi-Sprache in ihren Beziehungen zu derjenigen der Mon-Khmer-Sprachen, mit einem Anhang: Die Palaung-, Wa- und Riang-*

- Sprachen der mittleren Salwin.* München: Kaiserliche Akademie.
- . 1906. 'Die Mon-Khmer Völker, ein Bindeglied zwischen Völkern Zentral-Asiens und Austronesiens,' *Archiv für Anthropologie*, Neue Folge, V: 59-109.
- Sharma, G.R., Misra, V.D.; Mandal, D.; Misra, B.B.; and Pal, J.N. 1980. *Beginnings of Agriculture: From Hunting and Food Gathering to Domestication of Plants and Animals.* Allahabad: Abhinav Prakashan.
- Shorto, Harry. 2006 [posthumous]. *A Mon-Khmer Comparative Dictionary* (edited by Paul Sidwell, Doug Cooper and Christian Bauer). Canberra: Pacific Linguistics.
- Singh, Balwant. 1988. *Census of India, 1981, Series 24: Andaman and Nicobar Islands, Part IX: Special Tables for Scheduled Tribes.* Delhi: Controller of Publications
- Singh, K.S. 1994a. *The Scheduled Tribes (People of India, Volume III).* Madras: Anthropological Survey of India and Oxford University Press.
- . 1994b. *Andaman and Nicobar Islands (People of India, Volume XII).* Madras: Anthropological Survey of India.
- Singh, Simron Jit. 2003. *In The Sea of Influence: A World System Perspective of the Nicobar Islands* (Lund Studies in Human Ecology 6). Lund: Lunds Universitet.
- Sreenathan, M. 2001. *The Jarawas: Language and Culture.* Calcutta: Anthropological Survey of India.
- Temple, Sir Richard C. 1903. *Census of India, 1901, Vol. III: The Andaman and Nicobar Islands: Report on the Census.* Calcutta: Office of the Superintendent of Government Printing, India.
- . 1909. *Andaman and Nicobar Islands (Provincial Series, Imperial Gazetteer of India).* Calcutta: Superintendent of Government Printing.
- . 1930. 'Edward Horace Man' [obituary], *Man* 30. 11-12.

- Tewari, Rakesh, Srivastava, R.K.; Singh, K.K.; Saraswat, K.S.; and Singh, I.B. 2002. 'Preliminary report of the excavation at Lahuradewa, District Sant Kabir Nagar, U.P. 2001-2002: Wider archaeological implications'. *Prāgdhāra, Journal of the Uttar Pradesh Archaeological Department*, 13: 37-76.
- Thangaraj, Kumarasamy, Singh, Lalji; Reddy, Alla G.; Raghavendra Rao, V.; Sehgal, Subhash C.; Underhill, Peter A.; Pierson, Melanie; Frame, Ian G.; and Hagelberg, Erika. 2003. 'Genetic affinities of the Andaman Islanders, a vanishing human population'. *Current Biology* 13: 86-93.
- Thangaraj, Kumarasamy, Chaubey, Gyaneshwer; Kivisild, Toomas; Reddy, Alla G.; Singh, Vijay Kumar; Rasalkar, Avinash A.; and Singh, Lalji. 2005. 'Reconstructing the Origin of Andaman Islanders'. *Science* 308. 996.
- Trivedi, Rajni, Sitalaxmi, T.; Banerjee, Jheelam; Singh, Anamika; Sircar, P.K.; and Kashyap, V.K. 2006. 'Molecular insights into the origins of the Shompen, a declining population of the Nicobar archipelago'. *Journal of Human Genetics* 51.3: 217-226.
- Yasuda Yoshinori. 2002. 'Origins of pottery and agriculture in East Asia', pp. 119-142. In: Yoshinori Yasuda (ed.), *The Origins of Pottery and Agriculture*. New Delhi: Lusre Press and Roli Books, for the International Center for Japanese Studies.