Ancient genomes document multiple waves of migration in Southeast Asian prehistory

Mark Lipson1, Olivia Cheronet3,–5, Swapan Mallick3,–5, Nadin Rohland4, Marc Oxenham6, Michael Pietruensesky7, Thomas Oliver Pryce3,–9,–10, Anna Willis7, Hirofumi Matsunura2,–3, Hallie Buckley23, Kate Domett4,3, Giang Hai Nguyen15, Hoang Hiep Trinh13, Aung Aung Kyaw16, Tin Tin Win16, Baptiste Pradier9, Nasreen Broomandkhoshaib1,17, Francesca Candillio18,19, Piya Changmai20, Daniel Fernandez2,3,21, Matthew Ferry1,17, Beatriz Gamarra3,4, Eadaoin Harney1,17, Tatjup Kampanua22,3, Wibhu Kutanan22,3,4, Megan Michel17, Mario Novak3,23, Jonas Oppenheimer1,17, Kendra Sirak3,26, Kristin Stewardson1,17, Zhao Zhang1, Daniel Fernandes2,3,21, Matthew Ferry1,17, Beatriz Gamarra3,4, Eadaoin Harney1,17, Tatjup Kampanua22,3, Wibhu Kutanan22,3,4, Megan Michel17, Mario Novak3,23, Jonas Oppenheimer1,17, Kendra Sirak3,26, Kristin Stewardson1,17, Zhao Zhang1, Pavel Flegontov20,27, Ron Pinhasi2,3,9, David Reich1,4,17,‡

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 18 Southeast Asian individuals spanning from the Neolithic period through the Iron Age (4100 to 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 major 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 before the present (yr B.P.) and anatomically modern humans becoming widely established by 50,000 yr B.P. (1–3). Particularly profound changes in human culture were propelled by the spread of agriculture. Rice farming began in the region ~4500 to 4000 yr B.P. and was accompanied by a relatively uniform and widespread suite of tools and pottery styles displaying 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, 11–14). 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 and Fig. 1A): Man Bac (Vietnam, Neolithic; 4100 to 3600 yr B.P.), Nui Nap (Vietnam, Bronze Age; 2100 to 1900 yr B.P.), Oakaie 1 [Myanmar, Late Neolithic/Bronze Age; 3200 to 2700 yr B.P. (15)], Ban Chiang [Thailand, Late Neolithic through Iron Age; 3500 to 2400 yr B.P. (16),] and Vat Komnou [Cambodia, Iron Age; 1900 to 1700 yr B.P. (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 165 distinct individuals. For libraries with evidence of authentic ancient DNA, we generated wide data using in-solution enrichment, yielding sequences from 18 individuals (Table 1 and 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 dataset).

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 Asians, Australasians, Central Americans, and Europeans), 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 Khin), 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, whereas 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 speakers. Present-day Lao are intermediate between Austroasiatic speakers and Dai, and western Indonesians (Semende 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 f3-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 Khin; and Oakaie with Sino-Tibetan-speaking groups. We...
Man Bac individuals are additionally mostly similar to each other, except for one, VN29, which is significantly higher than the population mean [Bonferroni-corrected Z-test, \( p < 0.02 \) (19)]. Vat Kombou and Ban Chiang also yield high positive values, while Oakaie is modestly positive, and Nui Nap is close to zero (\( Z = 1.1 \)).

Next, we built admixture graph models to test the relationships between the Vietnam Neolithic samples and present-day Southeast Asians in a phylogenetic framework. We began with a scaffold model containing the Upper Paleolithic Siberian Ust'-Ishim individual (27) 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); \( f_2 \)-statistics indicate that this source is also not closely related to Papuans, South Asians, or the 40,000 yr B.P. Tianyuan individual (22); table S3]. The allele sharing demonstrated by outgroup \( f_2 \)-statistics can be accommodated along the farmer lineage, projecting instead owing to their large population-specific drift. Present-day colors indicate language family affiliation: green, Austroasiatic; blue, Austro-Asian; orange, Hmong-Mien; black, Sino-Tibetan; magenta, Tai-Kadai. Map data from www.freeworldmaps.net/asia/southeastasia/physical.html.

**Table 1. Sample information.** 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 (single-nucleotide polymorphism) targets; N, Neolithic; LN, Late Neolithic; BA, Bronze Age; IA, Iron Age.

<table>
<thead>
<tr>
<th>ID</th>
<th>Lib.</th>
<th>Date (yr B.P.)</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</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</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</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</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</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</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</td>
<td>20.1</td>
<td>106.0</td>
<td>M</td>
<td>M7b1a1</td>
<td>O2aC1b1a</td>
<td>0.102</td>
</tr>
<tr>
<td>VN40</td>
<td>6</td>
<td>3820–3615</td>
<td>Man Bac</td>
<td>Vietnam</td>
<td>20.1</td>
<td>106.0</td>
<td>M</td>
<td>M7b4b</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>OAIL/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>OAIL/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>F</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 Kombou</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>

**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

Lipson et al., Science 361, 92–95 (2018) 6 July 2018
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 and 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 and S4), with the first including Dai, Semende, Barito, Lebbo (from east-central Borneo), and Juang (an Austroasiatic-speaking group from India), and the second including Dai, Semende, Barito, and Lao. The western Indonesians could be fit well with three (but not two) sources of ancestry:

Astronomer-related, Austroasiatic-related, and Papuan-related (table S3), in respective proportions of ~67, 29, and 4% (Semende); ~37, 60, and 2% (Barito); and ~55, 23, and 22% (Lebbo) (19). The Austroasiatic-associated component was inferred to be closer to Nicobarese than to Mlabri or Man Bac, forming a "southern" Austroasiatic subclade (fig. 3B). For Juang, we also obtained a good fit with three ancestry components: one western Eurasian, one deep eastern Eurasian (interpreted as an indigenous South Asian lineage), and one from the Austroasiatic clade (fig. S3). The Austroasiatic source for Juang (proportion 35%) was inferred to be closer to Mlabri, as supported by statistics $f_4(Juang, Palliyar; Mlabri, X) > 0$ for $X = Atayal, Man Bac, or Nicobarese ($Z = 5.1, 2.8, 2.3$), creating a "northern" Austroasiatic subclade. Separately, we found that Lao also possesses ancestry from the Austroasiatic clade (47%; fig. S4) but cannot be fit as a simple mixture of the same two components found in Nicobarese and Mlabri (residual statistic $Z = 3.4$ without a source to explain distantly shared ancestry between Lao and Mixe) (19).

Our results provide genetic support for the hypothesis that agriculture was first practiced in mainland Southeast Asia by (proto-) Austroasiatic-speaking migrants from southern China (4–6, 11–13). We find that all eight of our sampled individuals from Man Bac (as well as individuals from Ban Chiang and Vat Komnou) are closely related to present-day Austroasiatic speakers, including a shared pattern of admixture, with one, VN29, exhibiting significantly elevated indigenous ancestry. By comparison, studies of cranial and dental morphology have placed Man Bac either close to present-day East and Southeast Asians (“Neolithic”), intermediate between East Asians and a cluster containing more ancient hunter-gatherers from the region plus present-day Onge and Papuan (“indigenous”), or split between the two clusters (7, 8, 23). The simplest explanation for our results is that the majority of our Man Bac samples represent a homogeneous Neolithic cluster, with recent local contact between farmers and hunter-gatherers leading to additional hunter-gatherer ancestry in VN29 and perhaps VN40 (7, 8). This model would imply that the incoming farmers had already acquired 25 to 30% hunter-gatherer ancestry, either in China or Southeast Asia, establishing the characteristic Austroasiatic-affiliated genetic profile seen in multiple populations today. The wide distribution of this profile across Southeast Asia (in some cases in admixed form) also supports a coherent migration with early shared admixture. The symmetric position of aboriginal Taiwanese and the majority East Asian ancestral lineage in Man Bac (and Austroasiatic speakers) with respect to Native Americans points to admixture in Lao from a mixture between resident Austroasiatic speaking (to an independent East Asian origin. Outside of mainland Southeast Asia, we document admixture events involving Austroasiatic-related lineages in India (where Austroasiatic languages

**Fig. 2. Relative amounts of deeply diverged ancestry**. The y axis shows $f_4(X, Kinh; Australasian, Han)$ (multiplied by $10^3$) 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 Kin (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).
continue to be spoken) and in Borneo and Sumatra (where all languages today are Austronesian). In the latter case, the shared ancestry with Nicobarese (in addition to separate Papuan-related and Austronesian-associated components) supports previous genetic results and archaeological hints of an early Austronesian-associated Neolithic expansion to western Indonesia (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

19. Materials and methods are available as supplementary materials.

ACKNOWLEDGMENTS


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/content/361/6397/92/suppl/DC1 Materials and Methods Figs. S1 to S4 Tables S1 to S3 References (27–70) 13 February 2018; accepted 3 May 2018 Published online 17 May 2018 10.1126/science.aat3188
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 Willis, Hirotumi Matsumura, Hallie Buckley, Kate Domett, Giang Hai Nguyen, Hoang Hiep Trinh, Aung Aung Kyaw, 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

Science 361 (6397), 92-95.
DOI: 10.1126/science.aat3188 originally published online May 17, 2018

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