The prehistoric peopling of Southeast Asia

Hugh McColl1, Fernando Racimo1, Lasse Vinner1, Fabrice Demeter1,2, Takashi Gakuhari3,4, J. Victor Moreno-Mayar4, George van Driem10, Uffe Gram Wilken5, Andaine Seguin-Orlando6,7, Constanza de la Fuente Castro7, Sally Wasef8, Rasmil Shoocoodje9, Viengkoe Ouksavat9

The human occupation history of Southeast Asia (SEA) remains heavily debated. Current evidence suggests that SEA was occupied by Hōabìnhian hunter-gatherers until ~4000 years ago, when farming economies developed and expanded, restricting foraging groups to remote habitats. Some argue that agricultural development was indigenous; others favor the "two-layer" hypothesis that posits a southward expansion of farmers giving rise to present-day Southeast Asian genetic diversity. By sequencing 26 ancient human genomes (25 from SEA, 1 Japanese), we show that neither interpretation fits the complexity of Southeast Asian prehistory: Both Hōabìnhian hunter-gatherers and East Asian farmers contributed to current Southeast Asian diversity, with further migrations affecting island SEA and Vietnam. Our results help resolve one of the long-standing controversies in Southeast Asian prehistory.

Anatomically modern humans expanded into Southeast Asia (SEA) at least 65 thousand years ago (2, 3). The two oldest samples—Hōabìnhian (4) and Natufian (5)—of present-day Southeast Asian populations suggest that diversity was influenced by later migrations involving rice and millet farmers from the north (4). These observations have generated two competing hypotheses: One states that the Hōabínhan hunter-gatherers adopted agriculture without substantial external gene flow (6, 7), and the other (the "two-layer" hypothesis) states that farmers from East Asia (EA) replaced the indigenous Hōabínhan inhabitants ~4 ka ago (8, 9). Studies of present-day populations have not resolved the extent to which migrations from EA affected the genetic makeup of SEA.

Obtaining ancient DNA evidence from SEA is challenging because of poor preservation conditions (10). We thus tested different whole-human-genome capture approaches and found that a modification of Methyl-Enrichment perfomed best (11). We applied this method together with standard shotgun sequencing to DNA extracted from human skeletal material from Malaysia, Thailand, the Philippines, Vietnam, Indonesia, Laos, and Japan dating between 0.2 and 8 ka ago (11). We obtained 26 low-coverage ancient whole genomes, including those of a Japanese Ikawazu Jōmon individual and Hōabínhan hunter-gatherers from Malaysia and Laos, as well as Late Neolithic, Bronze Age, and Iron Age farmers from across SEA (Fig. 1 and table S1) (11). We also sequenced mitochondrial DNA from 16 additional ancient individuals and high-coverage whole genomes from two present-day Jehai individuals from Northern Parak state, West Malaysia (table S3). All samples showed damage patterns typical of ancient DNA and minimal amounts of contamination (table S3) (11).

We performed a principal component analysis (PCA) of worldwide present-day populations (12, 13) to find the strongest axes of genetic variation in our data and projected the ancient individuals onto the first two principal components. The two oldest samples—Hōabínhan from Pha Faen, Laos (La368; 7950 to 7795 calendar years before the present [cal B.P.]) and Guà Cha, Malaysia (Ma911; 4415 to 4160 cal B.P.)—henceforth labeled "group 1," cluster most closely with present-day Onge from the Andaman Islands and away from other East Asian and Southeast Asian populations (Fig. 2), a pattern that differentiates them from all other ancient samples. We used ADMIXTURE (14) and fastNGSadiimix (15) to model ancient genomes as mixtures of latent ancestry components (11). Group

1Centre for GeoGenetics, Natural History Museum of Denmark, Copenhagen, Denmark. 2National Museum of Natural History, Écoanthropology and Ethnobiology, Musée de l’Homme, Paris, France. 3Center for Cultural Resource Studies, Kanazawa University, Kanazawa, Japan. 4Kitsato University School of Medicine, Sagamihara, Kanagawa, Japan. 5Institut für Sprachwissenschaft, Universität Bern, Bern, Switzerland. 6University of New England, Armidale, NSW, Australia. 7Laboratoire AMIS, Université Paul Sabatier (UPS), Toulouse, France. 8Australian Research Centre for Human Evolution, Griffith University, Nathan, QLD, Australia. 9Department of Archaeology, Faculty of Archaeology, Silpakorn University, Bangkok, Thailand. 10Department of Heritage, Ministry of Information, Culture and Tourism, Vientiane, Lao People’s Democratic Republic. 11Centre for Global Archaeological Research, Universiti Sains Malaysia, Penang, Malaysia. 12Department of Bioinformatics and Genomics, Graduate School of Advanced Preventive Medical Sciences, Kanagawa University, Kanagawa, Japan. 13Department of Computational Biology, University of Lausanne and SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland. 14Jeffrey Cheah School of Medicine & Health Sciences, Monash University Malaysia, Jalan Lagoon Selatan, Sunway City, Selangor, Malaysia. 15Department of Zoology, University of Cambridge, Cambridge, UK. 16Institute of Molecular Biology, National Academy of Sciences, Yerevan, Armenia. 17Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand. 18Anthropological and Paleoenvironmental Department, Institute of Archaeology, Hanoi, Vietnam. 19Department of Archaeology and Natural History, Australian National University, Canberra, ACT, Australia. 20Balai Archeology, Medan, Indonesia. 21Nara National Research Institute for Cultural Properties, Nara, Japan. 22University Museum, University of Tokyo, Tokyo, Japan. 23Graduate School of Medicine, University of the Ryukyus, Nishihara, Okinawa, Japan. 24Educational Committee of Tahara City, Tahara, Japan. 25National Museum of Japanese History, Sakura, Chiba, Japan. 26Division of Genomics, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan. 27Center for Information Biology, National Institute of Genetics, Mizamisato, Japan. 28National University of Singapore, School of Medicine, Trinity College Dublin, University of Dublin, Dublin, Ireland. 29Oxford Radiocarbon Accelerator Unit (GRAU), University of Oxford, Oxford, UK. 30Laboratoire AMIS, Université Paris Descartes, Faculté de Chirurgie Dentaire, Montirose, France. 31Ecole et Observatoire des Sciences des la Terre, Université de Strasbourg, Strasbourg, France. 32Institut de Physique du Globe de Strasbourg (IPGS) (CNRS/US 365 UMR 7516), Strasbourg, France. 33Laboratory "Image Ville et Environnement LIVE," UMR7362, CNRS and Université de Strasbourg, Strasbourg, France. 34Department of Anthropology, University of Illinois at Urbana-Champaign, Champaign, IL, USA. 35Natural History Museum of La Rochelle, La Rochelle, France. 36CNRS, UMR7055 "Préhistoire et Technologie," Musée Archéologie et Ethnologie, Nantes, France. 37CNRS, UMR7359 "Méthodes et Matériaux," Musée Archéologie et Ethnologie, Nantes, France. 38CNRS, UMR7359 "Méthodes et Matériaux," Musée Archéologie et Ethnologie, Nantes, France. 39Laboratoire "Image Ville et Environnement LIVE," UMR7362, CNRS and Université de Strasbourg, Strasbourg, France. 40Department of Anthropology, University of Illinois at Urbana-Champaign, Champaign, IL, USA. 41Natural History Museum of La Rochelle, La Rochelle, France. 42CNRS, UMR7055 "Préhistoire et Technologie," Musée Archéologie et Ethnologie, Nantes, France. 43Laboratoire "Image Ville et Environnement LIVE," UMR7362, CNRS and Université de Strasbourg, Strasbourg, France. 44Laboratory "Image Ville et Environnement LIVE," UMR7362, CNRS and Université de Strasbourg, Strasbourg, France. 45Department of Bioinformatics and Genomics, Graduate School of Advanced Preventive Medical Sciences, Kanagawa University, Kanagawa, Japan. 46Department of Anthropology, University of Cambridge, Cambridge, UK. 47Department of Anthropology and Archaeology, University of Otago, Dunedin, New Zealand. 48St. Catharine’s College, University of Cambridge, Cambridge, UK. 49Wellcome Trust Sanger Institute, Hinxton, UK. 50*These authors contributed equally to this work.

*Corresponding author. Email: ewillerslev@smn.ku.dk
Fig. 1. Maps of ages and differential ancestry of ancient Southeast Asian genomes. (A) Estimated mean sample ages for ancient individuals. (B to D) D statistics testing for differential affinity between (B) Papuans and Tiányuán (2240k dataset), (C) Önge and Tiányuán (2240k dataset), and (D) Mlabri and Hàn Chinese (Pan-Asia dataset).

Fig. 2. Exploratory analyses of relationships of ancient Southeast Asian genomes to those of present-day populations. Ancient samples are projected on the first two components of PCAs for (A) worldwide populations and (B) a subset of populations from EA and SEA. (C) fastNGSadmix plot at $K = 13$ (II). We refer to the following present-day language-speaking groups in relation to our ancient samples: Austroasiatic (bright green), Austroasiatic (pink), and Hmong-Mien (dark pink), along with a broad East Asian component (dark green). P.M., proto-Malay; M.N., Malaysian negrito; P.N., Philippines negrito; And. Is., Andaman Islands; NA, not applicable.
individuals differ from the other Southeast Asian ancient samples in containing components shared with the supposed descendants of the Hoba’inhans: the Önge and the Jehai (Peninsular Malaysia), along with groups from India and Papua New Guinea.

We also find a distinctive relationship between the group 1 samples and the Iwakawazumon from Japan (IK002). Outgroup D-statistics fit group 1 samples more closely than Tiányuán (Fig. 1) (17). This pattern of complex, localized admixture is evident in the Jehai, fitted as an admixed population receiving ancestry from group 1/Önge and a population related to East Asians (Amis), whereas present-day Japanese can be modeled as a mixture of Önge and an additional East Asian component (Fig. 3 and fig. S29).

Group 6 individuals (1880 to 299 cal B.P.) originate from Malaysia and the Philippines and cluster with present-day Austroasiatic (Fig. 2) (18). Group 6 also contains Ma554, having the highest amounts of Denisovan-like ancestry relative to the other ancient samples, although we observe little variation in this archaic ancestry in our samples from MSEA (18).

The remaining ancient individuals are modeled in fastNGSadmix as containing East Asian and Southeast Asian components present in high proportions in present-day Austroasiatic, Austroasiatic, and Hmong-Mien speakers, along with a broad East Asian component. A PCA including only East Asian and Southeast Asian populations that did not show considerable Papuan or Önge-like ancestry (fig. S31) separates the present-day speakers of ancestral language families in the region: Trans-Himalayan (formerly Sino-Tibetan), Austroasiatic, and Austroasiatic/Kradai (20). The ancient individuals form five slightly differentiated clusters (groups 2 to 6) (Fig. 1B), in accordance with fastNGSadmix and D-statistics (Fig. 2 and figs. S12 to S19) (17)

Group 2 contains late Neolithic and early Bronze Age individuals (4291 to 2184 cal B.P.), from Vietnam, Laos, and the Malay Peninsula who are closely related to present-day Austroasiatic language speakers such as the Mlabri and Htin (Fig. 1) (17). Compared with groups 3 to 6, group 2 individuals lack a broad East Asian ancestry component that is at its highest proportion in northern EA in fastNGSadmix. TreeMix analyses suggest that the two individuals with the highest coverage in group 2 (La364 and Ma912) form a clade resulting from admixture between the ancestors of East Asians and of La368 (Fig. 3 and figs. S24 to S27). This pattern of complex, localized admixture is also evident in the Jehai, fitted as an admixed population between group 2 (Ma912) and the branch leading to present-day Önge and La368 (fig. S28). Consistent with these results, La364 is best modeled as a mixture of a population ancestral to Amis and the group 1/Önge-like population (Fig. 3). The best model for present-day Dai populations is a mixture of group 2 individuals and a pulse of admixture from East Asians (fig. S39).

The Önge and the Jehai (Peninsular Malaysia), along with groups from India and Papua New Guinea. We also find a distinctive relationship between the group 1 samples and the Iwakawazumon from Japan (IK002). Outgroup D-statistics fit group 1 samples more closely than Tiányuán (Fig. 1) (17). This pattern of complex, localized admixture is evident in the Jehai, fitted as an admixed population receiving ancestry from group 1/Önge and a population related to East Asians (Amis), whereas present-day Japanese can be modeled as a mixture of Önge and an additional East Asian component (Fig. 3 and fig. S29).

Group 6 individuals (1880 to 299 cal B.P.) originate from Malaysia and the Philippines and cluster with present-day Austroasiatic (Fig. 2) (18). Group 6 also contains Ma554, having the highest amounts of Denisovan-like ancestry relative to the other ancient samples, although we observe little variation in this archaic ancestry in our samples from MSEA (18).

The remaining ancient individuals are modeled in fastNGSadmix as containing East Asian and Southeast Asian components present in high proportions in present-day Austroasiatic, Austroasiatic, and Hmong-Mien speakers, along with a broad East Asian component. A PCA including only East Asian and Southeast Asian populations that did not show considerable Papuan or Önge-like ancestry (fig. S31) separates the present-day speakers of ancestral language families in the region: Trans-Himalayan (formerly Sino-Tibetan), Austroasiatic, and Austroasiatic/Kradai (20). The ancient individuals form five slightly differentiated clusters (groups 2 to 6) (Fig. 1B), in accordance with fastNGSadmix and D-statistics (Fig. 2 and figs. S12 to S19) (17).

Group 2 contains late Neolithic and early Bronze Age individuals (4291 to 2184 cal B.P.), from Vietnam, Laos, and the Malay Peninsula who are closely related to present-day Austroasiatic language speakers such as the Mlabri and Htin (Fig. 1) (17). Compared with groups 3 to 6, group 2 individuals lack a broad East Asian ancestry component that is at its highest proportion in northern EA in fastNGSadmix. TreeMix analyses suggest that the two individuals with the highest coverage in group 2 (La364 and Ma912) form a clade resulting from admixture between the ancestors of East Asians and of La368 (Fig. 3 and figs. S24 to S27). This pattern of complex, localized admixture is also evident in the Jehai, fitted as an admixed population between group 2 (Ma912) and the branch leading to present-day Önge and La368 (fig. S28). Consistent with these results, La364 is best modeled as a mixture of a population ancestral to Amis and the group 1/Önge-like population (Fig. 3). The best model for present-day Dai populations is a mixture of group 2 individuals and a pulse of admixture from East Asians (fig. S39).

11 individuals from the other Southeast Asian ancient samples in containing components shared with the supposed descendants of the Hoba’inhans: the Önge and the Jehai (Peninsular Malaysia), along with groups from India and Papua New Guinea. We also find a distinctive relationship between the group 1 samples and the Iwakawazumon from Japan (IK002). Outgroup D-statistics fit group 1 samples more closely than Tiányuán (Fig. 1) (17). Using TreeMix and qpGraph (16, 19) to explore admixture graphs that could potentially fit our data, we find that group 1 individuals are best modeled as a sister group to present-day Önge (Fig. 3, and figs. S21 to S23 and S35 to S37). Finally, the Önge individual is best-modeled as a mix between a population related to group 1/Önge and a population related to East Asians (Amis), whereas present-day Japanese can be modeled as a mixture of Önge and an additional East Asian component (Fig. 3 and fig. S29).

The remaining ancient individuals are modeled in fastNGSadmix as containing East Asian and Southeast Asian components present in high proportions in present-day Austroasiatic, Austroasiatic, and Hmong-Mien speakers, along with a broad East Asian component. A PCA including only East Asian and Southeast Asian populations that did not show considerable Papuan or Önge-like ancestry (fig. S31) separates the present-day speakers of ancestral language families in the region: Trans-Himalayan (formerly Sino-Tibetan), Austroasiatic, and Austroasiatic/Kradai (20). The ancient individuals form five slightly differentiated clusters (groups 2 to 6) (Fig. 1B), in accordance with fastNGSadmix and D-statistics (Fig. 2 and figs. S12 to S19) (17).

Group 2 contains late Neolithic and early Bronze Age individuals (4291 to 2184 cal B.P.), from Vietnam, Laos, and the Malay Peninsula who are closely related to present-day Austroasiatic language speakers such as the Mlabri and Htin (Fig. 1) (17). Compared with groups 3 to 6, group 2 individuals lack a broad East Asian ancestry component that is at its highest proportion in northern EA in fastNGSadmix. TreeMix analyses suggest that the two individuals with the highest coverage in group 2 (La364 and Ma912) form a clade resulting from admixture between the ancestors of East Asians and of La368 (Fig. 3 and figs. S24 to S27). This pattern of complex, localized admixture is also evident in the Jehai, fitted as an admixed population between group 2 (Ma912) and the branch leading to present-day Önge and La368 (fig. S28). Consistent with these results, La364 is best modeled as a mixture of a population ancestral to Amis and the group 1/Önge-like population (Fig. 3). The best model for present-day Dai populations is a mixture of group 2 individuals and a pulse of admixture from East Asians (fig. S39).

Group 6 individuals (1880 to 299 cal B.P.) originate from Malaysia and the Philippines and cluster with present-day Austroasiatic (Fig. 2) (18). Group 6 also contains Ma554, having the highest amounts of Denisovan-like ancestry relative to the other ancient samples, although we observe little variation in this archaic ancestry in our samples from MSEA (18).

Group 5 (2304 to 1818 cal B.P.) contains two individuals from Indonesia, modeled by fastNGSadmix as a mix of Austroasiatic- and Austroasiatic-like ancestry, similar to present-day western Indonesians, a finding consistent with their position in the PCA (Fig. 2) (18). Indeed, after Mlabri and Htin, the present-day populations sharing the most drift with group 2 are western Indonesian samples from Bali and Java previously identified as having mainland Southeast Asian ancestry (21) (fig. S33). TreeMix models the group 5 individuals as an admixed population receiving ancestry related to group 2 (figs. S30 and S31) and Amis. Despite the clear relationship with the mainland group 2 seen in all analyses, the small ancestry components in group 5 related to Jehai and Papuans visible in fastNGSadmix may be remnants of ancient Sundaland ancestry. These results suggest that group 2 and group 5 are related to a...
we observe a change in ancestry by ~4 ka ago, supporting a demographic expansion from EA into SEA during the Neolithic transition to farming. However, despite changes in genetic structure coinciding with this transition, evidence of admixture indicates that migrations from EA did not simply replace the previous occupants. Additionally, late Neolithic farmers share ancestry with present-day Austronesian-speaking hill tribes, in agreement with the hypotheses of an early Austronesian farmer expansion (20). By 2 ka ago, Southeast Asian individuals carried additional East Asian ancestry components absent in the late Neolithic samples, much like present-day populations. One component likely represents the introduction of ancestral Kradai languages in MSEA (11), and another the Austronesian expansion into ISEA reaching Indonesia by 2.1 ka ago and the Philippines by 1.8 ka ago. The evidence described here favors a complex model including a demographic transition in which the original Hôabìnhians admixed with multiple incoming waves of East Asian migration associated with the Austronesian, Kra-dai, and Austronesian language speakers.

**REFERENCES AND NOTES**

4. C. Higham, Early mainland Southeast Asia: From First Humans to Angkor (River Books, 2014).

Supplementary text.


**ACKNOWLEDGMENTS**

We thank the National High-throughput DNA Sequencing Centre (Copenhagen Denmark) for advice and sequencing of samples, the Duckworth laboratory, University of Cambridge, for access to materials, K. Gregersen for making casts of teeth before sampling, and P. Tacon, ARCHE, Griffith University for assistance with sample transfer. E.W. thanks St. John’s College, University of Cambridge, for providing an inspiring environment for scientific thought.

**Funding:** This work was supported by the Lundbeck Foundation, the Danish National Research Foundation, and the KU2016 program. H.Mc. is supported by the University of Adelaide’s George Murray Scholarship, R.S. thanks the Thailand Research Fund (TRF) for support (grants RTA6080001 and RDGS5H0006). The excavation of the Jomón individual was supported by a Grant-in-Aid for Scientific Research (B) (25284517) to Y.Y. The Jomón genome project was organized by H.I., as well as T.H. and H.O., who were supported by MEXT KAKENI grants 16H06408 and 17H05132, and a Grants-in-Aid for Challenging Exploratory Research (23657167) and for Scientific Research (B) 17H03738. The Jomón genome sequencing was supported by JSPS KAKENHI grant 16H06270 to T.A. and partly funded by the CHOZEN project in Kanazawa University and the Cooperative Research Project Program of the Medical Institute of Bioregulation, Kyushu University. Computations for the Jomón genome were partially performed on the NIG supercomputer at RIOS National Institute of Genetics. M.M.L. is supported by the ERC award 295057. D.M.L. was supported by ARC grants LP120200144, LP150100583, and DP170103133. A.P. is supported by Leverhulme Project Research grant RPG-2016-235. M.E.P. acknowledges the Cardio-Metabolic research cluster at Jeffrey Cheah School of Medicine & Health Sciences, TMB research platform, Monash University Malaysia, and MOSTI Malaysia for research grant 100-RM/BiOTEK 16/6/28. A.S. was financed by the European Research Council (starting grant) and the Swiss National Science Foundation. **Author contributions:** E.W. initiated and led the study. E.W., D.M.L., L.V., M.E.A., H.O., M.E.D., A.S.M., L.D., H.Mc., and F.D. designed the study. E.W. and D.M.L. supervised the overall project. L.V., F.D., F.R., V.S., T.S., M.M.S., R.S., T.M.HN, C.H., K.W., E.P.E., J.C.G., R.H., H.B., C.P., H.I., T.H., M.E.D., F.A.A., A.S.M., and H.O. supervised specific aspects of the project. H.Mc., L.V., F.D., U.G.W., C.D., M.E.A., V.S., T.S., M.M.S., R.S., K.P., A.T., V.M.H., H.C.H., T.M.T., T.H.N., S.S., G.H.N., K.W., N.S., T.Y., A.M., B.D., J.L.P., L.S., E.P.E., N.A.T., B.B.P., J.C.G., R.H., M.E.D., A.F.A., and G.P. curated, cataloged, sampled, and/or provided samples. H.Mc., L.V., T.G., A.S.O., S.W., P.B.D., M.Y., A.Ta., H.S., A.To., S.R., T.D., M.E.D., F.A.A., and S.W. generated and produced data for analysis. H.Mc., F.R., L.V., T.G., J.V.M.M., C.D., K.T., S.S., H.Ma., S.W., S.A.M., A.Ta., M.E.D., L.D., M.E.S. analyzed or assisted in the analysis of data. H.Mc., F.R., L.V., F.D., T.G., A.M., L.O., M.S., N.S., A.T., F.D., and E.W. wrote the manuscript with considerable input from J.M.V.M., C.D., K.T., S.S., H.Ma., S.W., S.A.M., and M.E.S. M.A. supervised the final manuscript. Competing interests: The authors declare no competing interests. Data and materials availability: This study has been evaluated by the Danish Bioethical Committee (H-16018872) and the Department of Orang Asli Affairs, Malaysia, and the Department of Orang Asli Affairs, Malaysia (JHEOA PP.XX.052 [15.5.17]). MOU’s exist with local institutions where the sampling took place. Genomic data are available for download at the ENA (European Nucleotide Archive) with accession number PRJEB26712.

**SUPPLEMENTARY MATERIALS**

www.sciencemag.org/content/361/6397/88/suppl/DC1
Supplementary Text
Figs. S1 to S43
Tables S1 to S26
References (27–111)
21 February 2018; accepted 29 May 2018
10.1126/science.aai3628
The prehistoric peopling of Southeast Asia


Science 361 (6397), 88-92.
DOI: 10.1126/science.aat3628

Ancient migrations in Southeast Asia
The past movements and peopling of Southeast Asia have been poorly represented in ancient DNA studies (see the Perspective by Bellwood). Lipson et al. generated sequences from people inhabiting Southeast Asia from about 1700 to 4100 years ago. Screening of more than a hundred individuals from five sites yielded ancient DNA from 18 individuals. Comparisons with present-day populations suggest two waves of mixing between resident populations. The first mix was between local hunter-gatherers and incoming farmers associated with the Neolithic spreading from South China. A second event resulted in an additional pulse of genetic material from China to Southeast Asia associated with a Bronze Age migration. McColl et al. sequenced 26 ancient genomes from Southeast Asia and Japan spanning from the late Neolithic to the Iron Age. They found that present-day populations are the result of mixing among four ancient populations, including multiple waves of genetic material from more northern East Asian populations. Science, this issue p. 92, p. 88; see also p. 31