Ancient genomes document multiple waves of migration in Southeast Asian prehistory

Mark Lipson,1* Olivia Cheronet,2,3,4 Swapan Mallick,1,5 Nadin Rohland,1 Marc Oxenham,6 Michael Pietrusewsky,7 Thomas Oliver Pryce,8,9,10 Anna Willis,11 Hirofumi Matsumura,12 Hallie Buckley,13 Kate Domett,14 Nguyen Giang Hai,15 Trinh Hoang Hiep,15 Aung Aung Kyaw,16 Tin Tin Win,16 Baptiste Pradier,9 Nasreen Broomandkhoshbacht,1,17 Francesca Candido,18,19 Piya Changmai,20 Daniel Fernandes,2,3,21 Matthew Ferry,1,17 Beatriz Gamarra,3,4 Eadaoin Harney,1,17 Jatupol Kampuansai,22,23 Wibhu Kutanan,24 Megan Michel,1,17 Mario Novak,3,25 Jonas Oppenheimer,1,17 Kendra Sirak,3,26 Kristin Stewardson,1,17 Zhao Zhang,1 Pavel Flegontov,20,27† Ron Pinhasi,2,3,16 David Reich1,4,17†‡

1Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. 2Department of Anthropology, University of Vienna, 1090 Vienna, Austria. 3Earth Institute, University College Dublin, Dublin 4, Ireland. 4School of Archaeology, University College Dublin, Dublin 4, Ireland. 5Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. 6School of Archaeology and Anthropology, Australian National University, Canberra, ACT 0200, Australia. 7Department of Anthropology, University of Hawai‘i at Mānoa, Honolulu, Hawaii 96822, USA. 8Centre National de la Recherche Scientifique, 75016 Paris, France. 9UMR 7055 Préhistoire et Technologie, Université Paris Nanterre, 92023 Nanterre, France. 10CEA/CNRS UMR 3685 NIMBE, 91191 Gif-sur-Yvette, France. 11College of Arts, Society and Education, James Cook University, Townsville, Queensland 4811, Australia. 12School of Health Science, Sapporo Medical University, Sapporo O60-8556, Japan. 13Department of Anatomy, University of Otago, Dunedin 9054, New Zealand. 14Division of Tropical Health and Medicine, College of Medicine and Dentistry, James Cook University, Townsville, Queensland 4811, Australia. 15Department of Prehistoric Archaeology, Vietnam Institute of Archaeology, Hanoi, Vietnam. 16Department of Archaeology, Ministry of Religious Affairs and Culture, Mandalay, Myanmar. 17Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA. 18Soprintendenza Archeologia Belle Arti e Paesaggio per la Città Metropolitana di Cagliari e per le Province di Oristano e Sud Sardegna, 09124 Cagliari, Italy. 19Physicoanthropology Section, University of Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, PA 19104, USA. 20Department of Biology and Ecology, Faculty of Science, University of Ostrava, 70103 Ostrava, Czech Republic. 21CIAS, Department of Life Sciences, University of Coimbra, Coimbra 3000-456, Portugal. 22Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand. 23Center of Excellence in Bioresources for Agriculture, Industry and Medicine, Chiang Mai University, Chiang Mai 50200, Thailand. 24Department of Biology, Faculty of Science, Khon Kaen University, Khon Kaen, 40002, Thailand. 25Institute for Anthropological Research, 10000 Zagreb, Croatia. 26Department of Anthropology, Emory University, Atlanta, GA 30322, USA. 27Institute of Parasitology, Biology Centre, Czech Academy of Sciences, 37005 České Budějovice, Czech Republic.

*Corresponding author. Email: mlipson@genetics.med.harvard.edu (M.L.); ron.pinhasi@univie.ac.at (R.P.); reich@genetics.med.harvard.edu (D.R.)
‡P.F. holds part-time positions as a research assistant at the Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic and at the Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia.
§These authors contributed equally to this work.

Southeast Asia is home to rich human genetic and linguistic diversity, but the details of past population movements in the region are not well known. Here, we report genome-wide ancient DNA data from eighteen Southeast Asian individuals spanning from the Neolithic period through the Iron Age (4100–1700 years ago). Early farmers from Man Bac in Vietnam exhibit a mixture of East Asian (southern Chinese agriculturalist) and deeply diverged eastern Eurasian (hunter-gatherer) ancestry characteristic of Austroasiatic speakers, with similar ancestry as far south as Indonesia providing evidence for an expansive initial spread of Austroasiatic languages. By the Bronze Age, in a parallel pattern to Europe, sites in Vietnam and Myanmar show close connections to present-day majority groups, reflecting substantial additional influxes of migrants.

The archaeological record of Southeast Asia documents a complex history of human occupation, with the first archaic hominins arriving at least 1.6 million years ago (yBP) and anatomically modern humans becoming widely established by 50,000 yBP (1–3). Particularly profound changes in human culture were propelled by the spread of agriculture. Rice farming began in the region approximately 4500–4000 yBP and was accompanied by a relatively uniform and widespread suite of tools and pottery styles showing connections to southern China (4–7). It has been hypothesized that this cultural transition was effected by a migration of people who were not closely related to the indigenous hunter-gatherers of Southeast Asia (5, 7–10) and who may have spoken Austroasiatic languages, which today have a wide, but fragmented, distribution in the region (4, 5, 7–14). In this scenario, the languages spoken by the majority of present-day people in Southeast Asia (e.g., Thai, Lao, Myanmar, Malay) reflect later population movements. However, no genetic study has resolved the extent to which the spread of agriculture into the region and subsequent cultural and technological shifts were achieved by movement of people or ideas.
Here we analyze samples from five ancient sites (Table 1; Fig. 1A): Man Bac (Vietnam, Neolithic; 4100–3600 yBP), Nui Nap (Vietnam, Bronze Age; 2100–1900 yBP), Oakaie 1 (Myanmar, Late Neolithic/Brass Age; 3200–2700 yBP (15)), Ban Chiang (Thailand, Late Neolithic through Iron Age; 3500–2400 yBP (16)), and Vat Komnou (Cambodia, Iron Age; 1900–1700 yBP (17)). We initially screened a total of 350 next-generation sequencing libraries generated from petrous bone samples (specifically the high-yield cochlear region (18)) from 146 distinct individuals. For libraries with evidence of authentic ancient DNA, we generated genome-wide data using in-solution enrichment, yielding sequences from eighteen individuals (Table 1, table S1) (19). Because of poor preservation conditions in tropical environments, we observed both a low rate of conversion of screened samples to working data and also limited depth of coverage per sample, and thus we created multiple libraries per individual (102 in total in our final data set).

We initially analyzed the data by performing principal component analysis (PCA) using two different sets of present-day populations (19). First, compared to a set of diverse non-Africans (East and Southeast Asian, Australasian, Central American, and European), the ancient individuals fall close to present-day Chinese and Vietnamese when projected onto the first two axes, with Man Bac, Ban Chiang, and Vat Komnou shifted slightly in the direction of Onge (Andaman Islanders) and Papuan (fig. S1). To focus on East and Southeast Asian diversity, we then used a panel of 16 present-day populations from the region, with three primary directions in the first two dimensions represented by Han Chinese, Austroasiatic-speaking groups (Mlabri and Htin from Thailand, Nicobarese, and Cambodian, but not Kinh), and aboriginal (Austronesian-speaking) Taiwanese (right, left, and top, respectively; Fig. 1B; compare (20)). Man Bac, Ban Chiang (all periods), and Vat Komnou cluster with Austroasiatic speakers, while Nui Nap projects close to present-day Vietnamese and Dai near the center, and Oakaie projects close to present-day Myanmar and other Sino-Tibetan languages. Present-day Lao are intermediate between Austroasiatic speakers and Dai, and western Indonesians (Semene from southern Sumatra and Barito from southeastern Borneo) fall intermediate between Austroasiatic speakers and aboriginal Taiwanese.

We measured levels of allele sharing between populations via outgroup $f_2$-statistics and obtained results consistent with those from PCA (table S2). Nominal, the top sharing for each ancient population is provided by another ancient population, but this pattern may be an artifact due to correlated genotype biases between different ancient samples (table S3). Restricting to present-day comparisons, Man Bac, Ban Chiang, and Vat Komnou share the most alleles with Austroasiatic-speaking groups (as Austroasiatic-speaking groups do with each other); Nui Nap with Austronesian speakers, Dai, and Kinh; and Oakaie with Sino-Tibetan-speaking groups. We also computed statistics $f_3(X, Y, Z)$ for the 40,000 yBP Tianyuan individual (21) as an outgroup and present-day Mixe, Onge, and Atayal, to which we added Man Bac, Nicobarese, and Mlabri. The latter three were inferred to have ancestry from a Southeast Asian farmer-related source (∼70%, forming a clade with Atayal) and a deeply diverging eastern Eurasian source (∼30%, sharing a small amount of drift with Onge; $f_3$-statistics indicate that this source is also not closely related to Papuans, South Asians, or the 40,000 yBP Tianyuan individual (22); table S3). The allele sharing demonstrated by outgroup $f_2$-statistics can be accommodated along the farmer lineage, the deeply-splitting lineage, or a combination of the two, but given the closeness of the mixture proportions among the three groups, we found that the most parsimonious model (Fig. 3, fig. S2) involved a shared ancestral admixture event (29% deep ancestry; 28% omitting VN29), followed by divergence of Man Bac from the present-day Austroasiatic speakers, and lastly a second pulse of deep ancestry (5%) into Nicobarese (19).

Finally, to assess the relationships among additional present-day populations, we fit two extended admixture graphs (figs. S3, S4), with the first including Dai, Semene, Barito, Lebbo (from east-central Borneo), and Juang (an Austroasiatic-speaking group from India), and the second including Dai, Semene, Barito, and Lao. The western Indonesians could be fit well with three (but not two) sources of ancestry: Austronesian-related, Austroasiatic-related, and Papuan-related (table S3), in respective proportions of ∼67%, 29%, and 4% (Semene); ∼37%, 60%, and 2% (Barito); and ∼55%, 23%, and 22% (Lebbo) (19). The Austroasiatic-associated component was inferred to be closer to...
and Native American points to admixture in Lao from a 4.5). Conversely, the signal of allele-sharing between Lao population affected by Han Chinese migrations, with a plausible explanation for our results being mixture between resident Austroasiatic speakers and incoming Tai speakers within historical times (5).

Our findings also have implications for genetic transformations linked to later cultural and linguistic shifts in Southeast Asia and beyond. We observe substantial genetic turnover between the Neolithic period and Bronze Age in Vietnam, likely reflecting a new influx of migrants from China (24). Late Neolithic/Bronze Age Myanmar individuals from Oakaie also do not possess an Austroasiatic genetic signature, in their case being closer to populations speaking Sino-Tibetan languages (including present-day Myanmar), pointing to an independent East Asian origin. Outside of Mainland Southeast Asia, we document admixture events involving Austroasiatic-related lineages in India (where Austroasiatic languages continue to be spoken) and in Borneo and Sumatra (where all languages today are Austroasian).

In the latter case, the shared ancestry with Nicobarese (in addition to separate Papuan-related and Austroasiatic-associated components) supports previous genetic results and archaeological hints of an early Austroasiatic-associated Neolithic expansion to western Indonedia (25, 26). Overall, Southeast Asia shares common themes with Europe, Oceania, and sub-Saharan Africa, where ancient DNA studies of farming expansions and language shifts have revealed similar instances of genetic turnover associated with archaeologically attested transitions in culture.

REFERENCES AND NOTES
Materials and methods are available as supplementary materials.


Materials and methods are available as supplementary materials.


The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505, 43–49 (2014). doi:10.1038/nature12886 Medline


ACKNOWLEDGMENTS

We thank Iosif Lazaridis, Vagheesh Narasimhan, Iñigo Olalde, and Nick Patterson for technical assistance; Nicole Adamski and Ann-Marie Lawson for aiding with lab work; and Minh Tran Thi, Rona Ikehara-Quebral, Miriam Stark, Michele Toomay Douglas, and Joyce White for help with archaeological samples. Funding: This work was supported by the French Ministry for Europe and Foreign Affairs (T.O.P.), Japan Society for the Promotion of Science (grant 16H02527; H.M.), Statutory City of Ostrava (grant 0924/2016/ŠaS; P.C.), Moravian-Silesian Region (grant 01211/2016/RRC; P.C.), Irish Research Council (grant GOIPG/2013/36; D.F.), Thailand Research Fund (grant MRG5980146; W.K.), University of Ostrava (IRP projects; P.F. and P.C.), Czech Ministry of Education, Youth and Sports (project OPVVV 16_019/0000759; P.F. and P.C.), National Science Foundation (HOMINID grant BCS-1032255; D.R.), National Institutes of Health (NIGMS grant GM100233; D.R.), an Allen Discovery Center of the Paul Allen Foundation (D.R.), and the Howard Hughes Medical Institute (D.R.).


Competing interests: The authors declare no competing interests.

Data and materials availability: The aligned sequences are available through the European Nucleotide Archive under accession number PRJEB24939. Genotype datasets used in analysis are available at https://reich.hms.harvard.edu/datasets. All the skeletons for which we newly report ancient DNA data are curated by coauthors of this paper, who affirm that the sampling of the skeleton was performed with appropriate permissions.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aat3188/DC1

Materials and Methods

Figs. S1 to S4
Tables S1 to S3
References (27–70)
Fig. 1. Overview of samples. (A) Locations and dates of ancient individuals. Overlapping positions are shifted slightly for visibility. (B) PCA with East and Southeast Asians. We projected the ancient samples onto axes computed using the present-day populations (with the exception of Mlabri, who were projected instead due to their large population-specific drift). Present-day colors indicate language family affiliation: green, Austroasiatic; blue, Austronesian; orange, Hmong-Mien; black, Sino-Tibetan; magenta, Tai-Kadai. Map data from http://www.freeworldmaps.net/asia/southeastasia/physical.html.
Fig. 2. Relative amounts of deeply diverged ancestry. The Y-axis shows $f_4(X, \text{Kinh}; \text{Australasian, Han})$ (multiplied by $10^4$) for populations $X$ listed on the x-axis (present-day as aggregate; ancient samples individually, except for points labeled “all”). Symbols are as in Fig. 1. Bars give two standard errors in each direction; dotted lines indicate the levels in Man Bac (top, blue) and Kinh (zero, black). B. C., Ban Chiang.
Fig. 3. Schematics of admixture graph results. (A) Wider phylogenetic context. (B) Details of the Austroasiatic clade. Branch lengths are not to scale, and the order of the two events on the Nicobarese lineage in (B) is not well determined (19).
Table 1. Sample information.

<table>
<thead>
<tr>
<th>ID</th>
<th>Lib.</th>
<th>Date (yBP)</th>
<th>Site</th>
<th>Country/period</th>
<th>Lat.</th>
<th>Long.</th>
<th>Sex</th>
<th>Mt Hap</th>
<th>Y Hap</th>
<th>Cov.</th>
</tr>
</thead>
<tbody>
<tr>
<td>VN22</td>
<td>6</td>
<td>3835–3695</td>
<td>Man Bac</td>
<td>Vietnam N</td>
<td>20.1</td>
<td>106.0</td>
<td>F</td>
<td>M13b</td>
<td>..</td>
<td>0.048</td>
</tr>
<tr>
<td>VN29</td>
<td>9</td>
<td>3900–3600</td>
<td>Man Bac</td>
<td>Vietnam N</td>
<td>20.1</td>
<td>106.0</td>
<td>F</td>
<td>M7b1a1</td>
<td>..</td>
<td>0.049</td>
</tr>
<tr>
<td>VN31</td>
<td>1</td>
<td>3900–3600</td>
<td>Man Bac</td>
<td>Vietnam N</td>
<td>20.1</td>
<td>106.0</td>
<td>M</td>
<td>No call</td>
<td>No call</td>
<td>0.005</td>
</tr>
<tr>
<td>VN33</td>
<td>2</td>
<td>3900–3600</td>
<td>Man Bac</td>
<td>Vietnam N</td>
<td>20.1</td>
<td>106.0</td>
<td>M</td>
<td>B5a1a</td>
<td>O2a</td>
<td>0.028</td>
</tr>
<tr>
<td>VN34</td>
<td>10</td>
<td>4080–3845</td>
<td>Man Bac</td>
<td>Vietnam N</td>
<td>20.1</td>
<td>106.0</td>
<td>F</td>
<td>M7b1a1</td>
<td>..</td>
<td>0.106</td>
</tr>
<tr>
<td>VN37</td>
<td>4</td>
<td>3825–3635</td>
<td>Man Bac</td>
<td>Vietnam N</td>
<td>20.1</td>
<td>106.0</td>
<td>M</td>
<td>M7b1a1</td>
<td>CT</td>
<td>0.019</td>
</tr>
<tr>
<td>VN39</td>
<td>11</td>
<td>3830–3695</td>
<td>Man Bac</td>
<td>Vietnam N</td>
<td>20.1</td>
<td>106.0</td>
<td>M</td>
<td>M7b1a1</td>
<td>O2a1c1b1a</td>
<td>0.102</td>
</tr>
<tr>
<td>VN40</td>
<td>6</td>
<td>3820–3615</td>
<td>Man Bac</td>
<td>Vietnam N</td>
<td>20.1</td>
<td>106.0</td>
<td>M</td>
<td>M74b</td>
<td>O1b1a</td>
<td>0.041</td>
</tr>
<tr>
<td>VN41</td>
<td>5</td>
<td>2100–1900</td>
<td>Nui Nap</td>
<td>Vietnam BA</td>
<td>19.8</td>
<td>105.8</td>
<td>F</td>
<td>C7a</td>
<td>..</td>
<td>0.373</td>
</tr>
<tr>
<td>VN42</td>
<td>6</td>
<td>1995–1900</td>
<td>Nui Nap</td>
<td>Vietnam BA</td>
<td>19.8</td>
<td>105.8</td>
<td>M</td>
<td>M8a2a</td>
<td>F</td>
<td>0.042</td>
</tr>
<tr>
<td>OA11/S28</td>
<td>20</td>
<td>3200–2700</td>
<td>Oakaie 1</td>
<td>Myanmar LN/BA</td>
<td>22.4</td>
<td>95.0</td>
<td>F</td>
<td>D4q</td>
<td>..</td>
<td>0.178</td>
</tr>
<tr>
<td>OA11/S29</td>
<td>4</td>
<td>3200–2700</td>
<td>Oakaie 1</td>
<td>Myanmar LN/BA</td>
<td>22.4</td>
<td>95.0</td>
<td>F</td>
<td>D4h1c</td>
<td>..</td>
<td>0.011</td>
</tr>
<tr>
<td>BCES B67</td>
<td>1</td>
<td>3500–3200</td>
<td>Ban Chiang</td>
<td>Thailand LN/BA</td>
<td>17.4</td>
<td>103.2</td>
<td>F</td>
<td>F1f</td>
<td>..</td>
<td>0.005</td>
</tr>
<tr>
<td>BCES B38</td>
<td>1</td>
<td>3200–3000</td>
<td>Ban Chiang</td>
<td>Thailand BA</td>
<td>17.4</td>
<td>103.2</td>
<td>F</td>
<td>B5a1a</td>
<td>..</td>
<td>0.017</td>
</tr>
<tr>
<td>BCES B54</td>
<td>1</td>
<td>3200–3000</td>
<td>Ban Chiang</td>
<td>Thailand BA</td>
<td>17.4</td>
<td>103.2</td>
<td>M</td>
<td>B5a1c</td>
<td>CT</td>
<td>0.010</td>
</tr>
<tr>
<td>BCES B27</td>
<td>7</td>
<td>3000–2800</td>
<td>Ban Chiang</td>
<td>Thailand BA</td>
<td>17.4</td>
<td>103.2</td>
<td>M</td>
<td>M74b2</td>
<td>..</td>
<td>0.030</td>
</tr>
<tr>
<td>BCES B16</td>
<td>1</td>
<td>2600–2400</td>
<td>Ban Chiang</td>
<td>Thailand IA</td>
<td>17.4</td>
<td>103.2</td>
<td>M</td>
<td>M72a</td>
<td>F</td>
<td>0.017</td>
</tr>
<tr>
<td>AB40</td>
<td>7</td>
<td>1890–1730</td>
<td>Vat Komnou</td>
<td>Cambodia IA</td>
<td>11.0</td>
<td>105.0</td>
<td>M</td>
<td>B5a1a</td>
<td>O</td>
<td>0.047</td>
</tr>
</tbody>
</table>

Calibrated radiocarbon dates are shown in bold (95.4% confidence interval, rounded to nearest 5 years); dates in plain text are estimated from archaeological context. Lib., number of sequencing libraries; Cov., average coverage level for 1.2 million genome-wide SNP targets; N, Neolithic; LN, Late Neolithic; BA, Bronze Age; IA, Iron Age.
Ancient genomes document multiple waves of migration in Southeast Asian prehistory

Mark Lipson, Olivia Cheronet, Swapan Mallick, Nadin Rohland, Marc Oxenham, Michael Pietrusewsky, Thomas Oliver Pryce, Anna Willie, Hirofumi Matsumura, Hallie Buckley, Kate Domett, Nguyen Giang Hai, Trinh Hoang Hiep, Aung Aung Kyaw, Tin Tin Win, Baptiste Pradier, Nasreen Broomandkhoshbacht, Francesca Candilio, Piya Changmai, Daniel Fernandes, Matthew Ferry, Beatriz Gamarra, Eadaoin Harney, Jatupol Kampuansai, Wibhu Kutanan, Megan Michel, Mario Novak, Jonas Oppenheimer, Kendra Sirak, Kristin Stewardson, Zhao Zhang, Pavel Flegontov, Ron Pinhasi and David Reich

described online May 17, 2018