RESEARCH ARTICLE

Cranio-morphometric and aDNA corroboration of the Austronesian dispersal model in ancient Island Southeast Asia: Support from Gua Harimau, Indonesia

Hirofumi Matsumura[1], Ken-ichi Shinoda[2], Truman Shimanjuntak[3], Adhi Agus Oktaviana[4], Sofwan Noerwidi[4], Harry Octavianus Sofian[4], Dyah Prastiningtyas[3], Lan Cuong Nguyen[5], Tsuneo Kakuda[6], Hideaki Kanzawa-Kiriyama[2], Noboru Adachi[6], Hsiao-chun Hung[7], Xuechun Fan[8], Xiujie Wu[9], Anna Willis[10], Marc F. Oxenham[11]*

1 School of Health Science, Sapporo Medical University, Sapporo, Japan, 2 Department of Anthropology, National Museum of Nature and Science, Tokyo, Japan, 3 Center of Prehistory and Austronesian Studies, Jakarta, Indonesia, 4 The National Research Centre of Archaeology, Jakarta, Indonesia, 5 Institute of Archaeology, Vietnam Academy of Social Science, Hanoi, Vietnam, 6 Department of Legal Medicine, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Kofu, Japan, 7 Department of Archaeology and Natural History, Australian National University, Canberra, Australia, 8 Fujian Museum, Fuzhou, China, 9 Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing, China, 10 College of Arts, Society and Education, James Cook University, Townsville, Australia, 11 School of Archeology and Anthropology, Australian National University, Canberra, Australia

* marc.oxenham@anyu.edu.au

Abstract

The Austronesian language is spread from Madagascar in the west, Island Southeast Asia (ISEA) in the east (e.g. the Philippines and Indonesian archipelagoes) and throughout the Pacific, as far east as Easter Island. While it seems clear that the remote ancestors of Austronesian speakers originated in Southern China, and migrated to Taiwan with the development of rice farming by c. 5500 BP and onto the northern Philippines by c. 4000 BP (the Austronesian Dispersal Hypothesis or ADH), we know very little about the origins and emergence of Austronesian speakers in the Indonesian Archipelago. Using a combination of cranial morphometric and ancient mtDNA analyses on a new dataset from Gua Hairmau, that spans the pre-Neolithic through to Metal Period (5712—5591cal BP to 1864—1719 cal BP), we rigorously test the validity of the ADH in ISEA. A morphometric analysis of 23 adult male crania, using 16 of Martin's standard measurements, was carried out with results compared to an East and Southeast Asian dataset of 30 sample populations spanning the Late Pleistocene through to Metal Period, in addition to 39 modern samples from East and Southeast Asia, near Oceania and Australia. Further, 20 samples were analyzed for ancient mtDNA and assigned to identified haplogroups. We demonstrate that the archaeological human remains from Gua Harimau cave, Sumatra, Indonesia provide clear evidence for at least two (cranio-morphometrically defined) and perhaps even three (in the context of the ancient mtDNA results) distinct populations from two separate time periods. The results of these analyses provide substantive support for the ADH model in explaining the origins and population history of ISEA peoples.

Introduction

The Austronesian language is spread from Madagascar in the west, Island Southeast Asia
(ISEA) in the east (e.g. the Philippines and Indonesian archipelagoes) and throughout the
Pacific, as far east as Easter Island. Austronesian language dispersal models have been pro-
posed by Blust and Bellwood [1–3] and Bellwood has gone on to test these using archaeological
evidence as a proxy for human movement between 5000 and 1000 years ago [1, 4–7]. The most
widely recognized model that the remote ancestors of Austronesian speakers originated in
Southern China, and migrated to Taiwan with the development of rice farming by c. 5500 BP
and onto the northern Philippines by c. 4000 BP, is now broadly accepted [6, 7].

The subsequent Austronesian language speaking dispersals, from the Neolithic through to
later Metal periods, throughout Island Southeast Asia, including Malaysia and Indonesia, and
into the Pacific, referred to as the "Out of Taiwan" model or Austronesian Dispersal Hypothe-
sis (ADH), are similarly well attested to archaeologically [8]. Notwithstanding, unlike the case
in Mainland Southeast Asia (MSEA) [9, 10], human skeletal remains have not played a sub-
stantive role in human mobility debates in ISEA. The principle reason is the hitherto poor
preservation of human remains from key localities in the region. For instance, Niah Cave, is
the largest Neolithic mortuary site in ISEA, from which more than a hundred human skeletal
remains have been reported, including the earliest dated in the region: the 'deep skull' [11, 12].
Unfortunately, the very poor preservation of these remains, particularly the crania, has ham-
pered morphological analysis.

Recent excavations at the Gua Harimau cave site in southeastern Sumatra, provide an
assemblage spanning the pre-Neolithic to Metal periods and an opportunity to assess the ADH
model. The aim of this paper is to test the ADH using a combination of cranial morphometrics
and ancient mtDNA analyses on the Gua Harimau remains. Comparisons will be made with
appropriate modern and archaeological samples to better understand ancient patterns of
genetic exchange and human mobility patterns in the region.

Gua Harimau in context

The cave site of Gua Harimau is located in Padang Bindu, Oku district, in southeastern Suma-
tra, Indonesia (Fig 1). The cave, which formed several tens of meters above the present alluvial
plain, opens towards the southeast. The width of the main entrance is c. 30m and the average
horizontal depth is c. 15m. Since 2012, a substantive area of the floor of the single chambered
cave has been excavated, recovering 84 individual human skeletons dating from the pre-Neo-
lithic through to the Neolithic, Bronze and Iron Ages (Fig 2) or 5712—5591cal BP to 1864—
1719 cal BP (Table 1). Some level of continuity in artefact types and floral/faunal remains from
the Neolithic through to Bronze and Iron ages suggests continuity in occupation over this
period of time.

The main feature differentiating the Metal Period burials is the presence of bronze and/or
iron artefacts. Neolithic burials can be identified by way of characteristic paddle impressed
and incised ceramics. Somewhat intriguingly, cord marked ceramics were also identified
among some Neolithic burials. The pre-Neolithic layer, dated to between 5712–4434 cal BP, is
characterized by the presence of unifacial pebble tools, similar to the almond-shaped tools
often found in Hoabinhian cultural assemblages in MSEA, as well as a significant number of
flakes. The deeper late Pleistocene layer, dated using charcoal to 13,055 +/- 120 ^{14}C years [13]
or 14,061–13,312 cal BP (95.4%) (OxCal v4.3, IntCal 13, Bronk Ramsey [14]), is characterized
by numerous flakes originating from a range of materials, including obsidian. No human
remains have been recovered from this basal layer.

Fig 1. Location of Gua Harimau (star) in Southeast Sumatra, Indonesia.

https://doi.org/10.1371/journal.pone.0198689.g001

Fig 2. Views from Gua Harimau. Left: Metal Period (Bronze-Iron Age) extended burials (note, co-authors, from left to right: Nguyen Lan Cuong, Daya Prastingus; Hirofumi Matsumura, and Sofwan Neruwdi); Right Upper: Metal Period Burial No. 23 and 24; Right Lower: pre-Neolithic Period Burial No. 79.

Materials and methods

Cranial morphometric analysis

The excavation and analysis of Gua Harimau Cave was carried out with the permission of Dr H. Kuryana Azis (Head of the Local Government), Mr Pak Paisol (Head of the Local Tourism and Cultural Office) and Padang Bindu village. The male crania from Gua Harimau were analyzed to be consistent with standard recording protocols and generally male dominated comparative data sets available. Of the original 84 individuals, 23 were suitable for morphometric analysis and these are divided into Early and Late Gua Harimau (see representative crania in Fig 3). The Early Gua Harimau sample consists of two individuals buried in a flexed position, assigned to the pre-Neolithic period (individuals No. 74 and 79). Individual No. 74 was directly dated to between 4785–4434 cal BP. No. 79 could not be directly dated, but it was stratigraphically located between individuals No. 74 and No. 80 which was directly dated to 5712—5591cal BP. The Late Gua Harimau sample includes 20 individuals buried in a supine, extended position and one disturbed burial, the position of which could not be determined (individual No. 53), assigned to the Metal period, dated to between 2424–2333 and 1864–1719 cal BP. While three Neolithic burials could be dated (see Table 1), ranging between 3136–2953 and 2711–2379 cal BP, no Neolithic crania were sufficiently preserved to enable morphometric analysis.

The cranial data set included a subset of 16 measurements (Martin's method numbers M1, M8, 9, M17, M43(1), M43c, M45, M46b, M46c, M48, M51, M52, M54, M55, M57, M57a), that represent the most commonly available measurements among the comparative samples. The

Table 1. Results of AMS dating for human remains from the site of Gua Harimau.

ID	Lab. Code	Sample	^{14}C age (BP)	±	cal BP (95.4% range)	Period
27	BTN12002	bone	1852	20	1864–1719	Metal
3	BTN12003	bone	1880	20	1879–1737	Metal
56	BTN12004	bone	1910	20	1896–1820	Metal
4	BTN12008	bone	1925	20	1923–1823	Metal
8	BTN12009	bone	1995	20	1992–1896	Metal
58	BTN12005	bone	2015	20	2003–1899	Metal
13	BTN12001	bone	2048	20	2110–1945	Metal
2	BTN12010	bone	2150	25	2304–2046	Metal
54	BTN13023	bone	2190	20	2309–2142	Metal
43	BTN13022	bone	2290	20	2352–2206	Metal
11	WK 37248	tooth dentin	2290	20	2352–2206	Metal
40	BTN12007	bone	2305	25	2356–2206	Metal
18	BTN13035	bone	2350	20	2424–2333	Metal
53	IAAA-143261	tooth dentin	2463	26	2711–2379	Neolithic
44	BTN12006	bone	2575	30	2760–2518	Neolithic
26	IAAA-170200	tooth dentin	2890	20	3136–2953	Neolithic
74	IAAA-143262	tooth dentin	4054	28	4785–4434	Pre-Neolithic
80	Beta 450669	bone	4910	30	5712–5591	Pre-Neolithic

BTN Laboratorium Batan Indonesia; WK Waitako, New Zealand. IAAA Institute of Accelerator Analysis Ltd. Japan; Beta Beta Analytic Inc. USA

^{14}C ages are calibrated with OxCal v4.3, IntCal 13, Bronk Ramsey

https://doi.org/10.1371/journal.pone.0198689.t001

cranio-metric affinities of the comparative samples were assessed using Q-mode correlation coefficients [15], based on the above 16 cranial measurements. The comparative archaeological cranial series are listed in Tables 2 and 3 and included a total of 64 individuals from both archaeological and contemporary contexts in East, and Southeast Asia and the Pacific. The dataset includes individuals from the late Pleistocene, early to mid-Holocene, Neolithic

Individual No. 79   Individual No.48
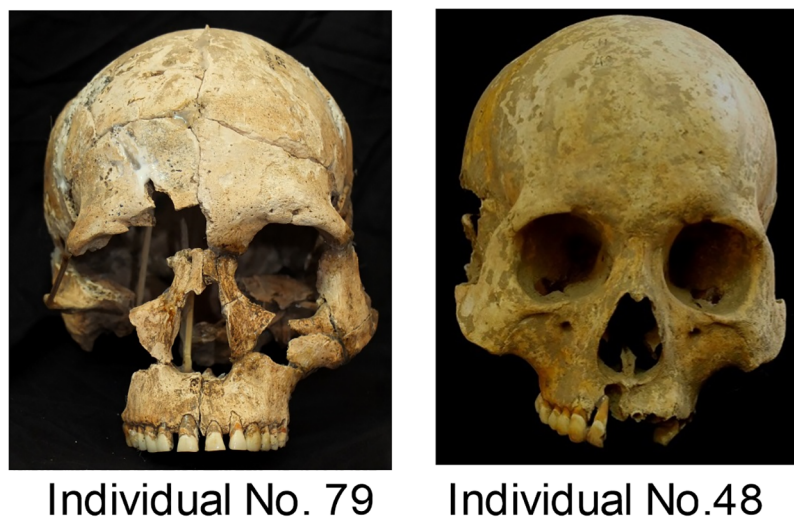
Fig 3. Representative pre-Neolithic (Left, individual No. 79) and Metal period (Right, individual No. 48) crania from Gua Harimau.

https://doi.org/10.1371/journal.pone.0198689.g003

Table 2. Prehistoric samples used in the present study.

Sample	Country	Period	District / Remarks / References	Data Source[a] / Storage[b] (n = sample size)
Late Pleistocene (except Hoabinhian)				
Liujiang		Late Pleistocene	Individual (no. PA89)	[17]; M43(1),43c,46b,46c, 57,57a by H.M. (cast).
Early Holocene Hoabinhian/Mesolithic				
Qihedong	China	Mesolithic (c. 9,500BP)	Individual (no.3), Site in Fujian Province. [18, 19]	H.M. / IVPP
Zengpiyan	China	Mesolithic (c. 8,000BP)	Guilin, Guangxi Region [20]	[20]
Lang Gao	Vietnam	Hoabinhian	Averages of two individuals (nos. 17 and 19) [21–23]	H.M. / MHO
Lang Bon	Vietnam	Hoabinhian (c. 8,000–7,000BP)	Individual (no number) [21–23]	H.M. / MHO
Mai Da Nuoc	Vietnam	Hoabinhian (c. 8,000BP)	Individual (no. 84MDNM1) [24]	H.M. / IAH
Hoabinhian	Vietnam	Hoabinhian (c. 10,000–7,000BP)	Specimens including other fragmental remains from above 3 sites and an incomplete skull of Mai Da Dieu (no.16) [24, 25]	H.M. / IAH, MHO (n = 6)
Bac Son	Vietnam	Epi-Hoabinhian (c. 8,000–7,000BP)	Pho Binh Gia, Cua Git, Lang Cuom, and Dong Thuoc [26]	H.M. / MHO (n = 8)
Con Co Ngua	Vietnam	Mesolithic (Da But Culture, c. 6,000BP)	Sites in Than Hoa Province [27, 28] (1980's excavated series including an individual from the site of Da But [27–29])	[27]; M5,43(1),43c,46b,46c, 57, 57a by H.M./ IAH (n = 7)
Gua Cha	Malaysia	Hoabinhian (c.8,000–6,000 BP)	Individual (no. H12), Site in Kelantan Province [30]	H.M. / CAM
Neolithic				
Weidun	China	Neolithic (Majiabang Culture, c.7,000–5,000 BP)	Jiangsu Province [31]	[31]
Xitou	China	Neolithic (Majiabang Culture, c.7,000–5,000 BP)	Site in Fujian Province [32, 33]	H.M. / FPM (n = 3)
Tanshishan	China	Neolithic (Majiabang Culture, c.7,000–5,000 BP)	Site in Fujian Province [34, 35]	H.M. / FPM (n = 7)
Hemudu	China	Neolithic (c. 6,300 BP, Hemudu Culture)	Individual (M23), Site in Zhejiang Province, Yangzi Delta region [36]	H.M. / HEMSM
Man Bac 1	Vietnam	Late Neolithic (c. 3,800–3,500 BP)	Ninh Binh Province (immigrant group) [16, 37]	H. M. / IAH
Man Bac 2	Vietnam	Late Neolithic (c. 3,800–3,500 BP)	Ninh Binh Province (indigenous group) [16, 37]	H. M. / IAH
An Son	Vietnam	Late Neolithic (c.3,800 BP)	Long An Province [38–40]	H.M. / LAPM (n = 4)
Ban Chiang	Thailand	Neolithic-Bronze Age (c. 4,100–2,300 BP)	Site in Udon Thani Province [41, 42]	[40]; M43(1),43c,46b,46c, 57,57a by H.M. / UH, SAC (n = 15)
Khok Phanom Di	Thailand	Late Neolithic (c. 3,800–3,500 BP)	Site in Chonburi Province [43, 44]	[45] H.M. / FAD (n = 19)
Tam Hang and Tam Pong	Laos	Neolithic (Tam Hang c. 3,500 BP, Tam Pong unknown abolute date)	[42, 46, 47] (C14 date recorded in 47 was later corrected to more modern by T. Higham)	H.M. / MHO (n = 3)
Neolithic Baikal	Russia	Neolithic [48]		[49]
Jomon	Japan	Neolithic (skeletal serie used: c. 5,000–2,300 BP) [50]	from whole Japan	[51, 52]
Bronze—Iron Age				
Anyang	China	Yin (Shan) Period (c. 1,500–1,027 BC)	Henan Province [53]	[54]; M43(1),43c,46b,46c, 57,57a by H.M. / AST (n = 26)
Giong Ca Vo	Vietnam	Iron Age (c. 300–0 BC)	Site in Can Gio District, Ho Chi Minh City [55]	[25]; M43(1),43c,46b,46c, 57,57a by H.M. / HCHM (n = 4)
Hoa Diem	Vietnam	Iron Age (123var yr AD-243 cal yr AD (IAAA-101437)	Khanh Hoa Province [56]	H.M. / KHPM (n = 6)
Dong Son	Vietnam	Dong Son Period (c. 1,000 BC-AD 300)	Sites of Dong Son Culture [57]	[57]; M43(1),43c,46b,46c, 57,57a by H.M. / IAH, CSPH (n = 21)
Rach Rung	Vietnam	2800 BP	Site in Moc Hoa District, Long An Province [58]	H.M. / LAPM (n = 2)
Jiangnan	China	Eastern Zhou—Former Han Periods (770 BC-AD 8)	Sites in Jiangnan Province along the Lower Basin of Yangtze River [31]	[31]

(Continued)

Table 2. (Continued)

Sample	Country	Period	District / Remarks / References	Data Source[a] / Storage[b] (n = sample size)
Jundushan	China	Spring and Autumn Period (c. 500 BC)	Site in Yanqing Prefecture near Beijing [59]	H.M. / PKU (n = 27)
Yayoi	Japan	Yayoi Period (c. 800 BC—AD 300)	Various sites in Northern Kyushu and Yamaguchi Districts [60]	[60]

Data source[a]: H.M. = the present first author Hirofumi Matsumura.

Storage[b]: AST = Academia Sinica of the Republic of China, Taipei; BMNH = Department of Paleontology, Natural History Museum, London; CAM = Division of Biological Anthropology, University of Cambridge; CSPH = Center for Southeast Asian Prehistory, Hanoi; FAD = the Fine Arts Department, Pimai; FPM = Fujian Provincial Museum, Fujian; HCHM = Ho Chi Minh Historical Museum, Ho Chi Minh; HEMSM = Hemudu Site Museum, Hemudu; IAH = Department of Anthropology, the Institute of Archaeology, Hanoi; IVPP = Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing; KHPM = Khanh Hoa Provincial Museum, Vietnam, Nha Trang; LAPM = Long An Provincial Museum, Tan An; MHO = Laboratoire d'Anthropologie Biologique, Musée de l'Homme, Paris; PKU = School of Archaeology and Museology, Peking University, Beijing; SAC = Princess Maha Chakri Sirindhorn Anthropology Centre, Bangkok; UH = Department of Anthropology, University of Hawaii, UNLV: Department of Anthropology, University of Nevada, Las Vegas

https://doi.org/10.1371/journal.pone.0198689.t002

(defined as early farming populations [16]), and Bronze and Iron Ages through to the proto-Historic and Historic periods. Space precludes a review of each sample in the dataset, however, the references in Tables 2 and 3 provide details on the majority of samples used in this analysis.

To aid the interpretation of any phenotypic affinities between the samples, Neighbor Net Split tree diagrams were generated using the software package Splits Tree Version 4.0, applied to the distance (1-r) matrix of Q-mode correlation coefficients (r) [68].

Mitochondrial DNA analysis

Tooth enamel forms a natural barrier to exogenous DNA contamination, and DNA recovered from teeth appears to lack most inhibitors of the enzymatic amplification of ancient DNA (aDNA) [69]. In addition, because recent research reveals that the temporal bone is an ideal region from which to analyze aDNA, samples were taken from both teeth and the temporal bone in this analysis [70]. In total, 20 samples (two pre-Neolithic, three Neolithic and 15 Metal Period) from well preserved teeth and temporal bones were selected for DNA analysis. A list of all samples used in this analysis are presented in Table 4 (see Results) along with their determined haplogroups.

Authentication methods for DNA extraction. Mitochondrial DNA (mtDNA) analyses were performed at the National Museum of Nature and Science, Tokyo, Japan, and at Yamanashi University, which have laboratories dedicated to aDNA analysis. Standard protocols were employed to avoid contamination, including the separation of pre- and post-PCR experimental areas, UV irradiation of equipment and benches, negative extraction, and PCR controls [71].

To prevent contamination from post-excavation handling, all samples were rinsed with DNA-decontamination agents (DNAaway; Molecular Bio Products, San Diego, CA, USA) or 13% bleach solution (Nacalai Tesque Inc., Kyoto, Japan), and then washed thoroughly with distilled water before drying. Next, tooth samples were encased in Exafine silicone rubber (GC, Tokyo, Japan). The tip of the root of each tooth was removed via a horizontal cut using a cutting disk, and the dentin within the dental pulp cavity was powdered and removed through the root tip using a dental drill [72]. Powdered samples were then decalcified using 0.5 M EDTA (pH 8.0) at room temperature overnight, samples were then decalcified for a further 48 hours in a fresh EDTA solution. Decalcified samples were lysed in 500 μl of Fast Lyse (Genetic

Table 3. Data sources of comparative modern population samples.

Population	Cranial metrics	Facial chord & subtense[a]	Remark (M = Martin's number)	Storage[b]
Aeta Negrito Philippines	H.M. (n = 11)	H.M. (n = 11)	-	MHO
Andaman Islands	[61]	H.M. (n = 5)	M9,51,55 by H.M (n = 22)	BMNH, CAM
Australian Aborigines	[47]	H.M. (n = 21)	-	BMNH
Bunun Taiwan	[62]	H.M. (n = 16)	M45,48,51,55 by H.M. (n = 23)	NTW
Cambodia	H.M. (n = 12)	H.M. (n = 12)	-	MHO
Celebes Island Indonesia	[63]	[48]	M17,45,48,51 by H.M. (n = 6)	BMNH
Hainan Island China	[61]	H.M. (n = 24)	M48,51,55 by H.M. (n = 24)	NTW
South China	H.M. (n = 7)	H.M. (n = 7)	Hong Kong	CAM
Japan	[47]	[48]	-	
Java Island Indonesia	[63]	[48]	M17,45,48,51 by H.M. (n = 20)	BMNH, CAM
Laos	[53]	H.M. (n = 10)	-	MHO
Loyalty Islands	H.M. (n = 17)	H.M. (n = 18)	-	MHO
Melanesia	[47]	[48]	Fiji, Tongans; New Hebrides; New Guinea	-
Myanmar	[63]	[48]	M17,45,48,51 by H.M. (n = 21)	BMNH
New Britain Island	H.M. (n = 20)	H.M. (n = 19)	-	CAM
New Guinea Tolai	[47]	H.M. (n = 26)	M9,48,51 by H.M. (n = 20)	USYD, CAM
Nicobar Islands	H.M. (n = 13)	H.M. (n = 9)	-	CAM
North China 1	[47]	[48]	Kiling Prov.	
North China 2	[47]	[48]	Manchuria Prov.	
Philippines	[64]	H.M. (n = 8)	-	NMP
Seman Negrito Malaysia	H.M. (n = 1)	H.M. (n = 1)	-	BMNH
South Moluccas Islands Indonesia	[63]	[48]	M17,45,48,51 by H.M. (n = 4)	-
Sumatra Island Indonesia	[63]	[48]	M17,45,48,51 by H.M. (n = 8)	BMNH, CAM
Thai	[65]	[48]	-	
Veddah Sri Lanka	H.M. (n = 2)	H.M. (n = 2)	-	CAM
Vietnam	H.M. (n = 27)	H.M. (n = 27)	-	MHO
Okhotsk Japan	[66]	[66]	AD c.400-1,000	
Hokkaido Ainu Japan	[67]	[45]		
Mongol	[67]	[45]		
Aleut, Asian Inuit, Buryat, Chukchi, Ekven, Nanay, Negidal, Nivkh, Oroch, Troitskoe, Ulch, Yakut, Yukagir	[67]	[45]	Russia	

Facial chord and subtense[a]: (M43(1) = frontal chord (FC); M43c = frontal subtense (FS); M57 = simotic chord (SC); M57a = simotic subtense (SS);

M46b = zygomaxillary chord (ZC); M46c = zygomaxillary subtense (ZS); M = Martin's cranial measurment number),

Storage[b]: institutions of materials studied by H.M. (H. Matsumura) BMNH = Department of Paleontology, Natural History Museum, London; MHO = Laboratoire d'Anthropologie Biologique, Musée de l'Homme, Paris; NTW = Department of Anatomy, National Taiwan University, NMP = Department of Archaeology, National Museum of the Philippines, Manila; USYD = Department of Anatomy, University of Sydney.

ID, Fairfield, IA, USA) with 30 µl of 20 mg/ml Proteinase K at 60°C for four hours. DNA was extracted from lysate using a FAST ID DNA Extraction Kit (Genetic ID) in accordance with the protocol described by Adachi et al. [73].

Data analysis and genotyping of mtDNA. mtDNA SNPs were detected using the amplified product length polymorphism (APLP) method [74, 75]. This method has been applied in

Table 4. Sample used for DNA extraction and the result of the DNA analysis from the Gua Harimau site.

Lab No	Individual No.	Sample	Dating Comments	Period	Haplogroup by APLP
1	No.1	Right temporal bone	Metal Period layer	Metal	E
2	No.2	Maxilla, Right, M2	2304–2046 cal BP	Metal	N.D.
3	No.3	Mandible, Right, M1	1879–1737 cal BP	Metal	N.D.
4	No.4	Maxilla, Right, M1	1923–1823 cal BP	Metal	N.D.
5		Right temporal bone	double burial with No. 3		N.D.
6	No.8	Mandible, Right, M2	1992–1896 cal BP	Metal	N.D.
7		Maxilla, Right, M3			N.D.
8	No.9	Mandible, Right, M2	Earlier layer than No. 13 (2110–1945 cal BP)	Metal	N.D.
9		Left temporal bone			B4a
10	No.10	Maxilla, Right, M1	Triple burrial with No.11 & 12	Metal	N.D.
11	No.11	Maxilla, Right, M3	2352–2206 cal BP	Metal	N.D.
12		Left temporal bone	Triple burrial with No. 10 &12		N.D.
13	No.12	Mandible. Left, M3	Triple burrial with No.10 & 11	Metal	N.D.
14		Left temporal bone			N9a
15	No.14	Mandible, Left, C	Same layer as No. 3 & 4 (1879–1737, 1923–1823 cal BP)	Metal	M7
15	No.19	Maxilla, Right, M3	Same layer as No. 2 (2304–2046 BP)	Metal	N.D.
17	No.21	Mandible, Left, M2	Metal Period layer	Metal	N.D.
18	No.23	Mandible, Right, M2	Triple burial with No. 24 & 25, same layer as burial No.11 (2352–2206 cal BP)	Metal	N.D.
19		Left temporal bone			Y2
20	No.24	Mandible, Right, M2	Triple burial with No. 23 & 25	Metal	N.D.
21		Right temporal bone			Y2
22	No.25	Right temporal bone	Triple burial with No. 23 & 24	Metal	N
23	No.26	Left temporal bone	3136–2953 cal BP	Neolithic	R
24	No.27	Mandible. Right, M3	1864–1719 cal BP	Metal	N.D.
25		Right temporal bone			M
26	No.36	Maxilla, Right, M3	Same layer as No.53 (2711–2379 cal BP)	Neolithic	N.D.
27		Left temporal bone			N.D.
28	No.38	Right temporal bone	Same layer as No.53 (2711–2379 cal BP)	Neolithic	R
29	No.39	Maxilla, Right, M3			N.D.
30	No.42	Right temporal bone	Same layer as No.43 (2352–2206 cal BP)	Metal	E
31	No.49	Right temporal bone	Same layer as No.58 (2003–1899 cal BP)	Metal	N.D.
32	No.57	Right temporal bone	Same layer as No.56 (1896–1820 cal BP)	Metal	R
33	No.60	Mandible, Right, M2	Earlier layer than No. 56 (1896–1820 cal BP)	Metal	B4c
34	No.74	Maxilla, Left, M3	4572–4514 cal BP	Pre Neolithic	N.D.
35	No.79	Maxilla, Right, M2	Layer between No. 74 & 80 (4434–5712 cal BP)	Pre-Neolithic	N.D.
36		Right temporal bone			N.D.

N.D. denotes Not Determined

aDNA analyses previously and has yielded successful results [71, 76]. In this study, 81 haplogroup-diagnostic SNPs and three deletion/insertion polymorphisms, and a 9 base-pair repeat variation in the non-coding cytochrome oxidase II/tRNALys intergenic region were analyzed using the multiplex APLP method and the primer sets described by Kakuda et al. [77]. Polymorphic sites included in this analysis are known to cover most haplogroup-defining mutations found in East and Southeast Asian mtDNAs. The constitution of the PCR reaction mixture, thermal conditions, and method for separating and detecting PCR products were undertaken following Kakuda et al. [73].

In addition to APLP analysis, next-generation sequencing (NGS) technology and the mtDNA capture method were applied to individuals, whose haplogroups were tentatively determined or ambiguously identified by APLP analysis, to determine the mtDNA haplogroup or haplotype more precisely. Libraries were prepared using 8 μl of DNA extracts and using the established protocols following Shinoda or Meyer and Kircher [78, 79]. For some of the badly degraded ancient DNA, Multiplex PCR kit (QIAGEN) was used instead of AccuPrime *Pfx* kit (Life Technologies) in the first round of PCR amplification. The TreSeq DNA LT Set A or HT (Illumina) barcode for was used for indexing. Bait preparation and mtDNA enrichment for libraries were conducted following the protocol of Maricic et al. [80] and sequenced on an Illumina MiSeq platform (MiSeq Reagent Kit 150 or 300 Cycles) with 75 or 150 cycles paired-end run.

Raw sequence reads were processed following a modified protocol of Shinoda et al. [79]. After adapter trimming and merging of paired reads with AdapterRemoval v2 (—trimns—trimqualities—minquality 25—minlength 35—collapse), the merged reads were mapped to a human reference genome (hg19) using the Burrows-Wheeler Aligner (BWA) (version 0.7.8) aln option (-l 1000) [81, 82]. Cross-contaminants among samples sequenced on the same sequence run were removed using the process outlined by Kanzawa-Kiriyama et al. [83]. The reads mapped to NUMT or mitochondrial genome were retrieved and remapped to the human mitochondrial genome (revised Cambridge Reference Sequence: rCRS) with the same criteria applied when mapping to hg19. PCR duplicates were removed using Picard MarkDuplicates (version 1.119) (http://broadinstitute.github.io/picard/), and only reads with mapping quality ≥20 were retained [84]. A mpileup file (-Q 30) was compiled using SAMtools (version 1.0), calculating coverage of width and average depth. The resulting bam file was also applied to Genome Analysis Toolkit (GATK) HaplotypeCaller [85] (-stand_emit_conf 10) to call SNPs and indels [81]. The sites with low depth (<3) and high mismatch to consensus sequences (>30 percent) were masked, allowing the manual determination of the mtDNA haplogroup based on PhyloTree-Build 17 [86]. Some SNPs that were characteristic of the haplogroup but masked because of low depth were manually re-identified. We also determined the mtDNA haplogroup by using HaploGrep2 as double check of the haplotyping [87].

We investigated the degree of terminal C to T misincorporation using PMDtools and read length distribution, both of which are characteristic of ancient DNA [88]. In order to estimate contamination frequency, we used Schmutzi software (contDeam.pl—lengthDeam 40—library double) [89].

Results

Cranial morphometric analysis

Basic statistics for the early and late Gua Harimau male series are presented in the S1 Appendix. Fig 4 presents the results of the Net Split analysis, applied to the distances of the Q-mode correlation coefficients based on 16 cranial measurements. Essentially, this unrooted network tree exhibited a straightforward dichotomization of the comparative group into two major clusters: (1) Northeast and East Asians, and several sets of Southeast Asians, ranging from the Neolithic to contemporary periods, occupy the upper left of the tree. The contemporary Southeast Asians are scattered adjacent to this cluster. (2) The Australo–Papuans, Veda of Sri Lanka, Nicobarese, Andaman Islanders, and early Holocene Southeast Asians, including Hoabinhian samples (who are morphologically quite distinct from Northeast and East Asians) form another major separate tree cluster on the lower right of the tree. It is quite interesting that Early Gua Harimau, a subset of the pre-Neolithic samples, is closely connected with the late Pleistocene and early Holocene populations, including Hoabinhian, early Bac Son, and Con
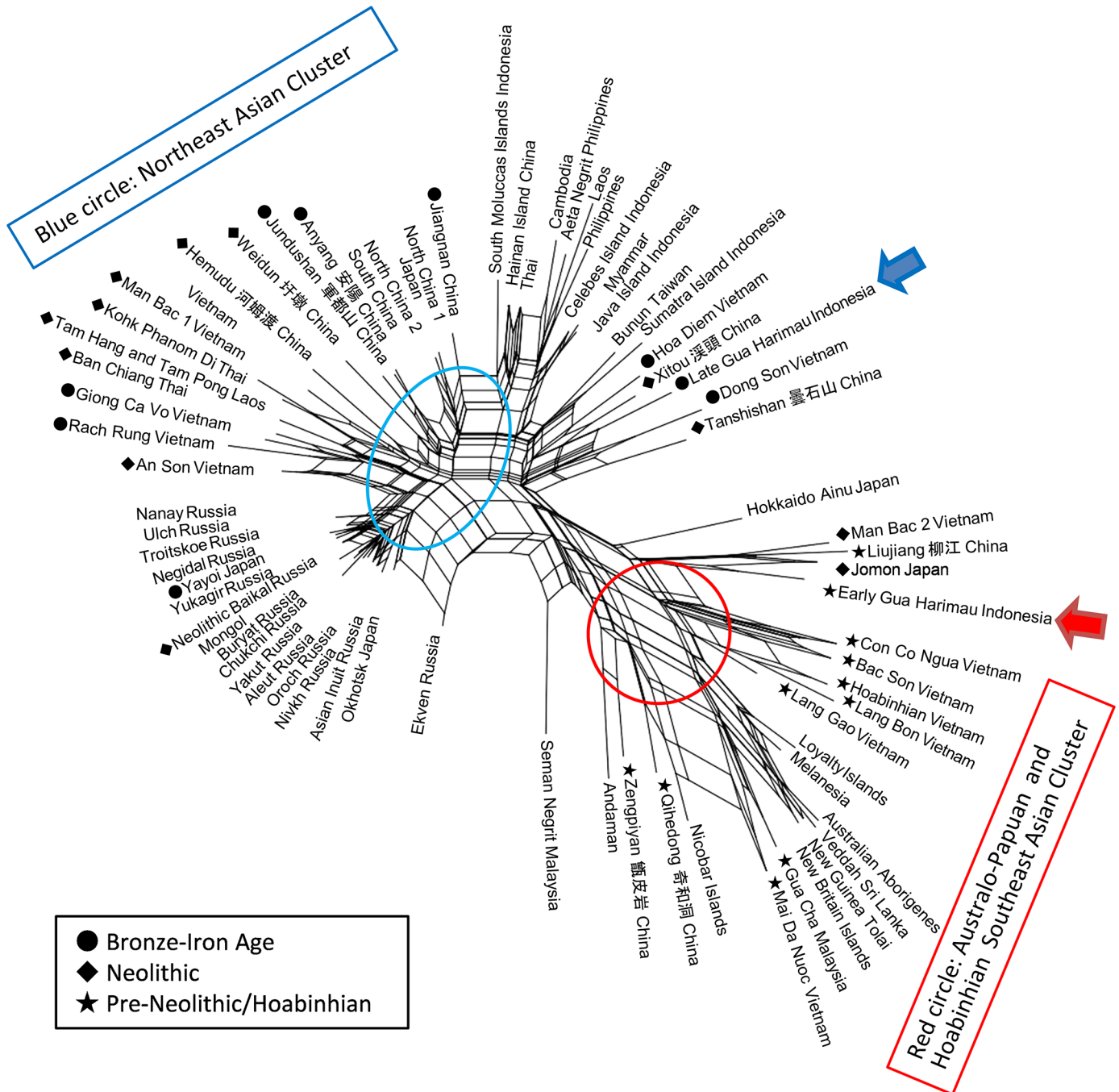
Fig 4. A neighbour net splits tree generated from a Q-mode correlation coefficients matrix, based on the craniometric data, comparing the archaeological and modern sample populations. The Late (pre-Neolithic) and Early Gua Harimau samples are boxed for ease of identification.

https://doi.org/10.1371/journal.pone.0198689.g004

Co Ngua. These samples form a mega cluster together with the Australo-Papuan and Gua Cha Malay series (Hoabinhian).

In the context of the ADH, it is notable that the Late Gua Harimau series (a Metal Period subset) exhibits close affinities with the Taiwan Formosa (Bunun), Sumatra, Moluccas, Philippines, and Celebes Island series. These current Austronesian speakers also closely cluster with

Northeast Asians. Moreover, the Xitou series rom Fujian Province, one of the representative groups of the Neolithic Southern Chinese, are tightly clustered with current Austronesian speakers, as well as the Late Gua Harimau series.

The Philippine Negritos, despite possessing phenotypically different features to surrounding populations, do not show any remarkable dissimilarity to the non-Negrito Philippine samples in terms of their craniometric morphology/affinities. The Iron Age Hoa Diem sample from central Vietnam shows a close affinity to the contemporary Late Gua Harimau series, while other Vietnamese samples, including Neolithic Man Bac, An Son, Metal Age Dong Son and Giong Ca Vo more closely resemble MSEA rather than ISEA Austronesian groups.

Mitochondrial DNA analysis

Table 4 presents the results of APLP analysis employed in the identification of mtDNA haplogroups. Of the 36 individuals considered in this analysis, we could successfully assign 13 mtDNAs of the individuals to the smallest named haplogroups and macro-haplogroups that can be identified by the APLP system employed in the present study.

In order to determine mtDNA haplogroups more precisely, from the highly fragmented DNA, we used NGS and the mtDNA capture method to investigate these tentative assigned sequences and one N.D. (not determined) sample (individual No. 79), which was the oldest of all the samples. Results of the NGS analysis are presented in Table 5. There were sufficient mtDNA reads to determine the mtDNA haplogroup or haplotype for 11 individuals. The lengths of the mtDNA fragments were very short, which is characteristic of aDNA (S2 Fig). Terminal C to T and G to A misincorporations were observed in all 11 individuals (S1 Fig). Contamination was less than two percent, so we expect that many fragments were endogenous human mtDNA. Thus, it is clear that the extracted solutions contained authentic human DNA.

Complete or nearly complete mitochondrial genome sequences were determined from these 11 individuals at 7.1~109.1-fold coverage. Each haplogroup and additional SNPs and indels from the hg node are also presented in Table 5. Although individual No. 14 was classified into M7b1a, the DNA fragments were relatively long and had little misincorporation (S1, S2 and S3 Figs). In addition, length of the DNA fragments having M7b1a specific mutations are relatively longer than other reads, especially C/T or G/A damaged reads at 3 bases sequence termini, which are considered authentic DNA [90]. Therefore, we considered it to be modern human DNA contamination. For No.12, while APLP analysis assigned its mtDNA to haplogroup N9a, NGS data was insufficient for evaluating the authenticity of this mtDNA. While APLP analysis can quickly and efficiently determine a haplogroup with low cost, the NGS analysis can verify the authenticity of a haplogroup or haplotype.

Discussion

Pre-Austronesian indigenous populations

Before assessing the ADH, it is necessary to review the evidence for the earliest human populations in the region. The date for anatomically modern human colonization of MSEA and ISEA is attested by way of assemblages excavated in Tam Pa Ling in Laos, Niah in Malaysia, and Tabon in the Philippines, ranging from 47,000 to 30,000 years BP [91–94]. Of these, the Niah and the Tabon series were excavated from sites now occupied by Austronesian speakers, and in the context of the ADH can be seen as representative of pre-dispersal indigenous populations. However, the poor preservation of such remains limit any attempts to assess their relationship to each other or later series in the region.

Table 5. Results of the NGS analysis.

Individual number	Sampling	Illumina index	Library preparation method	Kit for 1st PCR	Read length of sequence	Total paired reads	Merged reads	Hits to hg 19	(%)	Remove cross contamination	NUMT +mtDNA	Remap to rCRS with flag 0 or 16	Removal of PCR duplicates	> = mapq20	Coverage	Average depth	Haplogroup by HaploGrep2 (quality)	Haplogroup by manual determination	Haplogroup by APLP	Contamination frequency	Additional SNPs and indels from the hg node
No. 1	Right temporal bone	505–710	[28]	Multiplex	150	872,632	580,769	262,417	45.2%	-	26,017	25,294	10,133	10,133	99.99%	47.2	E1a1a1a (98.16%)	**E1a1a1a**	E	0% [0–0.5%]	None
No. 4	Maxilla, Right, M1	A006	[79]	Multiplex	75	683,144	534,229	23,648	4.4%	21,201	27,995	27,933	2,729	2,729	96.99%	9.3	E1a1a (77.19%)	**E1a1a1**	N.D.	0% [0–1%]	C10132T (a)
		A001	[79]	AccuPrime Pfx	75	223,019	117,125	3,774	3.2%	2,786											
		A012	[79]	Multiplex	75	215,102	168,576	6,143	3.6%	5,596											
		A014	[79]	Multiplex	75	136,703	46,658	2,140	4.6%	1,868											
No. 4	Right temporal bone	506–711	[28]	Multiplex	150	172,000	134,839	3,400	2.5%	-	329	320	26	26	-	-	-	**N.D.**	N.D.	-	-
No. 9	Left temporal bone	501–710	[28]	Multiplex	150	330,343	226,354	92,824	41.0%	-	18,789	18,572	3,332	3,332	99.94%	12.9	R+16189 (88.15%)	**B4a1a***	B4a	6.5% [4.5–8.5%]	A200G, GCA513G, T11830C, CAA16179C
No. 11	Left temporal bone	502–711	[28]	Multiplex	150	44,981	32,895	2,552	7.8%	-	576	576	60	60	-	-	-	**N.D.**	N.D.	-	-
No. 12	Left temporal bone	507–712	[28]	Multiplex	150	119,585	90,453	1,651	1.8%	-	276	273	67	67	-	-	-	**N.D.**	N9a	-	-
No. 14	Mandible, Left, C	A012	[79]	Multiplex	150	50,935	40,333	1,236	3.1%	1,071	1,104	1,097	288	288	71.19%	1.4	-	**(M7b1a)**	M7	1.5% [0.5–2.5%]	None
		A015	[79]	Multiplex	150	71,414	29,953	747	2.5%	544											
No. 23	Left temporal bone	504–711	[28]	Multiplex	150	322,776	222,042	75,379	33.9%	-	15,460	15,238	6,688	6,688	99.98%	30.7	Y2a1 (92.64%)	**Y2a1***	Y2	1.5% [0.5–2.5%]	None
No. 24	Right temporal bone	503–712	[28]	Multiplex	150	327,475	266,467	115,298	43.3%	-	26,055	25,809	6,775	6,775	99.98%	27.5	Y2a (97.25%)	**Y2a1***	Y2	1.5% [0.5–2.5%]	None
No. 25	Right temporal bone	508–701	[28]	Multiplex	150	247,176	123,683	34,674	28.0%	-	6,383	6,330	1,785	1,785	99.40%	7.1	F1 (88.62%)	**F1a1a1**	N	0% [0–0.5%]	GCA513G
No. 26	Left temporal bone	A007	[79]	AccuPrime Pfx	75	2,662,710	1,825,813	1,066,688	58.4%	1,012,747	828,262	817,900	21,377	21,376	100.00%	109.1	R (77.01%)	**R***	R	2% [1–3%]	C150T, A189G, T310TC, T450C, A339?G, G3483A, C3600A, A5484G, C6164T, A7271G, T9833C, G9966A, C10777T, G11150A, T14178C, T14311C, A15766G, T16304C
		A012	[79]	AccuPrime Pfx	75	2,499,051	1,745,258	904,864	51.8%	845,902											
No. 27	Right temporal bone	501–704	[28]	Multiplex	150	1,762,760	1,613,520	856,002	53.1%	-	104,366	102,986	10,381	10,381	99.98%	40.3	H2a2a1 (50.00%)	**M***	M	0% [0–0.5%]	C151T, T152C, T310TC, GCA513G, C3817T, T4336C, T4823C, G8592A, A9285G, G11176A, C12378T, G15172A, T16229C, C16294T, T16311C
No. 36	Left temporal bone	502–705	[28]	Multiplex	150	164,461	104,406	8,240	7.9%	-	2,024	2,024	152	152	-	-	-	**N.D.**	N.D.	-	-

(Continued)

Table 5. (Continued)

Individual number	Sampling	Illumina index	Library preparation method	Kit for 1st PCR	Read length of sequence	Total paired reads	Merged reads	Hits to hg 19	(%)	Remove cross contamination	NUMT +mtDNA	Remap to rCRS with flag 0 or 16	Removal of PCR duplicates	>= mapq20	Coverage	Average depth	Haplogroup by HaploGrep2 (quality)	Haplogroup by manual determination	Haplogroup by APLP	Contamination frequency	Additional SNPs and indels from the hg node
No. 38	Right temporal bone	503–706	[78]	Multiplex	150	1,279,220	1,095,331	504,191	46.0%	-	31,988	31,027	4,784	4,784	99.96%	19.3	B4c1b2a2 (81.63%)	**B4c1b2a2***	R	0% [0–0.5%]	None
No. 42	Right temporal bone	505–702	[78]	Multiplex	150	1,111,502	803,261	420,543	52.4%	-	36,986	35,534	16,614	16,614	100.00%	88.8	E1a1a1a (98.22%)	**E1a1a1a**	E	0% [0–0.5%]	A302AC, T310TC
No. 49	Right temporal bone	504–707	[78]	Multiplex	150	174,450	109,241	5,247	4.8%	-	374	364	40	40	-	-	-	N.D.	N.D.	-	-
No. 57	Right temporal bone	505–708	[78]	Multiplex	150	360,438	306,470	101,159	33.0%	-	5,942	5,779	850	850	-	-	-	N.D.	R	-	-
No. 60	Mandible, Right, M2	A018	[79]	Multiplex	150	87,144	39,084	18,568	47.5%	16,288	118,606	118,533	8,273	8,273	99.95%	35.7	B4c1b2a2 (85.68%)	**B4c1b2a2**	B4c	0% [0–0.5%]	C146T, A302ACC, A16183C
		A002	[79]	AccuPrime Pfx	75	792,360	446,015	101,601	22.8%	89,515											
		A005	[78]	AccuPrime Pfx	150	34,517	24,477	7,639	31.2%	7,519											
		A016	[78]	AccuPrime Pfx	150	31,402	16,391	3,843	23.4%	3,765											
No. 79	Maxilla, Right, M2	A019	[79]	Multiplex	150	19,867	14,176	232	1.6%	144	-	-	-	-	-	-	-	N.D.		N.D.	-	-
No. 79	Right temporal bone	506–709	[78]	Multiplex	150	186,433	143,691	29,993	20.9%	-	1,760	1,730	420	420	-	-	-	N.D.	N.D.	-	-

N.D. denotes Not Determined

(a) Ambiguous SNP because of low depth (4) and high mismatch rate (25%)

* The haplogroup of the individual in question is ancestral, and as such was not classified into a sub-haplogroup

It is not until the late Pleistocene to early/mid-Holocene (often referred to the Hoabinhian in MSEA), c. 23,000–8000 BP [95–98], that we have a robust sample of ostensibly pre-ADH crania. Key specimens derive from cave sites in Vietnam and Malaysia (for instance, Lang Gao, Lang Bon, Pho Binh Gia, Lang Cuom, Cua Gi, Mai Da Nuoc and Mai Da Dieu in Vietnam and Gua Cha on the central Malay Peninsula [10, 21–24, 26, 30, 96]). As shown in Fig 4, all the available Hoabinhian specimens are consistently defined as having close Australo-Papuan affinities in terms of their cranio-metrically expressed morphology. While the focus of this analysis is on male crania (see Methods), female material (e.g., Hang Cho, Gua Gunung Runtuh, and Moh Khiew) have also demonstrated remarkable cranial and dental similarities to Australian and/or Melanesian samples, suggesting a close biological affinity [99–101]. The network tree diagram (Fig 4) further indicates that some Pleistocene and early Holocene samples from China (Liujiang and Zenpiyang from Guangxi) share morphological similarities with MSEA Hoabinhian samples. Furthermore, the cranial traits characterizing these early indigenous inhabitants in the region (for instance, in northern Vietnam), were retained through the subsequent pre-Neolithic Da But Culture (c. 6700–4500 BP), clearly suggesting that such pre-agricultural foraging communities were likely direct lineal descendants of Hoabinhian foragers.

The earliest reliably dated anatomically modern humans in the region have been found in Southeast Asia, suggesting the initial colonization of the region via India, rather than north and inland through Siberia (see discussion in Buckley and Oxenham [102]). Moreover, these first colonists shared a common ancestry with the earliest settlers of continental Sahul (Australian and New Guinea). Indeed, there is a long history of scholarship [9, 10] suggesting morphological similarities, with implied genetic affinities, between Australian Aboriginals, Papuans, Melanesians and (poorly preserved) pre-Neolithic populations in Southeast Asia (e.g., Tabon in Philippines and Niah, Gua Cha, Guar Kepha, and Gua Kerbau in Malaysia). The current analysis of a more extensive cranial dataset finds further support for close affinities between early Southeast Asians, including Hoabinhian samples and Australian and Papuan-Melanesian groups, as well as the Andaman and Nicobar Indians. These observed close biological ties linking Sahul, early mainland Southeast Asia, and Eastern India, strongly suggest that the first anatomically modern human colonizers of this region migrated to the southern rim of Eurasia and then dispersed into late Pleistocene Sundaland (Southeastern Asia), including what is now ISEA. Pre-Austronesian indigenous populations may, in turn, share a common ancestry with early Hoabinhian populations in MSEA and present-day Australian Aboriginal, Papuan, and Melanesian peoples. In fact, as depicted in Fig 4, the pre-Neolithic samples from Gua Harimau (Early Gua Harimau) show a close affinity with these early settlers of MSEA and Sahul, or the first anatomically modern humans in the region.

Austronesian dispersal

The cranio-metric analysis (see Fig 4) demonstrates a close association between the Late Gua Harimau (Metal period) and contemporary Taiwan (Bunun), Sumatra, Moluccas, Philippines, and Celebes Island samples. The morphological affinities between these series suggests a significant level of genetic interaction among neighboring inhabitants of ISEA in the past. The clustering of the Hoa Diem sample with the aforementioned series is worth discussing in more detail.

The large mortuary site at Hoa Diem, located in Khanh Hoa Province in central Vietnam, is interesting in terms of its assumed ancestry to the Chamic people of the same region. The excavation of this site has produced a large number of jar burials and associated mortuary ceramics that are strikingly similar to those from Kalanay Cave in the Philippines [52].

Similarities in material culture between the Philippines and central coastal Vietnam, as well as cranial morphometric clustering of Indonesian (Late Gua Harimau) and coastal Vietnam (Hoa Diem) populations collectively suggest substantive connections and interactions among Island populations bordering the South China Sea during the Iron Age.

Regionally, the prehistoric dispersal of Austroasiatic speakers across MSEA and Austronesian speakers throughout ISEA and the Pacific has been explicitly associated with the spread of farming during the Neolithic and subsequent early Bronze and Iron ages [1, 3, 5, 103–117]. Linguistic data indicate that Southern China and Taiwan were the origin of many of the existing language families of Southeast Asia, while archaeological evidence places the origins of Neolithic farming societies in the Yangzi and Yellow River Basins during the early Holocene, prior to their subsequent expansion into Southern China, Southeast and eastern Asia [5, 118–122].

These major Neolithic demographic transitions (NDT) in the region are often referred to in terms of the two-layer-model, whereby the first layer refers to late Pleistocene occupation of East and Southeast Asia, with the second layer being characterized by the NDT and the arrival of the ancestors of contemporary Austroasiatic (MSEA) and Austronesian (ISEA and the Pacific) speakers. Modern day Australians, Papuans and Melanesians represent the direct descendants of the first layer populations, while the descendants of the second layer include the somewhat heterogenous populations characterizing the Neolithic through to modern times.

The results from the cranial morphometric analysis in this study clearly supports the two-layer-model for both MSEA and ISEA by demonstrating close morphological associations between widely dispersed pre-NDT samples (or first layer populations) in the broader region. For instance, the early Holocene Qihedong series from Fujian Province, China, and Early Gua Harimau sample from Sumatra, Indonesia, cluster together within the Australo-Papuan group (see Fig 4). On the other hand, evidence for the spread of second layer (or NDT populations) can be identified by way of the close affinities between the Late Gua Harimau series, a number of Austronesian speaking assemblages from ISEA, and the Neolithic Southern Chinese sample from Xitou.

Turning to the genetic evidence, in disagreement with the Austronesian Dispersal Hypothesis (ADH), or Out-of-Taiwan model, is Cox and colleagues work [123, 124] which argued for a significant genetic cline across ISEA and the Pacific, ostensibly traced back to incoming populations from MSEA. Cox et al. [124] concluded that the phenotypic gradient likely reflects a mixing of two major ancestral source populations; one descended from the initial occupants of the region who were related to modern Melanesians, and the other related to Asian immigrants since the Neolithic period. Other research has also rejected the idea of large-scale demographic movement during the Neolithic, advocating for local evolutionary processes in the context of evidence for a common genetic heritage derived from the late Pleistocene colonization of Sundaland [125, 126]. As for the Austronesian expansion into mainland Southeast Asia, mtDNA analysis of Austronesian-speaking Cham individuals in central Vietnam suggests that cultural, rather than genetic, links were more a factor in this case [127]. Other DNA studies have argued that Southeast Asia was a major geographic source of East Asian populations, within which the roots of all present-day East Eurasians are historically united via a single primary wave of migration into the region [128, 129].

In this study we were able to determine mtDNA haplogroups for 11/36 samples. Three individuals (Nos. 9, 38 and 60) were identified as a subgroup of haplogroup B, a common haplogroup in ISEA that is comprised of two main clades: B4 and B5. Most of these B lineages in ISEA fall within haplogroup B4, while B5 is relatively rare. The bulk of B4 in ISEA is B4a with its major branch, B4a1, including the so-called 'Polynesian motif'. Although Hill et al.'s [125]

mtDNA analysis indicated that the dispersal of haplogroup B4a1 was triggered by postglacial flooding in the late Pleistocene or early Holocene, B4a1a has a similar distribution to that of Austronesian speakers. Gua Harimau individual No. 9 was assigned to haplogroup B4a1a, suggesting their ancestry may be Austronesian.

Arguably, lineages of haplogroup B are largely the result of a second wave of dispersal of proto-Austronesian speakers. The ancestral forms of haplogroups B4b, B4c, and B5b are found in South Chinese populations, a mainland origin, and subsequent dispersal into ISEA. The B4c haplogroup has been found in samples of ancient Negrito hair, potentially indicating a diffusion of this haplogroup from the mainland [130].

Two Gua Harimau individuals (No. 38 and 60) were classified into sub-haplogroups of B4 using whole-mtDNA sequence analysis: B4c1b2a2. Haplogroup B4c was found to have an age between 32,000 BP and 25,000 BP, with sub-haplogroup B4c1 originating between 27,000 BP and 24,000 BP, B4c1b2 to between 16,000 BP and 14,000 BP, with the origin of B4c1b2a2 dating to the Neolithic [131]. According to the DNA Database in Japan, haplogroup B4c1b2a is found in South China (Liaoning and Zhejiang provinces), as well as among aboriginal Taiwanese, the Philippines, and Indonesia. Given this demographic distribution, sub-haplogroup B4c1b2a appears to be the group associated with the Austronesian expansion during the Neolithic and/or post-Neolithic periods.

Haplogroup E is common in ISEA [125], and is frequently carried by aboriginal Taiwanese, however, it is otherwise almost absent in China and the Pacific. It has been proposed as a potential hypothetical genetic marker of Austronesian-speaking people [126]. Notwithstanding, others have attributed the origin of this haplogroup to an early Holocene population expansion originating within ISEA, which is inconsistent with the Neolithic agriculturally-driven population dispersal hypothesized in the ADH model [125, 126]. In fact, there are two major subclades, E1 and E2. Of these, E1 comprises two additional subclades, E1a and E1b, the former almost entirely restricted to Taiwan and ISEA, while the latter is found predominantly in the ISEA but absent in Taiwan.

Previously, haplogroup E itself dates to over 25,000 BP and lineages within haplogroup E have dates ranging from 6,000 BP to 16,000 BP, while a recent study based on ancient DNA calibration and Bayesian dating suggests that haplogroup E probably arose 8,136–10,933 ya (95% highest posterior density, HPD) and the majority of E lineages show a coalescence at 5–8 kya with a higher mean probability at about 6 kya [132]. According to this new time frame, Ko et al. (2014) reconstructed a history of early Austronesians arriving in Taiwan in the north ~6,000 ya, spreading rapidly to the south, and leaving Taiwan ~4,000 ya to spread throughout ISEA, Madagascar, and Oceania. Based on the demographic distribution and new time depth, E1a1a is a candidate for the presumed out-of-Taiwan dispersal. Spatial frequency distribution and diversity suggest that this haplogroup arose within ISEA, while some of its subclades subsequently spread to Taiwan [126]. This haplogroup probably evolved within the descendants of the Austronesian-speaking groups originating from Taiwan.

Relatively poor mtDNA preservation of Gua Harimau individual No. 4 (Metal period, c. 2,196–1,786 BP) makes identification of its sub-haplogroup difficult. Notwithstanding, individual No. 4's sequences were tentatively classified as E1a1a based on diagnostic coding site changes. The greater diversity of haplogroup E in ISEA compared to Taiwan is consistent with the expansion of populations from the south [125, 126]. However, E1a1a has a lower diversity in Philippine and Sulawesi populations than it does among Taiwanese aboriginals, despite making up a larger proportion of these populations [126]. While haplogroup E may be a marker of postglacial expansion, clades within this haplogroup, such as E1a1a, possibly reflect the impact of later population events [133]. The most plausible explanation for this observation

is that the diffusion of the haplogroup E1a1a in the Gua Harimau population occurred after the Neolithic expansion.

Haplogroup Y2a1 was observed in Gua Harimau individuals No. 23 and 24. This haplogroup is also common in ISEA and shared by Philippine, Taiwanese aboriginal, and other ISEA populations [133]. Y2 has a slightly higher frequency in the Philippines compared with surrounding groups. The existence of this haplogroup suggests a genetic link between ISEA and Gua Harimau populations.

The two somewhat common and widespread Southeast Asian mtDNA haplogroups are B and R9, the latter encompassing haplogroup F, with F1a being widespread in Southeast Asia. The "Early Train" hypothesis [134] claimed large scale late Pleistocene/early Holocene dispersals from MSEA into Sunda, and helps explain the distribution of F1a. Gua Harimau individual No. 25 belongs to haplogroup F1a1a1, which has been observed in high frequencies in MSEA, suggesting a link between Gua Harimau and the mainland. F1a1a1 of individual No. 25 was probably introduced by the early train, although it is still consistent with the possibility that the haplogroup entered into the Gua Harimau population by way of the Neolithic Austronesian expansion, since it is unknown whether No. 25 pre-dates the expansion. Unclassified haplogroups R* (Gua Harimau individual No. 26) and M* (Gua Harimau individual No. 27) appear unrelated to any other global lineages, are the new basal R and M haplogroups, and represent indigenous haplogroups in ISEA. Table 5 presents the complete genome substitutions of these cases. Individual No. 26 has the diagnostic polymorphisms of macrohaplogroup N (rCRS positions at 8701, 9540, 10398, 10873, and 15301), macrohaplogroup R (rCRS positions at 12705 and 16223), and 18 specific nucleotide substitutions. Individual No. 27 has the diagnostic polymorphisms of macrohaplogroup M (rCRS positions at 489, 10400, 14783 and 15043) and 15 specific nucleotides substitutions.

There are several rare ancient haplogroups within macrohaplogroup N and its sub-haplogroups R and M in ISEA. The C14 AMS dating of Gua Harimau individual No. 26 (group R) places it at c. 3000 BP, or prior to major settlement by Metal period migrants, while individual No. 27 (group M) dates to the metal period at c. 2000 BP. It seems likely that these haplogroups are relics of the original Pleistocene inhabitants of ISEA. This view is based on evidence from the persistence of mtDNA ostensibly characterizing the earliest settlers of the region. Indeed, as discussed above in the context of cranial morphometric analysis, the pre-Neolithic indigenous Gua Harimau population can potentially trace their maternal ancestry back to the first anatomically modern settlers of ISEA. It is noteworthy that the Gua Harimau gene pool consists of Austronesian (B4a1, B4c, E1a and Y2), mainland (F1a), and putative indigenous (R* and M*) forms. The mtDNA analysis is limited in estimating the composition of the three lineages making up the Gua Harimau population as well as the manner in which they genetically changed over time at the site. Nuclear genome analyses are required in order to gain further detail.

Conclusions

The archaeological human remains from Gua Harimau cave, Sumatra, Indonesia provide evidence for at least two (cranially defined) and perhaps three (in the context of the ancient mtDNA results) distinct populations from two separate time periods, thus supporting the ADH or two-layer-model. The cranial data indicate that the pre-Neolithic occupants of (Early) Gua Harimau, who cluster with the Australo-Papuan series, were subsequently replaced by a population with close cranial affinities to present-day Austronesian speakers, including Taiwanese aboriginals, who possess Northeast Asian features to a certain extent. Further, it is apparent that the Neolithic Southern Chinese, represented by Xitou in Fujian Province, share close cranial affinities with both Austronesian speaking samples and the (Late) Gua Harimau

series, supporting the view that their remote homeland was somewhere in Southern China. The results from the mtDNA is not consistent with the view (based on DNA studies of modern populations, [123, 125]) of a single origin, stretching back into the Pleistocene, for the Gua Harimau population. While the two-later-model is well supported for MSEA [9, 10, 16], this study now provides substantive support for the value of the two-layer-model in also explaining the population history of ISEA.

Supporting information

S1 Appendix. Cranial and mandibular measurements (mm) for the Gua Harimau series. (DOCX)

S1 Fig. Pattern of postmortem misincorporation. C to T indicates C in reference genome and T in Gua Harimau samples, and G to A indicates G in reference genome and A in Gua Harimau samples. For No. 26, reduction of the misincorporation in 5' end compared to 3' end is explained by the inability of AccuPrime Pfx to bypass uracils, which is frequent in sequence termini.
(JPG)

S2 Fig. Fragment size distribution of sequence reads mapped to rCRS. Only sequences having mapping quality equal or larger than 20 were used. PCR duplicates were removed.
(JPG)

S3 Fig. Fragment size distribution of GH14. GH14 includes all mapped reads, and GH14 damaged includes the reads having C/T or G/A changes at 3 bases of sequence termini. White circle indicates the reads having mutations relating to haplogroup M7b1a. Those reads are relatively longer than other reads, and we considered that these are contaminants.
(JPG)

Acknowledgments

Author Contributions

Conceptualization: Hirofumi Matsumura.

Formal analysis: Hirofumi Matsumura, Ken-ichi Shinoda, Sofwan Noerwidi.

Funding acquisition: Hirofumi Matsumura, Truman Shimanjuntak, Marc F. Oxenham.

Investigation: Hirofumi Matsumura, Truman Shimanjuntak, Adhi Agus Oktaviana, Sofwan Noerwidi, Harry Octavianus Sofian, Dyah Prastiningtyas, Lan Cuong Nguyen, Tsuneo Kakuda, Hideaki Kanzawa-Kiriyama, Noboru Adachi, Hsiao-chun Hung, Xuechun Fan, Xiujie Wu, Anna Willis, Marc F. Oxenham.

Methodology: Hirofumi Matsumura, Ken-ichi Shinoda, Marc F. Oxenham.

Project administration: Hirofumi Matsumura.

Resources: Hsiao-chun Hung.

Writing – original draft: Hirofumi Matsumura, Marc F. Oxenham.

Writing – review & editing: Anna Willis, Marc F. Oxenham.

References

1. Bellwood P. Prehistory of the Indo-Malaysian Archipelago. 3rd ed. Honolulu, Hawaii: University of Hawai'i Press; 2007.

2. Blust R. Austronesian culture history: The window of language. TAPS. 1996; 86(5):28–35. https://doi.org/10.2307/1006619

3. Blust R. Beyond the Austronesian homeland: The Austric hypothesis and its implications for archaeology. TAPS. 1996; 86(5):117–58. https://doi.org/10.2307/1006623

4. Bellwood P. The origins and dispersals of agricultural communities in Southeast Asia. In: Glover I, Bellwood PS, editors. Southeast Asia: From prehistory to history. New York: RoutledgeCurzon; 2004. p. 21–40.

5. Bellwood P. Examining the farming/language dispersal hypothesis in the East Asian context. In: Blench R, Sagart L, Sanchez-Mazas A, editors. The peopling of East Asia: Putting together archaeology, linguistics and genetics. London: RoutledgeCurzon; 2005. p. 17–30.

6. Bellwood P, Dizon E. The Batanes archaeological project and the "Out of Taiwan" hypothesis for Austronesian dispersal. Journal of Austronesian Studies. 2005; 1(1):1–33.

7. Bellwood P, Dizon E. Austronesian cultural origins: Out of Taiwan, via the Batanes Islands, and onwards to Western Polynesia. In: Sanchez-Mazas A, Blench R, Ross M, Peiros I, Lin M, editors. Past human migrations in East Asia. London: Routledge; 2008. p. 23–39.

8. Bellwood P, Hung H-C. The dispersals of early food producers from Southern China into Southeast Asia. Hemudu Culture International Forum, China. Beijing: China Modern Economic Publishing House; 2013. p. 160–75.

9. Matsumura H, Oxenham MF. Demographic transitions and migration in prehistoric East/Southeast Asia through the lens of nonmetric dental traits. Am J Phys Anthropol. 2014; 155:45–65. https://doi.org/10.1002/ajpa.22537 PMID: 24954129

10. Matsumura H, Oxenham MF, Nguyen LC. Hoabinhians: A key population with which to debate the peopling of Southeast Asia. In: Kaifu Y, Izuho M, Goebel T, Sato H, Ono A, editors. Emergence and Diversity of Modern Human Behavior in Paleolithic Asia. Texas: Texas A&M University Press; 2015. p. 117–32.

11. Harrison B. A classification of stone age burials from Niah Great Cave, Sarawak. Sarawak Mus J. 1967; 15(30–31):126–200.

12. Harrisson T. Early dates for "seated" burial and burial matting at Niah Caves, Sarawak (Borneo). Asian Persp. 1975; 18(2):161.

13. Simanjuntak HT. Gua Harimau cave and the long journey of Oku civilization. Yogyakarta: Gadjah Mada University Press; 2016.

14. Bronk Ramsey C. Bayesian analysis of radiocarbon dates. Radiocarbon. 2009; 51(1):337–60.

15. Sneath PH, Sokal RR. Numerical taxonomy. The principles and practice of numerical classification. San Francisco: WH Freeman; 1973.

16. Oxenham MF, Matsumura H, Nguyen KD. Man Bac: The excavation of a Neolithic site in northern Vietnam. The biology. Canberra: ANU E Press; 2011.

17. Woo JK. Human fossils found in Liukiang, Kwapai, China. Vertebr Palasiat. 1959; 3:109–18.

18. Fang Y, Fan X, Li S. Body size of Neolithic human remains from the Qihe Cave, Zhangping, Fujian. Acta Anthropol Sin. 2015; 34(2):202–15.

19. Wu X, Fan X, Li S, Gao X, Zhang Y, Fang Y, et al. The early Neolithic human skull from the Qihe cave, Zhangping, Fujian. Acta Anthropol Sin. 2015; 33:448–59.

20. Institute of Archaeology Chinese Academy of Social Science (IACAS), Archaeological Team of Guangxi Zhuang Municipality (ATGZM), Zengpiyan Museum (ZM), Archaeological Team of Guilin City (ATGC). Zengpiyan—a prehistoric site in Guilin, Beijing. Beijing: Cultural Relics Publishing House; 2003.

21. Colani M. L'age de la pierre dans la province de Hoa-Binh (Tonkin). Mémoires du Service Géologique de l'Indochine. 1927; 14(1).

22. Colani M. La grotte sépulcrale de Lang Gao. L'Anthropologie. 1927; 37:227–9.

23. Colani M. La civilisation Hoabinhienne extrême-orientale. Bulletin de la Société Préhistorique Française. 1939; 36(170–174).

24. Nguyễn LC. Two early Hoabinhian crania from Thanh Hoa province, Vietnam. Z Morphol Anthropol. 1986; 77(1):11–7. PMID: 3564631

25. Nguyễn LC. Palaeoanthropology in Vietnam. Khảo cổ học. 2007; 2:23–41.

26. Mansuy H, Colani M. Néolithique inférieur (Bacsonien) et Néolithique supérieur dans le Haut-Tonkin. Bulletin du Service Géologique de l'Indochine. 1925; 12:1–45.

27. Nguyễn LC. Ancient human bones in Da But culture—Thanh Hoa province. Khảo cổ học. 2003; 3:66–79.

28. Oxenham MF, Trinh HH, Willis A, Jones RK, Domett K, Castillo C, et al. Between foraging and farming: Strategic responses to the Holocene thermal maximum in Southeast Asia. Antiquity.Forthcoming.

29. Nguyễn KT. Ancient human skeletons at Con Co Ngua. Khảo cổ học. 1990; 3:37–48.

30. Sieveking GdG. Excavations at Gua Cha, Kelantan 1954. Part 1. Federation Museums Journal. 1954; 1–2:75–138.

31. Nakahashi T, Li M. Ancient people in the Jiangnan region, China. Fukuoka: Kyushu University Press; 2002.

32. Fujian Provincial Museum. Excavation report of Minhou Baisha Xitou site1980.

33. Fujian Provincial Museum. Second excavation report of Minhou Baisha Xitou site1984.

34. Fujian Provincial Museum, Tanishishan Site Museum. Report on excavation of the Tanshishan site in Minhou in 2004. Fujian Wenbo. 2004; 1:1–12.

35. Zhejiang Cultural Relics Archaeological Research Institute (ZCARI). Hemudu: Report on the excavation of the Neolithic settlement site. Beijing: Wenwu, 2003.

36. Bellwood P, Oxenham MF, Bui CH, Nguyen KD, Willis A, Sarjeant C, et al. An Son and the Neolithic of Southern Vietnam. Asian Persp. 2011; 50:144–74.

37. Nguyễn LC. About the ancient human bones at An Son (Long An) through the third excavation. Khảo cổ học. 2006; 6:39–51.

38. Nishimura M, Nguyễn KD. Excavation of An Son: a Neolithic mound site in the middle Vam Co Dong valley, southern Vietnam. Bulletin of the Indo-Pacific Prehistory Association. 2002; 22:101–9.

39. Gorman C, Charoenwongsa P. Ban Chiang: A mosaic of impressions from the first two years. Expedition. 1976; 18:14–26.

40. Pietrusewsky M, Douglas MT. Ban Chiang, a prehistoric village site in Northeast Thailand I: The skeletal remains. Philadelphia: University of Pennsylvania Museum of Archaeology and Anthropology; 2002.

41. Higham CFW, Thosarat R. The excavation of Khok Phantom Di, a prehistoric site in Central Thailand. Volume III: The material culture. London: Society of Antiquaries of London; 1993.

42. Tayles N. The excavation of Khok Phanom Di, a prehistoric site in Central Thailand. Volume V: The people. London: Society of Antiquaries of London; 1999.

43. Bayard DT. Non Nok Tha: The 1968 excavation, procedure, stratigraphy and summary of the evidence. Dunedin: University of Otago, Department of Anthropology; 1972.

44. Debets G. Anthropological studies in the Kamchatka region. Trudy Instituta of Ethnografii. 1951; 17:1–263.

45. Ishida H. Craniometric variation of the Northeast Asian populations. Homo. 1997; 48:106–24.

46. Habu J. Ancient Jomon of Japan. Cambridge: Cambridge University Press; 2004.

47. Hanihara T. Craniofacial features of Southeast Asians and Jomonese: A reconsideration of their microevolution since the Late Pleistocene. Anthropol Sci. 1993; 101(1):25–46. https://doi.org/10.1537/ase.101.25

48. Hanihara T. Frontal and facial flatness of major human populations. Am J Phys Anthropol. 2000; 111 (1):105–34. https://doi.org/10.1002/(SICI)1096-8644(200001)111:1<105::AID-AJPA7>3.0.CO;2-O PMID: 10618591

49. Institute of History and Institute of Archaeology (IHIA), Chinese Academy of Social Science (CASS). Contributions to the study on human skulls from the Shang sites at Anyang. Beijing: Cultural Relics Publishing House; 1982.

50. Han KX, Qi PF. The study of the human bones of the middle and small cemeteries of Yin sites, Anyang. In: Institute of History and Institute of Archaeology (IHIA), Chinese Academy of Social Science (CASS), editors. Contributions to the study on human skulls from the Shang sites at Anyang. Beijing: Cultural Relics Publishing House; 1982. p. 50–81.

51. Đặng VT, Vũ QH. Excavation at Giong Ca Vo site, Can Gio district, Ho Chi Minh city. Journal of Southeast Asian Archaeology. 1997; 17(6):30–44.

52. Yamagata M, Bùi CH, Nguyễn KD. The excavation of Hòa Diêm in Central Vietnam. Tokyo: Showa Women's University Institute of International Culture; 2012.

53. Nguyễn LC. Anthropological characteristics of Dông Sơn population in Vietnam. Hanoi: Social Sciences Publishing House; 1996.

54. Higham CFW, Kijngam A. The origins of the civilization of Angkor. Volume III: The Excavation of Ban Non Wat. Bangkok: The Thai Fine Arts Department; 2009.

55. Higham CFW, Kijngam A. The origins of the civilization of Angkor. Volume IV: The Excavation of Ban Non Wat. Part II: the Neolithic Occupation. Bangkok: Thai Fine Arts Department; 2011.

56. Higham CFW, Kijngam A. The origins of the civilization of Angkor. Volume V: The Excavation of Ban Non Wat. Part III: the Bronze Age. Bangkok: Thai Fine Arts Department; 2012.

57. Higham CFW, Kijngam A. The origins of the civilization of Angkor. Volume VI: The Excavation of Ban Non Wat. Part IV: the Iron Age, summary and conclusions. Bangkok: Thai Fine Arts Department; 2012.

58. Bùi PD, Đào LC, Vương TH. Archaeology in Long An province: The first C.E. centuries. Long An: Long An Provincial Museum, 2001.

59. Beijing Cultural Relic Insitute (BCRI). Jundushang burial grounds. Beijing: Wenwu, 2007.

60. Nakahashi T. The Yayoi people. In: Nagai M, Nasu T, Kanaseki Y, Sahara M, editors. Research on Yayoi culture. Tokyo: Yuzankaku; 1989. p. 23–51.

61. Howells WW. Skull shapes and the map: Cranio-metric analysis in the dispersion of modern *Homo*. Papers of the Peabody Museum of Archaeology and Ethnology. 1989; 79.

62. Pietrusewsky M, Chang C-f. Taiwan aboriginals and peoples of the Pacific-Asia region: Multivariate craniometric comparisons. Anthropol Sci. 2003; 111(3):293–332. https://doi.org/10.1537/ase.111.293

63. Pietrusewsky M. Cranial variation in early metal age Thailand and Southeast Asia studied by multivariate procedures. Homo. 1981; 32(1):1–26.

64. Suzuki H, Mizoguchi Y, Conese E. Craniofacial measurement of artificially deformed skulls from the Philippines. Anthropol Sci. 1993; 101(1):111–27. https://doi.org/10.1537/ase.101.111

65. Sangvichien S. Physical anthropology of the skull of Thai [PhD Thesis]. Bangkok1971.

66. Ishida H. Metric and nonmetric cranial variation of the prehistoric Okhotsk people. Anthropol Sci. 1996; 104(3):233–58. https://doi.org/10.1537/ase.104.233

67. Ishida H. Cranial morphology of several ethnic groups from the Amur basin and Sakhalin. J Anthrop Soc Nippon. 1990; 98(2):137–48.

68. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006; 23(2):254–67. https://doi.org/10.1093/molbev/msj030 PMID: 16221896

69. Woodward SR, King MJ, Chiu NM, Kuchar MJ, Griggs CW. Amplification of ancient nuclear DNA from teeth and soft tissues. Genome Res. 1994; 3(4):244–7.

70. Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-fortes G, Mattiangeli V, et al. Genome flux and stasis in a five millennium transect of European prehistory. Nat Commun. 2014; 5:5257. http://dx.doi.org/10.1038/ncomms6257. PMID: 25334030.

71. Shinoda K-i, Adachi N, Guillen S, Shimada I. Mitochondrial DNA analysis of ancient Peruvian highlanders. Am J Phys Anthropol. 2006; 131(1):98–107. https://doi.org/10.1002/ajpa.20408 PMID: 16485299

72. Gilbert MTP, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, Lynnerup N, et al. Distribution patterns of postmortem damage in human mitochondrial DNA. Am J Hum Genet. 2003; 72(1):32–47. https://doi.org/10.1086/345378 PMID: 12489041

73. Adachi N, Shinoda K-i, Umetsu K, Matsumura H. Mitochondrial DNA analysis of Jomon skeletons from the Funadomari site, Hokkaido, and its implication for the origins of Native American. Am J Phys Anthropol. 2009; 138(3):255–65. https://doi.org/10.1002/ajpa.20923 PMID: 18951391

74. Umetsu K, Tanaka M, Yuasa I, Adachi N, Miyoshi A, Kashimura S, et al. Multiplex amplified product-length polymorphism analysis of 36 mitochondrial single-nucleotide polymorphisms for haplogrouping of East Asian populations. Electrophoresis. 2005; 26(1):91–8. https://doi.org/10.1002/elps.200406129 PMID: 15624129

75. Umetsu K, Tanaka M, Yuasa I, Saitou N, Takeyasu T, Fuku N, et al. Multiplex amplified product-length polymorphism analysis for rapid detection of human mitochondrial DNA variations. Electrophoresis. 2001; 22(16):3533–8. https://doi.org/10.1002/1522-2683(200109)22:16<3533::AID-ELPS3533>3.0.CO;2-S PMID: 11669538

76. Adachi N, Umetsu K, Takigawa W, Sakaue K. Phylogenetic analysis of the human ancient mitochondrial DNA. J Archaeol Sci. 2004; 31(10):1339–48. https://doi.org/10.1016/j.jas.2004.02.011.

77. Adachi N, Shinoda K-i, Umetsu K, Kitano T, Matsumura H, Fujiyama R, et al. Mitochondrial DNA analysis of Hokkaido Jomon skeletons: Remnants of archaic maternal lineages at the southwestern edge of former Beringia. Am J Phys Anthropol. 2011; 146(3):346–60. https://doi.org/10.1002/ajpa.21561 PMID: 21953438

78. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc. 2010; 2010(6):pdb.prot5448. https://doi.org/10.1101/pdb.prot5448 PMID: 20516186

79. Shinoda K-i, Adachi N. Ancient DNA Analysis of Palaeolithic Ryukyu Islanders. In: Piper P, Matsumura H, Bulbeck D, editors. New Perspectives in Southeast Asian and Pacific Prehistory. Terra Australis 45. Canberra: ANU Press; 2017. p. 51–60.

80. Maricic T, Whitten M, Pääbo S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS One. 2010; 5(11):e14004. https://doi.org/10.1371/journal.pone.0014004 PMID: 21103372

81. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009; 25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

82. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes. 2016; 9(1):88. https://doi.org/10.1186/s13104-016-1900-2 PMID: 26868221

83. Kanzawa-Kiriyama H, Kryukov K, Jinam TA, Hosomichi K, Saso A, Suwa G, et al. A partial nuclear genome of the Jomons who lived 3000 years ago in Fukushima, Japan. J Hum Genet. 2016; 62:213. https://doi.org/10.1038/jhg.2016.110 PMID: 27581845

84. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 1999; 23:147. https://doi.org/10.1038/13779 PMID: 10508508

85. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20(9):1297–303. https://doi.org/10.1101/gr.107524.110 PMID: 20644199

86. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat. 2009; 30(2):E386–E94. https://doi.org/10.1002/humu.20921 PMID: 18853457

87. Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt H-J, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. 2016; 44(W1):W58–W63. https://doi.org/10.1093/nar/gkw233 PMID: 27084951

88. Skoglund P, Northoff BH, Shunkov MV, Derevianko AP, Pääbo S, Krause J, et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. PNAS. 2014; 111 (6):2229–34. https://doi.org/10.1073/pnas.1318934111 PMID: 24469802

89. Renaud G, Slon V, Duggan AT, Kelso J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. 2015; 16(1):224. https://doi.org/10.1186/s13059-015-0776-0 PMID: 26458810

90. Meyer M, Arsuaga J-L, de Filippo C, Nagel S, Aximu-Petri A, Nickel B, et al. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. Nature. 2016; 531:504–7. https://doi.org/10.1038/nature17405 PMID: 26976447

91. Brothwell DR. Upper Pleistocene human skull from Niah Caves, Sarawak. Sarawak Mus J. 1960; 9:323–49.

92. Demeter F, Shackelford LL, Bacon A-M, Duringer P, Westaway K, Sayavongkhamdy T, et al. Anatomically modern human in Southeast Asia (Laos) by 46 ka. PNAS. 2012; 109(36):14375–80. https://doi.org/10.1073/pnas.1208104109 PMID: 22908291

93. Kennedy KAR. The deep skull of Niah: An assessment of twenty years of speculation concerning its evolutionary significance. Asian Persp. 1977; 20(1):32–50.

94. Macintosh NWG, Barker BCW. The Tabon cave mandible. Archaeology and Physical Anthropology in Oceania. 1978; 13(2–3):143–66.

95. Hà VT. The Hoabinhian in the context of Vietnam. Vietnam Stud. 1976; 46:127–97.

96. Trevor JC. The human remains of Mesolithic and Neolithic date from Gua Cha, Kelantan. Federation Museums Journal. 1962; 7:6–22.

97. White JC. Emergence of cultural diversity in mainland Southeast Asia: a view from prehistory. In: Enfield N, editor. Dynamics of human diversity. Canberra: Pacific Linguistics; 2011. p. 9–46.

98. Yi S, Lee J-J, Kim S, Yoo Y, Kim D. New data on the Hoabinhian: Investigations at Hang Cho Cave, northern Vietnam. Bulletin of the Indo-Pacific Prehistory Association. 2008; 28:73–9.

99. Matsumura H, Pookajorn S. A morphometric analysis of the Late Pleistocene human skeleton from the Moh Khiew Cave in Thailand. Homo. 2005; 56(2):93–118. http://dx.doi.org/10.1016/j.jchb.2005.05.004. PMID: 16130834

100. Matsumura H, Yoneda M, Dodo Y, Oxenham MF, Nguyễn LC, Nguyễn KT, et al. Terminal Pleistocene human skeleton from Hang Cho Cave, northern Vietnam: implications for the biological affinities of Hoabinhian people. Anthropol Sci. 2008; 116(3):201–17.

101. Matsumura H, Zuraina M. Metric analyses of an early Holocene human skeleton from Gua Gunung Runtuh, Malaysia. Am J Phys Anthropol. 1999; 109(3):327–40. https://doi.org/10.1002/(SICI)1096-8644(199907)109:3<327::AID-AJPA4>3.0.CO;2-5 PMID: 10407463

102. Oxenham MF, Buckley HR. The population history of Mainland and Island Southeast Asia. In: Oxenham MF, Buckley H, editors. The Routledge Handbook of Bioarchaeology in Southeast Asia and the Pacific. London: Routledge; 2016. p. 9–23.

103. Bellwood P. First farmers: The origins of agricultural societies. Malden: Blackwell publishers; 2005.

104. Bellwood P. First migrants: ancient migration in global perspective. Chichester: Wiley-Blackwell; 2013 2013.

105. Bellwood P. 36 Southeast Asian islands: Archaeology. In: Ness I, editor. The encyclopedia of global human migration. Hoboken: Blackwell Publishing; 2013.

106. Bellwood P. First islanders: Prehistory and human migration in Island Southeast Asia. Oxford: Wiley Blackwell; 2017.

107. Bellwood P, Gillespie R, Thompson GB, Vogel JS, Ardika IW, Datan I. New dates for prehistoric Asian rice. Asian Persp. 1992; 31(2):161–70.

108. Diamond J, Bellwood P. Farmers and their languages: The first expansions. Science. 2003; 300 (5619):597–603. https://doi.org/10.1126/science.1078208 PMID: 12714734

109. Glover IC, Higham CF. New evidence for early rice cultivation in South, Southeast and East Asia. In: Harris DR, editor. The origins and spread of agriculture and pastoralism in Eurasia. London: University College London Press; 1996. p. 413–41.

110. Higham CFW. Archaeology, linguistics and the expansion of the East and Southeast Asian Neolithic. In: Blench R, Spriggs M, editors. Archaeology and Language II: Archaeological data and linguistic hypotheses. London: Routledge; 1998. p. 103–14.

111. Higham CFW. Prehistory, language and human biology: Is there a consensus in East and Southeast Asia? In: Jin I, Seielstad M, Xiao CJ, editors. Genetic, Linguistic and Archaeological Perspectives on Human Diversity in Southeast Asia. Recent Advances in Human Biology. Singapore: World Scientific; 2001. p. 3–16.

112. Higham CFW. 34 Southeast Asian mainland: archaeology. In: Ness I, editor. The Encyclopedia of Global Human Migration. Hoboken: Blackwell Publishing; 2013.

113. Renfrew C. Models of change in language and archaeology. TPS. 1989; 87(2):103–55. https://doi.org/10.1111/j.1467-968X.1989.tb00622.x

114. Renfrew C. Archaeology and language: the puzzle of Indo-European origins. Cambridge: Cambridge University Press; 1990.

115. Renfrew C. World languages and human dispersals: a minimalist view. In: Hall JA, Jarvie IC, editors. Transition to Modernity: essays on power, wealth and belief. Cambridge: Cambridge University Press; 1992. p. 11–68.

116. Sagart L. The expansion of Setaria farmers in East Asia: a linguistic and archaeological model. In: Sanchez-Mazas A, Blench R, Ross MD, Peiros I, Lin M, editors. Past human migrations in East Asia: Matching archaeology, linguistics and genetics. New York: Routledge; 2008. p. 133–58.

117. Bellwood PS, Renfrew C. Examining the farming/language dispersal hypothesis. Cambridge: McDonald Institute for Archaeological Research; 2002.

118. Crawford GW, Shen C. The origins of rice agriculture: recent progress in East Asia. Antiquity. 1998; 72(278):858–66. Epub 2015/01/02. https://doi.org/10.1017/S0003598X00087494

119. Lu TLD. The occurrence of cereal cultivation in China. Asian Persp. 2006; 45(2):129–58.

120. Xingcan C. On the earliest evidence for rice cultivation in China. Bulletin of the Indo-Pacific Prehistory Association. 1999; 18:81–94.

121. Zhang C, Hung H-C. The emergence of agriculture in Southern China. Antiquity. 2010; 83:1–15.

122. Zhang C, Hung H-c. 26 Eastern Asia: Archaeology. In: Ness I, editor. The Encyclopedia of Global Human Migration. Hoboken: Blackwell Publishing; 2013.

123. Cox MP. 37 Southeast Asian islands and Oceania: Human genetics. In: Ness I, editor. The encyclopedia of global human migration. Hoboken: Blackwell Publishing; 2013.

124. Cox MP, Karafet TM, Lansing JS, Sudoyo H, Hammer MF. Autosomal and X-linked single nucleotide polymorphisms reveal a steep Asian–Melanesian ancestry cline in eastern Indonesia and a sex bias in admixture rates. Proceedings of the Royal Society B: Biological Sciences. 2010.

125. Hill C, Soares P, Mormina M, Macaulay V, Clarke D, Blumbach PB, et al. A mitochondrial stratigraphy for Island Southeast Asia. AJHG. 2007; 80(1):29–43. https://doi.org/10.1086/510412.

126. Soares P, Trejaut JA, Loo J-H, Hill C, Mormina M, Lee C-L, et al. Climate change and postglacial human dispersals in Southeast Asia. Mol Biol Evol. 2008; 25(6):1209–18. https://doi.org/10.1093/molbev/msn068 PMID: 18359946

127. Peng M-S, Quang HH, Dang KP, Trieu AV, Wang H-W, Yao Y-G, et al. Tracing the Austronesian footprint in Mainland Southeast Asia: A perspective from mitochondrial DNA. Mol Biol Evol. 2010; 27(10):2417–30. https://doi.org/10.1093/molbev/msq131 PMID: 20513740

128. Capelli C, Wilson JF, Richards M, Stumpf MPH, Gratrix F, Oppenheimer S, et al. A Predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. AJHG. 2001; 68(2):432–43. https://doi.org/10.1086/318205.

129. The HUGO Pan-Asian SNP Consortium. Mapping human genetic diversity in Asia. Science. 2009; 326 (5959):1541–5. https://doi.org/10.1126/science.1177074 PMID: 20007900

130. Ricaut F-X, Bellatti M, Lahr MM. Ancient mitochondrial DNA from Malaysian hair samples: Some indications of Southeast Asian population movements. Amer J Hum Biol. 2006; 18(5):654–67. https://doi.org/10.1002/ajhb.20535 PMID: 16917897

131. Derenko M, Malyarchuk B, Denisova G, Perkova M, Rogalla U, Grzybowski T, et al. Complete mitochondrial DNA analysis of Eastern Eurasian haplogroups rarely found in populations of northern Asia and eastern Europe. PLoS One. 2012; 7(2):e32179. https://doi.org/10.1371/journal.pone.0032179 PMID: 22363811

132. Ko Albert M-S, Chen C-Y, Fu Q, Delfin F, Li M, Chiu H-L, et al. Early Austronesians: Into and Out Of Taiwan. AJHG. 2014; 94(3):426–36. https://doi.org/10.1016/j.ajhg.2014.02.003.

133. Tabbada KA, Trejaut J, Loo J-H, Chen Y-M, Lin M, Mirazón-Lahr M, et al. Philippine mitochondrial DNA diversity: A populated viaduct between Taiwan and Indonesia? Mol Biol Evol. 2010; 27(1):21–31. https://doi.org/10.1093/molbev/msp215 PMID: 19755666

134. Jinam TA, Hong L-C, Phipps ME, Stoneking M, Ameen M, Edo J, et al. Evolutionary History of Continental Southeast Asians: "Early Train" Hypothesis Based on Genetic Analysis of Mitochondrial and Autosomal DNA Data. Mol Biol Evol. 2012; 29(11):3513–27. https://doi.org/10.1093/molbev/mss169 PMID: 22729749