


Unveiling the Genetic History of the Maniq, a Primary Hunter-Gatherer Society

Tobias Göllner^{1,2*}, Maximilian Larena^{3,*}, Wibhu Kutanan ⁴, Helmut Lukas⁵, Martin Fieder^{1,2}, and Helmut Schaschl¹

¹Department of Evolutionary Anthropology, Faculty of Life Sciences, University of Vienna, Austria

²HEAS - Network of Human Evolution and Archeological Sciences, University of Vienna, Vienna, Austria

³Human Evolution, Department of Organismal Biology, Uppsala University, Sweden

⁴Department of Biology, Faculty of Science, Khon Kaen University, Thailand

⁵Institute for Social Anthropology, Austrian Academy of Sciences, Vienna, Austria

*Corresponding authors: E-mails: tobias.goellner@univie.ac.at; maximilian.larena@ebc.uu.se.

Accepted: February 1, 2022

Abstract

The Maniq of southern Thailand is one of the last remaining practicing hunter-gatherer communities in the world. However, our knowledge on their genetic origins and demographic history is still largely limited. We present here the genotype data covering ~2.3 million single nucleotide polymorphisms of 11 unrelated Maniq individuals. Our analyses reveal the Maniq to be closely related to the Semang populations of Malaysia (Malay Negritos), who altogether carry an Andamanese-related ancestry linked to the ancient Hòabìnhiàn hunter-gatherers of Mainland Southeast Asia (MSEA). Moreover, the Maniq possess ~35% East Asian-related ancestry, likely brought about by recent admixture with surrounding agriculturist communities in the region. In addition, the Maniq exhibit one of the highest levels of genetic differentiation found among living human populations, indicative of their small population size and historical practice of endogamy. Similar to other hunter-gatherer populations of MSEA, we also find the Maniq to possess low levels of Neanderthal ancestry and undetectable levels of Denisovan ancestry. Altogether, we reveal the Maniq to be a Semang group that experienced intense genetic drift and exhibits signs of ancient Hòabìnhiàn ancestry.

Key words: Southeast Asia, Semang, Negritos, population genetics, genetic ancestry.

Significance

Our knowledge on the genetic origins and identity of the Maniq, a hunter-gatherer society of Mainland Southeast Asia, is still largely limited. We generated and comprehensively analyzed the genome-wide autosomal data of the Maniq, and demonstrate that the Maniq exhibit high levels of genetic drift likely due to long periods of geographic isolation, share an Andamanese-related ancestry with the Semang hunter-gatherer populations of Malaysia, are related to the ancient Hòabìnhiàn individuals of MSEA, possess ~35% East Asian-related ancestry brought about by interbreeding with recent migrants, and have undetectable levels of Denisovan ancestry. Altogether, our findings reveal for the first time, the genetic origins, demographic history, and the genetic relationships of the Maniq relative to other populations in the region.

© The Author(s) 2022. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Introduction

The Maniq (also known as Kensiu) are a society of ~250 hunter-gatherers (Eberhard et al. 2019) who reside in the remaining arboreous hills of southern Thailand (Kricheff and Lukas 2015). They are culturally classified as one of the “Semang” groups of Mainland Southeast Asia (MSEA), which include the Orang Semang indigenous cultural communities of Peninsular Malaysia, such as the Batek (“Bateq”), Jehai (“Jahai”), Kintaq, and Mendriq. The Semang are loosely labeled together as the “Negritos,” grouped with the other indigenous cultural communities of the region including the Andamanese (e.g., Onge or Jarawa) and the Philippine Negritos (e.g., Ayta, Agta, or Mamanwa) (Barrows 1910; Radcliffe-Brown 2013). This classification is mainly based on the early anthropological descriptions of the characteristic Negrito phenotype of these populations: on average short stature, frizzier hair, and darker skin color (Barrows 1910; Evans 1968; Radcliffe-Brown 2013).

The peopling of MSEA has been a topic of interest for many anthropologists and historians for more than a century (Skeat et al. 1907). Based on the phenotypic and ethnographic descriptions of indigenous groups in the area, one of the earliest hypotheses posited three sequential waves of human migration into Peninsular Malaysia (Schebesta 1927), starting with the hunter-gatherer Semang as the first people, followed by the Senoi who mainly practice slash and burn agriculture, and then by the Proto-Malay who are largely regarded as the settled agriculturists. This hypothesis was partially based on the current location where these populations reside, with the Semang in the north, followed by the Senoi in the middle of the Peninsular Malaysia, and lastly the Proto-Malay down in the south. Alternatively, a Two-Layer hypothesis was also proposed, which is mostly based on morphometric and dental analyses of human remains from various archaeological sites in MSEA (Matsumura et al. 2008; Reich et al. 2011; Jinam et al. 2017). This archaeological theory suggested a two-layered migration of culturally and phenotypically distinct groups of modern humans across different periods, first by Palaeolithic Australo-Melanesian hunter-gatherers followed by Neolithic East Asian agriculturists (Matsumura et al. 2019). The former is associated with the pre-Neolithic Hòabinhian cultural assemblage, whereas the latter is associated with the spread of cereal agriculture by farmer communities from southern China into MSEA.

For the past decade or so, a series of genomic studies shed some light on the peopling of MSEA. A large study covering 73 present-day Asian populations pointed towards one single wave of migration into Asia (HUGO Pan-Asian SNP Consortium et al. 2009). A subsequent study with a more comprehensive coverage of indigenous Malaysian populations revealed at least three waves of migration into the Malay peninsula, with the ancestors of the genetically distinct Semang groups depicted as the first modern human migrants

of the region (Aghakhanian et al. 2015). More complexity was revealed following two related studies that utilized multiple ancient DNA samples from different sites (Lipson et al. 2018; McColl et al. 2018). Altogether, both studies inferred at least four major waves of migration into Southeast Asia: the Hòabinhian, Austroasiatic, Austronesian, and another East Asian-related group.

The Maniq may add interesting insights on the origins and demographic history of early modern humans in MSEA. They have a unique position among the Semang, given their long history of geographical isolation from the other Semang groups. This fact, which has been noted in an earlier study (Brandt 1961), is even more pronounced in recent times, provided the existence of a political border that separates the Maniq of Thailand from the other Semang groups of Malaysia. The information on the genetic identity of the Maniq and their genetic relationships to other Semang and non-Semang groups of Malaysia is still largely limited to an analysis of uniparental markers (Kutanan et al. 2018). To advance our understanding, we present here a genome-wide autosomal data for 11 unrelated individuals of Maniq from Trang, Phatthalung, and Satun provinces of Thailand (fig. 1A). The samples were genotyped for ~2.3 million single nucleotide polymorphisms (SNPs), and were comprehensively analyzed together with the available genome-wide data for the present-day populations and ancient individuals of the region (supplementary table 1A and B, Supplementary Material online).

Results

Genetic Affiliation of the Maniq

To determine the population genetic relationships of the Maniq, we performed a PCA on populations of Africa, Europe, East Asia, and Oceania (fig. 1B). The Maniq individuals form a cluster close to the Malay Semang, on a cline between the Andamanese and East Asian population clusters, indicating an admixture between Andamanese-related and East Asian-related ancestral sources. We then performed a PCA restricted to Asian and Oceanian populations (fig. 1C). Eigenvector 1 spans from East Asian to Australasian populations, whereas the second Eigenvector is defined by the Oceanian versus Maniq populations. The Maniq forms a separate cluster at the polar opposite of East Asians, where Malay Semang and non-Semang populations lie in between in a cline.

We performed a clustering approach implemented in ADMIXTURE to identify the population structure of Maniq in relation to other Asian and Oceanian populations (fig. 2). The earliest K ($K = 2$) is defined by the East Asian-related versus Australasian-related (or Negrito-Australopapuan) clusters, where Maniq appears to be largely affiliated with the latter. Notably, as early as $K = 3$, Maniq already forms its own

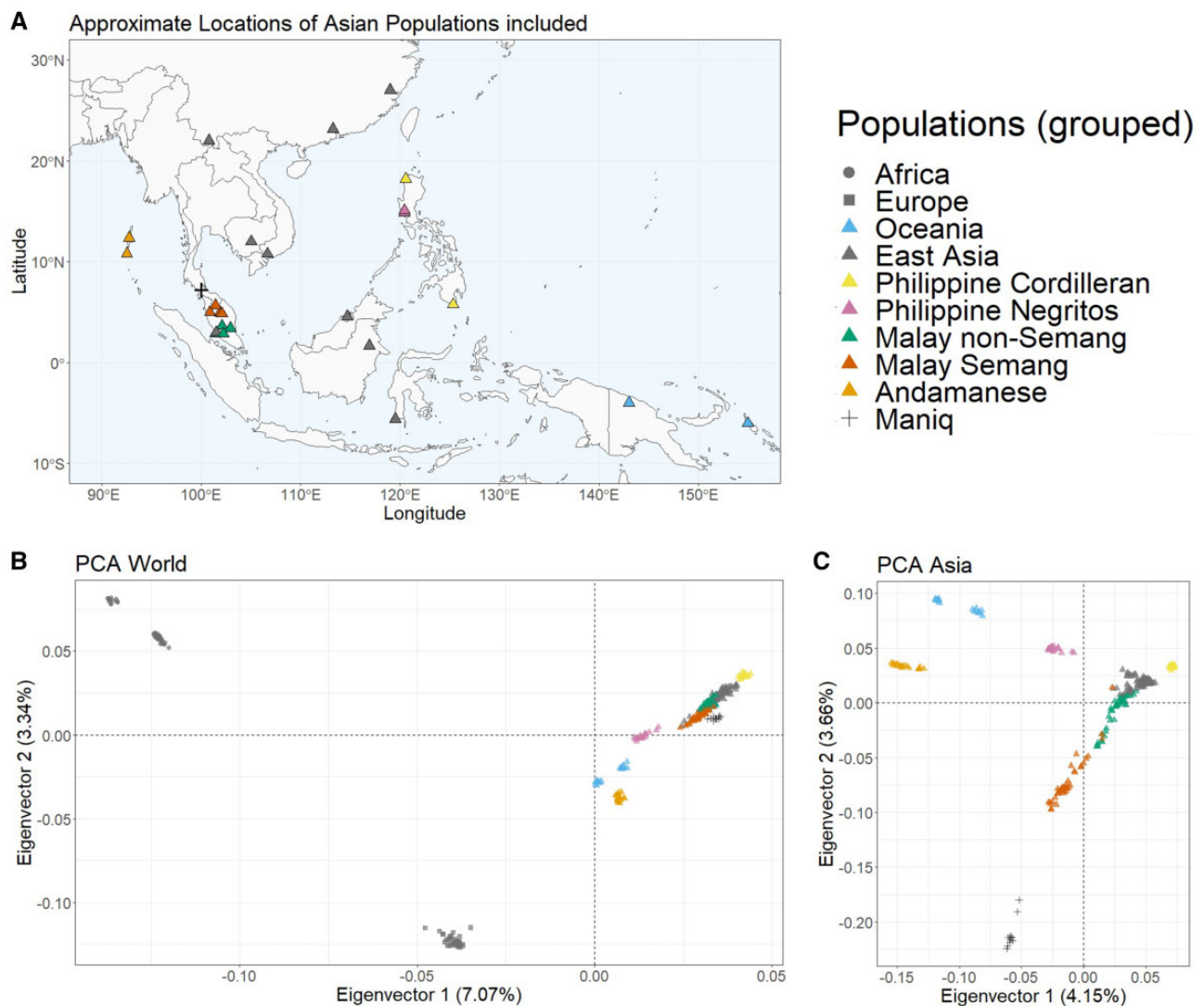


Fig. 1.—Maniq form a genetic cluster close to Malay Semang populations. (A) Map with approximate locations of all Asian populations included in the study. (B) Eigenvector 1 and 2 of a PCA on a worldwide set of populations, number of SNPs used = 320,032. (C) Eigenvector 1 and 2 of a PCA restricted to Asian populations, number of SNPs used = 319,845.

distinct cluster, separating them from Onge, Jarawa, Papuan, and Bougainville Islanders. The Maniq largely and consistently retained its own distinct cluster up to the optimal $K = 7$, the K with the lowest cross-validation error and high consistency of iterated runs (50 consistent replicates out of 50 runs). The lack of admixture among the Maniq in this structure-based analysis may be attributed to the high degree of genetic drift as discussed below.

Maniq Exhibit High Degree of Genetic Differentiation

To determine the degree of genetic differentiation in Maniq in relation to other populations in the Asia-Pacific region, we estimated the in between population F_{ST} of our data set panel (supplementary figs. 1–3, Supplementary Material online).

We find that the Maniq have experienced intense genetic drift, given that they exhibit very high genetic distance based on pairwise F_{ST} when compared with any other reference population. The degree of genetic drift in Maniq is even higher than that of Andamanese or Philippine Negritos, who were previously shown as drifted populations (Mondal et al. 2016; Larena, Sanchez-Quinto, et al. 2021). Moreover, the degree of genetic drift is comparable to that of Surui of Brazil, even higher than the Mangyan Buhid of Philippines (supplementary fig. 3, Supplementary Material online), all of whom were previously demonstrated to have one of the highest levels of genetic drift relative to a worldwide set of populations (Mallick et al. 2016; Larena, Sanchez-Quinto, et al. 2021). The evidence on the extreme genetic drift observed in Maniq is also supported by our analysis on the ROH. The

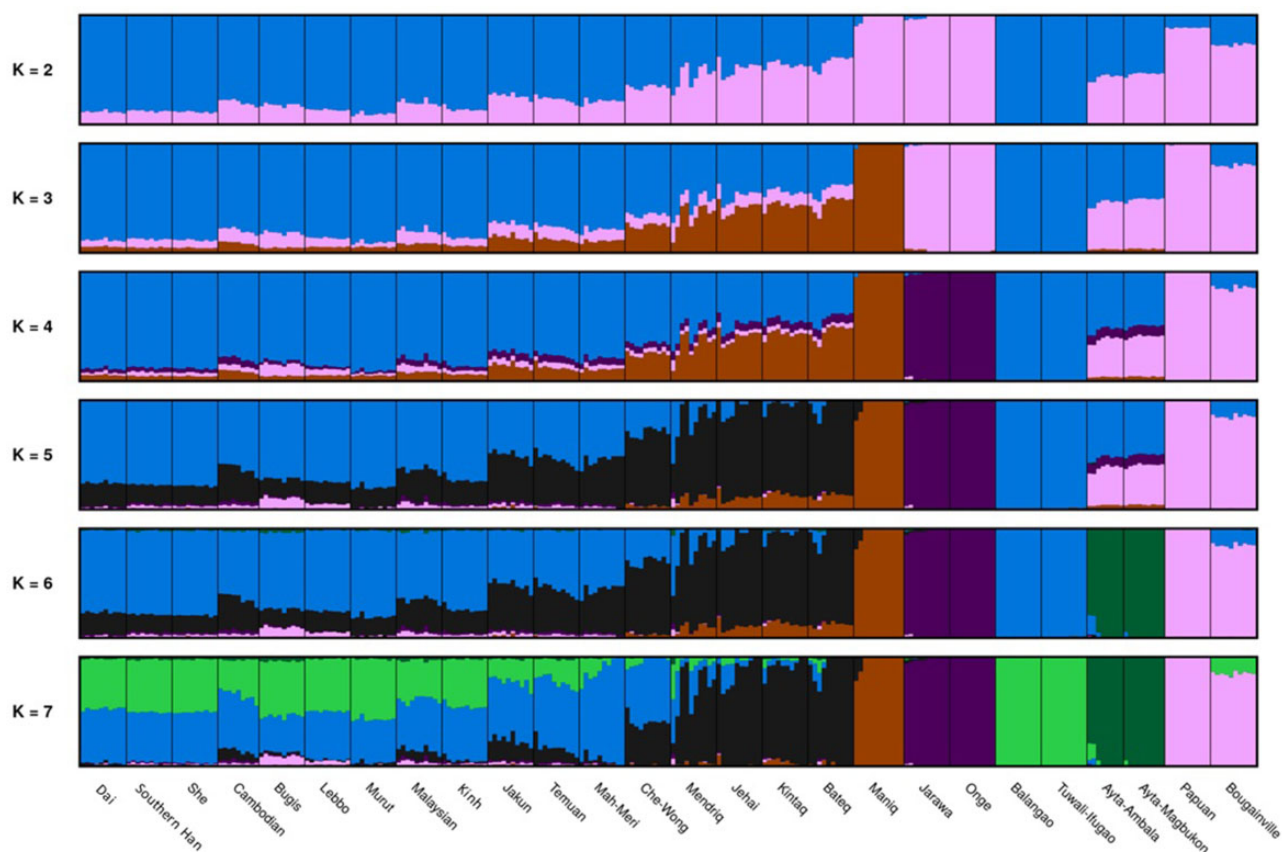


Fig. 2.—Admixture analysis for Asian populations. admixture plots for $K=2-7$ with 170,513 SNPs used ($K=7$ has the lowest cross-validation error).

length and the number of ROH can provide insights on a population's demographic history. High number of long tract ROH can be indicative of a population bottleneck or largely inbred population of small size (Ceballos et al. 2018). Except for Africans, all other populations including the Maniq displayed high number of short segment ROH, which was argued as an evidence for a shared out-of-Africa bottleneck (Kirin et al. 2010). Interestingly, compared with all other populations, the Maniq consistently exhibit high numbers of ROH in each segment length category of more than 1 Mb (fig. 3B). Moreover, when plotting the total number versus the total length of ROH segments (fig. 3A), the Maniq exhibit high values for both, a pattern that was also observed in other highly drifted populations such as the Surui of the Americas and the Mangyan Buhid of the Philippines (supplementary fig. 4A and B, Supplementary Material online). Accordingly, these aforementioned populations as well as the Maniq possess similar population characteristics: small population size and historical evidence of endogamy (Brandt 1961; Mallick et al. 2016; Larena, Sanchez-Quinto, et al. 2021). Thus, it is not surprising to observe that these highly drifted populations also demonstrate to possess the highest values of inbreeding coefficient (supplementary table 2, Supplementary Material online).

Maniq Are One of the Hòabìnhan-Related Populations of MSEA

It was previously argued that the ancient hunter-gatherers of the Hòabìnhan cultural complex are regarded as the ancestors of present-day hunter-gatherers of MSEA (McCull et al. 2018). This was recently supported by an ancient DNA data, where among the ancient individuals, Ornge and Jarawa were shown to be most genetically affiliated with an 8,000-year-old Hòabìnhan individual from Pha Faen, Laos and a 4,000-year-old Hòabìnhan individual from Gua Cha, Malaysia (McCull et al. 2018). Using the test $f_3(\text{Mbuti}; \text{Maniq}, \text{Ancient})$, we also find the Maniq to share high levels of genetic drift with the Lao Hòabìnhan, besides the ~8,000-year-old Liangdao 2 individual and the Malay and Lao Neolithic individuals (supplementary fig. 5 and supplementary table 3, Supplementary Material online), indicating a combination of Hòabìnhan-related and East Asian-related ancestries for the Maniq. We then formally tested the presence of Hòabìnhan-related ancestry in the Maniq, other Semang, and Andamanese populations with $f_4(\text{Mbuti}, \text{Lao Hòabìnhan}; \text{Liangdao 2}, X)$ and $f_4(\text{Mbuti}, \text{Lao Hòabìnhan}; \text{Balangao}, X)$, where we examine which X population share alleles with the Lao Hòabìnhan relative to a least admixed East Asian individual/population, Liangdao 2 or Cordilleran Balangao (Larena,

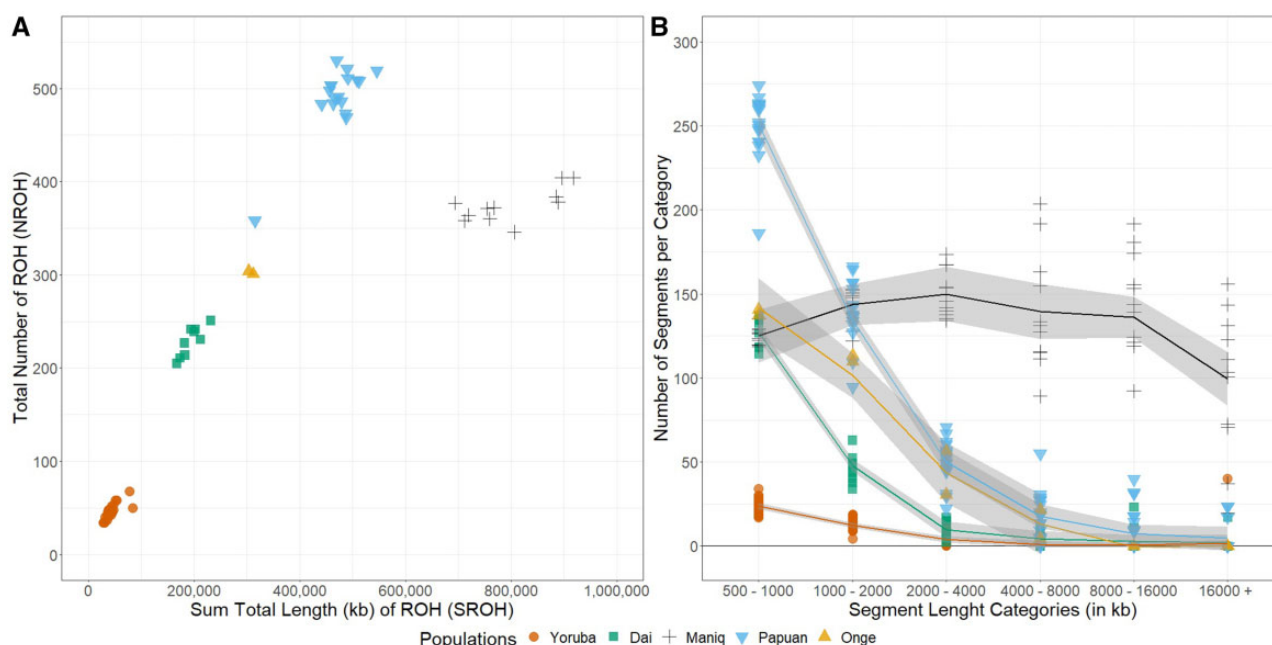


Fig. 3.—High levels of genetic differentiation in the Maniq. (A) Total length of ROH versus total number of ROH per individual and (B) Number of ROH per segment length category covering the Yoruba, Dai, Maniq, Papuan, and Onge populations (for additional populations see [supplementary figs. 4A and B, Supplementary Material online](#)). Number of SNPs used = 779,956.

Sanchez-Quinto, et al. 2021). As expected and consistent with previous observations (McCull et al. 2018), Onge and Jehai displayed significant levels of Hòabìnhiàn-related ancestry. Looking at the remaining Semang populations, the Maniq displayed also high levels of Hòabìnhiàn-related ancestry (fig. 4A and B).

Examining the genetic relationships of present-day populations, our PCA and Admixture analysis indicate that the Maniq are affiliated to the Semang populations of Malaysia. We investigated this further with an outgroup f_3 statistics, using the test $f_3(\text{Mbuti}; \text{Maniq}, X)$, where X is any other population in the Asia-Pacific region. Evidently, Maniq shares the most drift with the Malay Semang, followed by the Senoi ([supplementary fig. 6, Supplementary Material online](#)), suggesting a shared history among these populations. These findings are also consistent when we test $f_4(\text{Mbuti}, \text{Maniq}; X, \text{Malay Semang/Andamanese})$, where the Maniq consistently forms a clade with the Malay Semang, relative to any other reference Asia-Pacific population ([supplementary fig. 7A–F, Supplementary Material online](#)). Moreover, using the tests $f_4(\text{Mbuti}, \text{Papuan}; \text{Balangao}, X)$ and $f_4(\text{Mbuti}, \text{Onge}; \text{Balangao}, X)$, we find the Maniq and Malay Semang populations to share alleles with Papuans and Andamanese, highlighting the shared Basal Australasian ancestry and the deep historical relationships among these populations (fig. 4E and F). Last, we find the Maniq to share more alleles with Andamanese relative to any other Malay Semang population ([supplementary fig. 8, Supplementary Material online](#)).

Maniq Display Evidence for Recent Admixture with East Asians

East Asian gene flow was previously observed among Malay Semang of MSEA (Aghakhanian et al. 2015). Ancient East Asian gene flow, represented by the Liangdao 2 sample, is present in Malay Semang, but to a lower extent in Maniq (fig. 4C). To formally investigate present-day gene flow in the Maniq, we performed the test $f_4(\text{Mbuti}, \text{Balangao}; \text{Onge}, X)$, where we examine whether any X of Asia-Pacific populations have gene flow with East Asians, represented by Balangao Cordillerans (fig. 4D). We find that Maniq, together with Malay Semang, exhibit significant gene flow with East Asians. This is supported by the implementation of f -statistical estimations in qpAdm (Harney et al. 2021), where we find that the most plausible model for the ancestral source of Maniq is a combination of both Andamanese and East Asian-related ancestries, 65% and 35%, respectively (fig. 5 and [supplementary table 5, Supplementary Material online](#)). Furthermore, we estimated the date of admixture using a weighted LD statistic-based method, MALDER (Loh et al. 2013), and revealed that Maniq had an admixture event between East Asian- and Andamanese-related ancestries which is dated to 709 years ago (95% CI: 475–944 years, assumed generation time of 25 years; [supplementary table 6, Supplementary Material online](#)).

To gain understanding on the phylogenetic relationships between Maniq and other MSEA populations, we implemented TreeMix which utilizes allele frequency data together

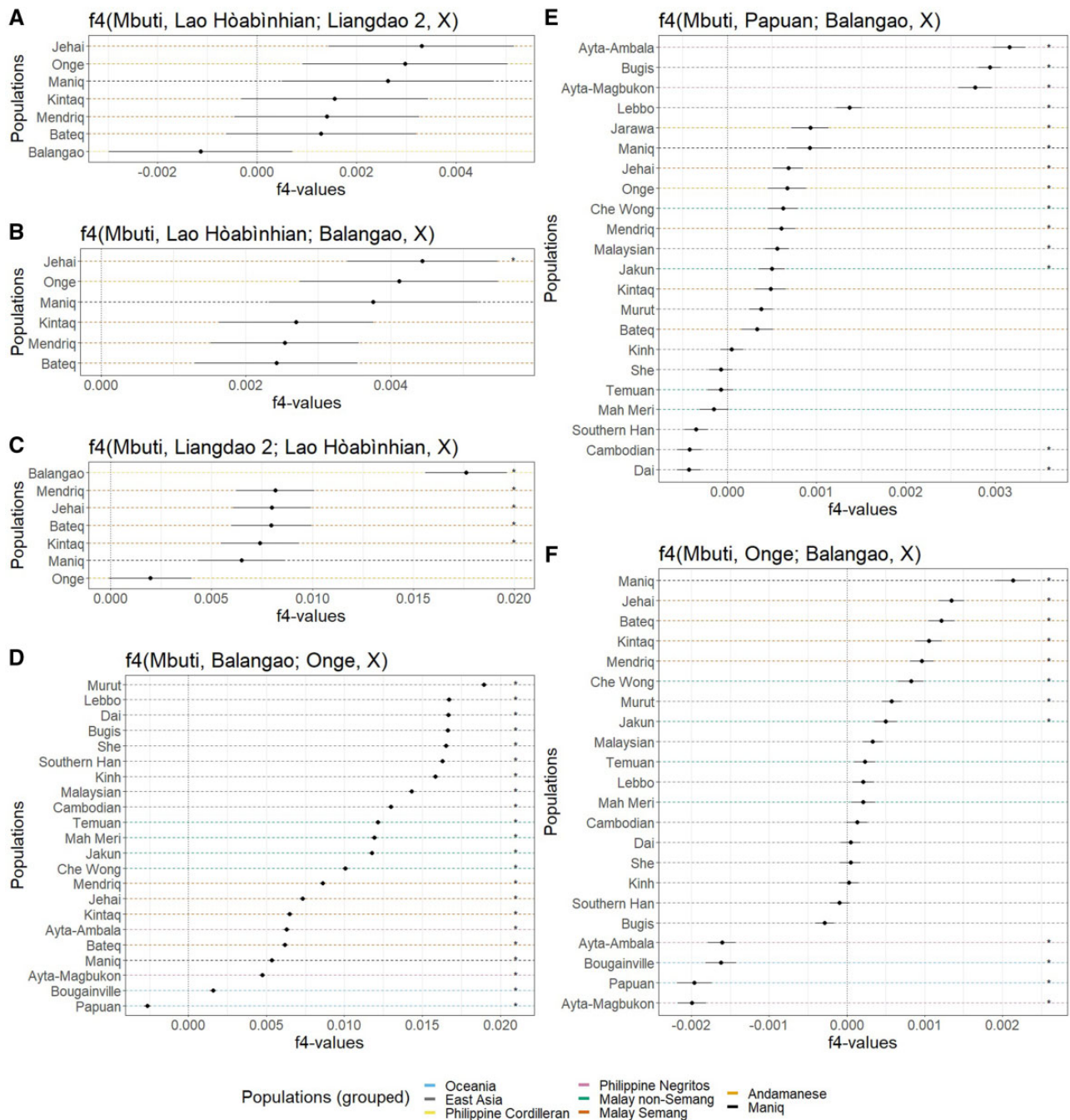


Fig. 4.—Different ancestry components of the Maniq. Results of f_4 statistics including standard errors for (A) $f_4(\text{Mbuti, Lao Hòabinhian; Liangdao 2, X})$, (B) $f_4(\text{Mbuti, Lao Hòabinhian; Balangao, X})$, (C) $f_4(\text{Mbuti, Liangdao 2; Lao Hòabinhian, X})$, and (D) $f_4(\text{Mbuti, Balangao; Onge, X})$, (E) $f_4(\text{Mbuti, Papuan; Balangao, X})$ and (F) $f_4(\text{Mbuti, Onge; Balangao, X})$ sorted from highest to lowest f_4 values. A line marked with an asterisk indicates an absolute Z value of greater than 3. Number of SNPs used for (A–C) = 16,428, for (D–F) = 319,386. Please see [supplementary table 4, Supplementary Material online](#), for the exact results including standard errors of (D).

with a Gaussian approximation for genetic drift (Pickrell and Pritchard 2012). Our analyses consistently reveal the significant genetic drift observed in Maniq relative to other populations, as well as the presence of Andamanese gene flow into the common ancestral branch of Semang groups (fig. 6).

Additionally, using qpGraph, we explicitly tested the possible models of admixture history for the Maniq (fig. 7). A model, where Maniq is exclusively a clade of Hòabinhian or East Asian-related groups were rejected (worst fitting Z of 8.3743 and 16.0049, respectively (see [supplementary figs. 9](#)

and 10, [Supplementary Material online](#)). A model that fits the data is when Maniq forms a clade with Onge and Lao Hòabinhian, and later received admixture with East Asian-related populations (worst fitting $Z = 0.788$), providing additional support for the dual ancestral source for the Maniq.

No Evidence for Elevated Levels of Denisovan Ancestry among Maniq

Using the direct f_4 -ratio estimation of Neanderthal ancestry: $f_4(\text{Neanderthal(Altai), Chimp; X, Yoruba})/f_4(\text{Neanderthal(Altai), Chimp; Neanderthal(Vindija), Yoruba})$, we find that the Maniq to possess $\sim 1.9\%$ Neanderthal ancestry, which is expected for a population outside of Africa, ([supplementary table 7](#) and [supplementary fig. 11, Supplementary Material online](#)). Likewise, the results are also consistent when we use the test $f_4(\text{Chimp, Neanderthal; Mbuti, X})$ ([supplementary table 8, Supplementary Material online](#)). Using $f_4(\text{Chimp, Denisovan; European, X})$ and the f_4 -ratio test $f_4(\text{Chimp, Neanderthal; Southern Han, X})/f_4(\text{Chimp, Neanderthal; Southern Han, Denisovan})$, we find no detectable signal of Denisovan ancestry among the Maniq, as well as other Malay Semang populations (Z scores < 1 or even negative). This holds also true for the Malay non-Semang, Andamanese, Indonesian, Cordilleran, and mainland East Asian populations ([supplementary tables 9 and 10, Supplementary Material online](#)). This is in contrast to the high levels of Denisovan ancestry found among Philippine Negritos and Australopapuans ([Jinam et al. 2017](#); [Jacobs et al. 2019](#)), indicating that the Denisovan introgression event among these groups likely occurred in Island Southeast Asia and the Oceania ([Jacobs et al. 2019](#); [Choin et al. 2021](#); [Larena, Sanchez-Quinto, et al. 2021](#)).

Discussion

We present in this study the first comprehensive investigation on the genetic origins and ancestry of the Maniq of Thailand, who are considered as one of the last remaining primary hunter-gatherers in the region ([Kelly 2013](#)). Our genome-wide analyses reveal that the Maniq population form a clade with the Semang groups of MSEA, who altogether carry high levels of Andamanese-related genetic ancestry. Among the Semang, the Maniq cluster more closely to the Bateq and the Kintaq, than to Jehai and Mendriq ([fig. 6](#) and [supplementary fig. 12, Supplementary Material online](#)). This is in line with the earlier anthropological descriptions of the Maniq ([Brandt 1961](#)), where they were regarded to share common cultural features with the Semang groups of Peninsular Malaysia, indicating a shared history and origins. Both the Maniq and the Malay Semang do trade their goods with other non-Semang groups of MSEA. In addition, they also speak a language that is altogether classified under the Aslian branch of the Austroasiatic language family. In contrast to the Malay Semang, the Maniq do not subsist on agriculture or practice

any animal husbandry, they solely rely on hunting and gathering as their mode of subsistence ([Lukas 2004](#)). Of course, this is only true for those Maniq that still live according to their traditional lifestyle and did not transition to a sedentary way of life.

The foraging form of subsistence can be traced back to the earliest “first layer” migrants of MSEA, the ancient Hòabinhian hunter-gatherers. Hence, like the Malay Semang, we find the Maniq to possess strong genetic affiliation with the 8,000-year-old Lao Hòabinhian individual. Historically, populations with high levels of Hòabinhian-related genetic ancestry were more widespread in East Asia; aside from Laos, they are also found in southern China ([Wang et al. 2021](#)) and as far as the Japanese archipelago ([McColl et al. 2018](#)). Due to the recent expansion of East Asian-related groups, the Hòabinhian-related cultural communities were either displaced, replaced, or absorbed into the larger population of farmer migrants. Though this is not the case with the Maniq, who were remained to be largely isolated and retained to be hunter gatherers, making them one of the few groups in mainland Asia carrying high levels of Hòabinhian-related ancestry.

Like other Semang groups of Peninsular Malaysia, the Maniq also exhibit admixture with populations bearing East Asian-related ancestry. This may be attributed to the recent southward expansion of Austroasiatic speakers via the Mekong River and/or the more recent northward expansion of Austronesian speakers from Malay Peninsula into southern Thailand. The latter may, in part, explain some Austronesian lexical items found in the Maniq language. Alternatively, this could also be attributed to the expansion of Tai-Kadai speakers into Southeast Asia around 1,000–2,000 years ago ([O'Connor 1995](#); [Pittayaporn 2014](#)). We additionally cannot entirely exclude a substrate of deep East-ancient related ancestry in the Maniq, which was recently detected in a $\sim 7,300$ -year-old Leang Panninge individual of Sulawesi ([Carlhoff et al. 2021](#)). Hence, the complex series of Holocene migrations into MSEA likely had a profound impact on the variable levels of East Asian-related ancestry among the Semang populations ([figs. 4 and 5](#)).

Interestingly, despite the presence of East Asian-related admixture, the Maniq consistently exhibit the highest amount of Andamanese-related ancestry in MSEA, levels that are higher than any other present-day Semang populations in the region. This implies that the impact of East Asian admixture in the Maniq is more limited relative to other Semang groups, likely attributed to their long periods of geographical and cultural isolation. Accordingly, our findings fit with the narrative of Maniq demographic history as a hunter-gatherer Hòabinhian-related population who arrived in MSEA and remained largely distinct, and who later received limited admixture with neighboring populations carrying East Asian ancestry.

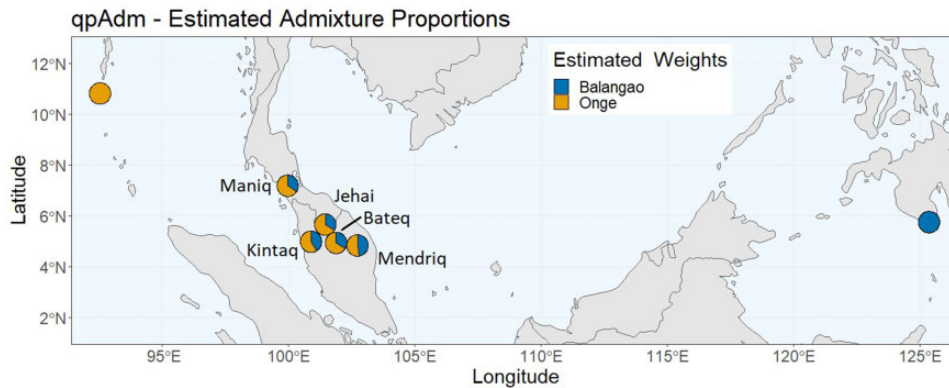


Fig. 5.—Estimated admixture proportions of Andamanese-related and East Asian-related ancestries of the Semang Populations of MSEA. Pie chart results of qpAdm with 319,386 SNPs used. The position of the pie charts corresponds to the approximate location of their respective populations (Aghakhanian et al. 2015). Pie chart positions were slightly adjusted to prevent overlapping, notably the Mendriq were shifted to the right.

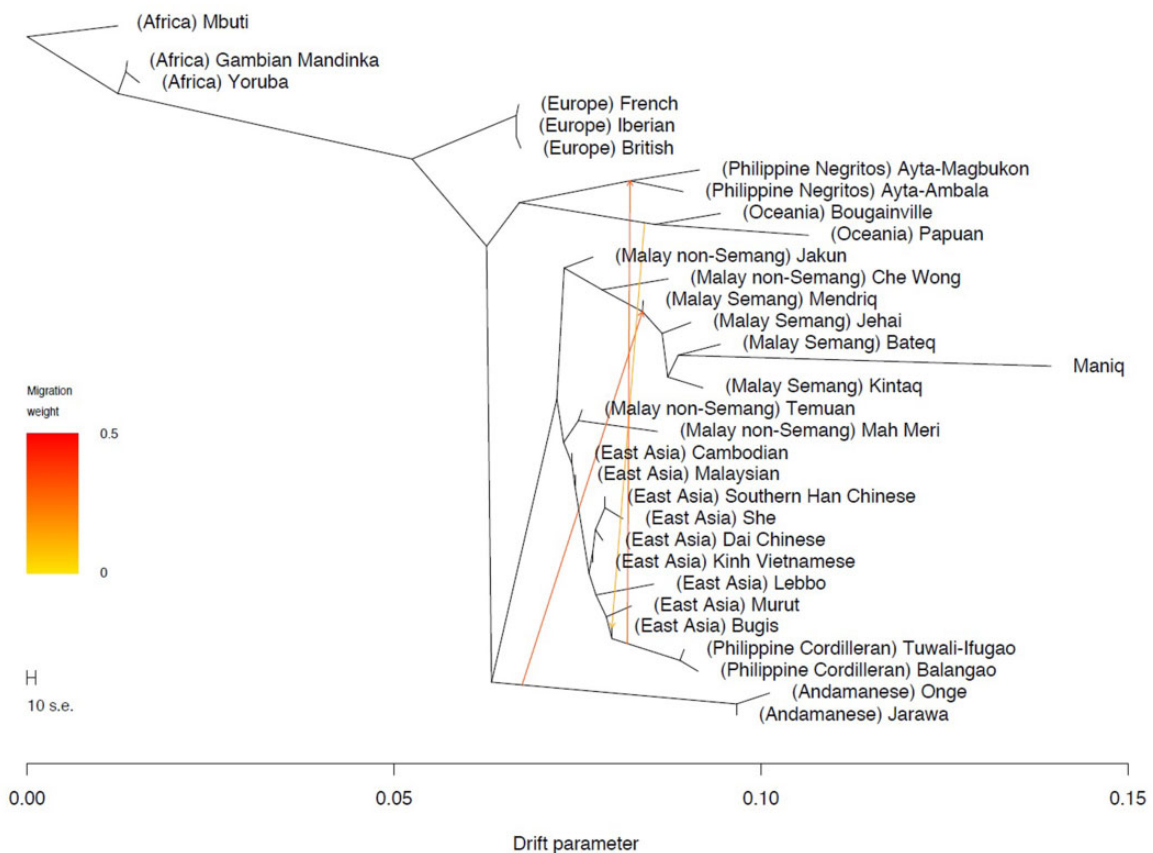


Fig. 6.—Genetic relationship of the Maniq relative to worldwide set of populations. TreeMix analysis with three added migration edges (over 99% explained variance). Number of SNPs used = 190,283.

The long periods of isolation and limited admixture with other populations had evidently an impact on the level of genetic drift among the Maniq. Though high mobility and regular interaction between Semang groups were observed in the earliest anthropological reports, it was noted back then that the Maniq were more isolated than the others, provided

their distance to the any other Semang population (Schebesta 1927; Brandt 1961). Consequently, we find, in this study, the Maniq to be currently one of the populations in the world with the highest levels of population F_{ST} as well as the highest levels of long ROH tracts (supplementary figs. 3, 4A, and B, Supplementary Material online). This is consistent with the

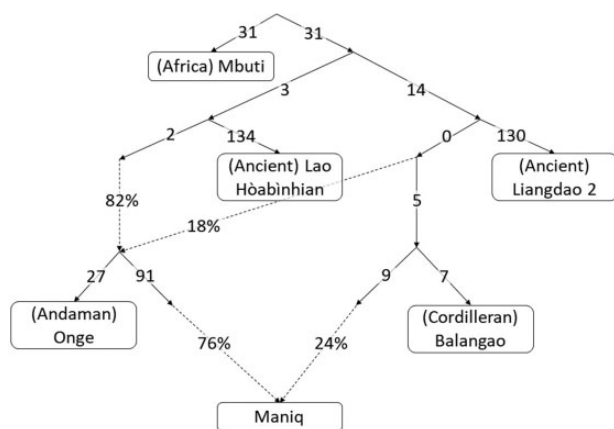


Fig. 7.—Inferred admixture graph model for the Maniq. A possible qpGraph scenario to explain the topology around the Maniq. Worst fitting $Z = 0.788$, number of SNPs used = 56,590.

earlier reports on the cultural practices Maniq, who practice endogamy, and where intermarriage with other non-Maniq (e.g., with Thai or Malay) is largely limited and is still rare until today. This consequently has an impact on the Maniq population size, which was estimated to range only from 100 to 300 individuals prior to 1960s, and was still recently estimated to be around 250 individuals (Brandt 1961; Eberhard et al. 2019). Other factors which impacted the population size in the past remain unknown, but a more recent factor is deforestation, which reduces not only their access to fruits and tubers for gathering, but also their game area for hunting (Lukas 2004). Though there were limited intermarriages, the interactions between the Maniq with other non-Maniq have recently influenced the lifestyle of the Maniq via cultural diffusion. Although the sedentary lifestyle is not looked upon favorably among the Maniq, recently, some Maniq are transitioning from a foraging to a more sedentary Thai lifestyle. With more comprehensive data an investigation on the effects of these high levels of autozygosity on specific phenotypic traits could prove insightful (Clark et al. 2019).

The undetectable levels of Denisovan ancestry among the Maniq supports the evidence that the ancestors of Australopapuans and Philippine Negritos likely experienced a Denisovan introgression event east of Wallace line (Jacobs et al. 2019; Carlhoff et al. 2021; Larena, McKenna, et al. 2021; Teixeira et al. 2021). For the Denisovan introgression event to occur in MSEA, we would expect high levels of Denisovan ancestry among the present-day Negrito groups of MSEA and Borneo. However, our analyses, together with previous findings (Reich et al. 2011; Jinam et al. 2017), do not show the Maniq or other Semang groups of Peninsular Malaysia to possess high levels of Denisovan ancestry, and the levels were in fact undetectable based on standard f_4 -ratio or D tests (supplementary tables 9 and 10, Supplementary Material online). Likewise, all Bornean populations exhibit low to undetectable levels of Denisovan

ancestry (Larena, McKenna, et al. 2021; Teixeira et al. 2021). Additionally, all ancient individuals in the region, including the 4,000-year-old Malay and 8,000-year-old Lao Hòabìnhian hunter gatherers, also did not show significantly high levels of Denisovan ancestry (McColl et al. 2018). Given these findings, it is less likely that the ancestors of Australopapuans or Philippine Negritos picked up the high levels of Denisovan ancestry along the course of their migration in MSEA and Borneo, and it is more parsimonious that they likely experienced an independent introgression event along the islands east of Wallace line.

Like most population genetic studies (e.g., Bergström et al. 2020; Mondal et al. 2016), our findings are limited by the small sample size. Although we have covered more than 3% of all Maniq (11 out of ~300), future studies with larger sample size or with higher resolution via full genome sequencing will be valuable to validate our demographic analyses including the estimates on the impact of archaic introgression.

Materials and Methods

Ethical Considerations and Sample Collection

This study was approved by the Ethics Committee of the University of Vienna (Reference No. 00444) and the Khon Kaen University Ethics Committee for Human Research (Reference No. HE622223). All study participants self-identify as belonging to the Maniq and are at least 18 years old. Consent process was conducted through a series of consultations, where the aims and rationale of the study were thoroughly explained and discussed with the Maniq. Saliva samples were collected from each individual who provided an informed consent ($n = 21$), using the Oragene DNA (OG-500) collection kit by DNA Genotek Inc. (Ontario, Canada). DNA isolation and genome-wide array-based genotyping was carried out by the Life & Brain Genomic Center (Bonn, Germany). We used the Infinium Omni2.5Exome-8 BeadChip (Illumina Inc., San Diego, CA) array which includes about 2.6 million variants, encompassing SNPs on all chromosomes as well as from the mtDNA. This array is essentially a combination of the Infinium Omni 2.5-8 Kit and the Infinium Exome-24 BeadChip (both Illumina Inc., San Diego, CA). The array was analyzed on the iScan system (Illumina Inc.). The results of this study will be discussed together with the Maniq in a future fieldwork trip (a planned visit in 2021 had to be canceled due to the coronavirus disease 2019 pandemic). We are especially interested in their opinions on their long isolation and relationships with other Semang.

Quality Control and Data Preparation

The newly generated genotyped data were processed for quality check and filtering using the PLINK v1.9 software (Purcell et al. 2007). Duplicate samples, duplicated SNPs, and mitochondrial and sex chromosomal SNPs were removed.

All SNPs with a high missing rate ($-\text{geno } 0.001$) as well as all SNPs which failed the Hardy-Weinberg Exact test ($P = 10^{-6}$) were excluded. No sample had to be excluded because of high missing call rates ($-\text{mind } 0.01$). We used the R program version 4.0.3 (R Core Team 2013) to run different R packages. In order to only use unrelated samples for the analyses, we used the R package “SNPrelate” (Zheng et al. 2012). First, we pruned the data set to exclude markers which were in linkage disequilibrium (LD) (`snpGdsLDpruning(x, ld.threshold = 0.5)`, using the default composite measure), then we used the Identity-by-Descent Estimation using a Maximum Likelihood Estimation approach (IBD-MLE) with the following parameters (`snpGdsIBDML(x, snp.id=snpset_pruned_05, maf = 0.01, missing.rate = 0.05)`). A kinship cutoff of 0.125 was applied to identify the related samples, where one of the pair of related samples with the higher ID number was removed, resulting in a base Maniq data set containing 11 unrelated individuals and 2,333,374 SNPs spanning across the 22 autosomes. The base data set was merged with the populations from the published 1000 Genomes Phase 3 (Auton et al. 2015), Simons Genome Diversity Project (Mallick et al. 2016), Human Genome Diversity Project (Bergström et al. 2020), Andamanese (Mondal et al. 2016), Malaysian (Teo et al. 2009; Aghakhanian et al. 2015), Indonesian (Pierron et al. 2014; Mörseburg et al. 2016; Kusuma et al. 2017), and Philippine (Larena, Sanchez-Quinto, et al. 2021) data sets to generate a panel with 320,032 SNPs. In addition, the base data set was also merged with the published ancient DNA data (Lipson et al. 2018; McColl et al. 2018; Larena, Sanchez-Quinto, et al. 2021) to generate panel with 67,094 SNPs for transversion-sites only. All exome SNPs were excluded following QC and merging with other available data sets.

Population Genetic Analysis

Runs of homozygosity (ROH) and inbreeding coefficient were estimated using the `-homozyg` and `-het` flags of PLINK v1.9, respectively. The number of ROH for each tract length category for each population was plotted to infer demographic history (Ceballos et al. 2018), where long tracts of ROH are characteristic of genetic drift caused by recent bottlenecks, founder events, or endogamy. Principal component analysis (PCA) and estimation of between population F_{ST} were performed using EIGENSOFT v7.1 (Patterson et al. 2006; Price et al. 2006). The heat maps based on the F_{ST} values were plotted using the R package “pheatmap” (Kolde and Kolde 2015). The accompanying dendrograms were achieved using the default hierarchical clustering method (`pheatmap(..., clustering_method = "complete")`). The `lsq` project function of EIGENSOFT v7.1 was also utilized to perform PCA on present-day populations and project ancient individuals. Population structure was analyzed using the unsupervised clustering approach implemented in ADMIXTURE v1.3

(Alexander et al. 2009), and limited the number of individuals to a maximum of 10 per population (except for Maniq, where we used all 11 samples). Using the default settings, Admixture analyses were run for 50 iterations for each of the 12 clusters, and the common modes of replicates were ascertained using the CLUMPP program (Jakobsson and Rosenberg 2007). For each K cluster, the major mode was visualized using Pong v1.4 (Behr et al. 2016). To create potential phylogenies with gene flow events we used the TreeMix software (Pickrell and Pritchard 2012). SNPs were pruned beforehand with PLINK, and the trees were computed with Mbuti as outgroup and grouping together 200 SNPs to account for potential LD. Migration edges were added sequentially from 0 to 9. The R package of Admixtools (Patterson et al. 2012) ADMIXTOOLS 2 version 2.0.0 (<https://github.com/uqmaie1/admixtools>, last accessed February 27, 2022), was utilized to model admixture in populations (`qpGraph`), to estimate admixture proportions (`qpAdm`), to estimate shared genetic drift between populations based on outgroup f_3 statistics (`qp3Pop`), and to formally test presence of admixture in populations with f_4 and f_4 -ratio statistics (`qpDstat` and `qpF4Ratio`, respectively). A Z score of >3 was set as a threshold for statistical significance. The default parameters set by ADMIXTOOLS 2 were used in all computations except for the `qpGraph` computation (`qpgraph(..., lsqmode = TRUE, return_fstats = TRUE)`). With these settings the analyses run should be equal to the common settings for the classic Admixtools: `outpop` (NULL), `blgsz` (0.05), `diag` (0.0001), `lsqmode` (YES), and `allsnps` (YES). For an initial assessment of shared drift, we calculated f_3 statistics with the following ancient samples. For Africa, we used “baa001” (Schlebusch et al. 2017) and “Mota” (Llorente et al. 2015), for Europe we used “Bichon” (Jones et al. 2015), “Ust'-Ishim” (Fu et al. 2014, 00), “Loschbour” (Lazaridis et al. 2014), and in Southeast Asia we used “Th521,” “Ma555,” “La368,” “La364,” “Ma912” (McColl et al. 2018), and “Liangdao 2” (Larena, Sanchez-Quinto, et al. 2021). To construct admixture graph models, we incorporated into the model the African Mbuti as an outgroup, the ancient Lao Hòabnhian (“La368”) individual (McColl et al. 2018) and present-day Onge individuals to represent Andamanese-related ancestry, and the ancient Liangdao 2 individual (Larena, Sanchez-Quinto, et al. 2021) and present-day Balangao Cordilleran individuals to represent East Asian-related ancestry. For the implementation of `qpAdm`, we utilized Balangao Cordilleran of the Philippines as the surrogate for the least admixed East Asian source and Onge as the surrogate for the least admixed Andamanese source, and utilized Mbuti, Yoruba, French, Iberian, Surui, and Karitiana (Bergström et al. 2020) as the “right” or outgroup populations. To estimate the date of East Asian and Hòabnhian-related admixture among Maniq and other related populations, we implemented a weighted LD statistic-based method, MALDER (Loh et al. 2013). From the two main data sets (present-day and archaic) we derived subsets of

populations for analysis data sets if needed. Wherever applicable the SNPs were pruned before the analysis with PLINK using `-indep-pairwise 50 10 0.5`. A special data set was created for the ROH analysis, to achieve maximum SNP overlap before pruning. The list of populations, number of SNPs, and number of samples used in the study is detailed in [supplementary table 1A and B](#), [Supplementary Material online](#).

Supplementary Material

[Supplementary data](#) are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank our Ethno-Linguist Pacchira Chindaritha, Tingsabath Charit (Chulalongkorn University), and Khaled Hakami (University of Vienna) for their continued support. We especially thank the Maniq people for their interest and participation. Research reported in this publication was jointly supported via travel grant (grant number: ASEA 2019/Uni Wien/4) by the ASEAN-European Academic University Network (ASEA-UNINET), the Austrian Federal Ministry of Education, Science and Research and the Austrian Agency for International Cooperation in Education and Research (OeAD-GmbH).

Data Availability

The data generated in this study will be made available to researchers after review and approval by the Data Access Committee led by the corresponding authors.

Literature Cited

- Aghakhanian F, et al. 2015. Unravelling the genetic history of Negritos and indigenous populations of Southeast Asia. *Genome Biol Evol.* 7(5):1206–1215.
- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–1664.
- Auton A, et al.; 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* 526(7571):68–74.
- Barrows DP. 1910. The Negrito and allied types in the Philippines. *Am Anthropol.* 12(3):358–376.
- Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S. 2016. Pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* 32(18):2817–2823.
- Bergström A, et al. 2020. Insights into human genetic variation and population history from 929 diverse genomes. *Science* 367(6484):eaay5012.
- Brandt JH. 1961. The Negrito of peninsular Thailand. *J Siam Soc.* 49:123–158.
- Carlhoff S, et al. 2021. Genome of a middle Holocene hunter-gatherer from Wallacea. *Nature* 596:543–547.
- Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. 2018. Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet.* 19(4):220–234.
- Choin J, et al. 2021. Genomic insights into population history and biological adaptation in Oceania. *Nature* 592(7855):583–589.
- Clark DW, et al. 2019. Associations of autozygosity with a broad range of human phenotypes. *Nat Commun.* 10(1):4957.
- Eberhard DM, Simons GF, Fennig CD. 2019. *Ethnologue: languages of Asia*. Dallas (TX): Sil International.
- Evans IHN. 1968. *The Negritos of Malaya*. Hove: Psychology Press.
- Fu Q, et al. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514(7523):445–449.
- Harney É, Patterson N, Reich D, Wakeley J. 2021. Assessing the performance of qpAdm: a statistical tool for studying population admixture. *Genetics* 217(4):iyaa045.
- HUGO Pan-Asian SNP Consortium, et al. 2009. Mapping human genetic diversity in Asia. *Science* 326:1541–1545.
- Jacobs GS, et al. 2019. Multiple deeply divergent Denisovan ancestries in Papuans. *Cell* 177(4):1010–1021.e32.
- Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23(14):1801–1806.
- Jinam TA, et al. 2017. Discerning the origins of the Negritos, first Sundaland people: deep divergence and archaic admixture. *Genome Biol Evol.* 9(8):2013–2022.
- Jones ER, et al. 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun.* 6:1–8.
- Kelly RL. 2013. *The lifeways of hunter-gatherers: the foraging spectrum*. Cambridge: Cambridge University Press.
- Kirin M, et al. 2010. Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5(11):e13996.
- Kolde R, Kolde MR. 2015. Package ‘pheatmap’. *R Package* 1:790.
- Kricheff DA, Lukas H. 2015. Being Maniq: practice and identity in the forests of Southern Thailand. *Hunt Gatherer Res.* 1(2):139–155.
- Kusuma P, et al. 2017. The last sea nomads of the Indonesian archipelago: genomic origins and dispersal. *Eur J Hum Genet.* 25(8):1004–1010.
- Kutanan W, et al. 2018. Contrasting maternal and paternal genetic variation of hunter-gatherer groups in Thailand. *Sci Rep.* 8(1):1536.
- Larena M, McKenna J, et al. 2021. Philippine Aytá possess the highest level of Denisovan ancestry in the world. *Curr Biol.* 31:4219–4230.e10.
- Larena M, Sanchez-Quinto F, et al. 2021. Multiple migrations to the Philippines during the last 50,000 years. *Proc Natl Acad Sci U S A.* 118:e2026132118.
- Lazaridis I, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513(7518):409–413.
- Lipson M, et al. 2018. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Genetics* 361:92–95.
- Llorente MG, et al. 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* 350(6262):820–822.
- Loh P-R, et al. 2013. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* 193(4):1233–1254.
- Lukas H. 2004. Can “They” save “Us”, the Foragers? Indonesian and Thai Hunter-Gatherer Cultures under Threat from Outside. Südostasien Working Papers Band 2. doi: 10.1553/soawp2.
- Mallick S, et al. 2016. The Simons genome diversity project: 300 genomes from 142 diverse populations. *Nature* 538(7624):201–206.
- Matsumura H, et al. 2008. Morphometric affinity of the late Neolithic human remains from Man Bac, Ninh Binh Province, Vietnam: key skeletons with which to debate the ‘two layer’ hypothesis. *Anthropol Sci.* 116:0802160030.
- Matsumura H, et al. 2019. Craniometrics reveal “two layers” of prehistoric human dispersal in eastern. *Eurasia Sci Rep.* 9:1–12.
- McColl H, et al. 2018. The prehistoric peopling of Southeast Asia. *Science* 361(6397):88–92.
- Mondal M, et al. 2016. Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nat Genet.* 48(9):1066–1070.

- Mörseburg A, et al. 2016. Multi-layered population structure in Island Southeast Asians. *Eur J Hum Genet.* 24(11):1605–1611.
- O'Connor RA. 1995. Agricultural change and ethnic succession in Southeast Asian states: a case for regional anthropology. *J Asian Stud.* 54:968–996.
- Patterson N, et al. 2012. Ancient admixture in human history. *Genetics* 192(3):1065–1093.
- Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet.* 2 (12):e190.
- Pickrell J, Pritchard J. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8:e1002967.
- Pierron D, et al. 2014. Genome-wide evidence of Austronesian–Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. *Proc Natl Acad Sci U S A.* 111(3):936–941.
- Pittayaporn P. 2014. Layers of Chinese loanwords in protosouthwestern Tai as evidence for the dating of the spread of southwestern Tai. *MANUSYA* 17(3):47–68.
- Price AL, et al. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 38(8):904–909.
- Purcell S, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81(3):559–575.
- R Core Team. 2013. *R: a language and environment for statistical computing.* Vienna (Austria): R Foundation for Statistical Computing.
- Radcliffe-Brown AR. 2013. *The Andaman Islanders.* Cambridge: Cambridge University Press.
- Reich D, et al. 2011. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am J Hum Genet.* 89(4):516–528.
- Schebesta P. 1927. *Bei den Urwaldzweigen von Malaya.* Leipzig: FA Brockhaus.
- Schlebusch CM, et al. 2017. Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science* 358(6363):652–655.
- Skeat WW, Blagden O, Ridley HN. 1907. The Pagan Races of the Malay Peninsula. *J Straits Branch R Asiat Soc.* 49:1–5.
- Teixeira JC, et al. 2021. Widespread Denisovan ancestry in Island Southeast Asia but no evidence of substantial super-archaic hominin admixture. *Nat Ecol Evol.* 5(5):616–624.
- Teo Y-Y, et al. 2009. Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations. *Genome Res.* 19(11):2154–2162.
- Wang T, et al. 2021. Human population history at the crossroads of East and Southeast Asia since 11,000 years ago. *Cell* 184(14):3829–3841.e21.
- Zheng X, et al. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics.* 28(24):3326–3328.

Associate editor: David Enard