



DISSERTATION | DOCTORAL THESIS

Titel | Title

Genomic Insights on the Maniq: Population Genetics, Natural
Selection and Immunogenetics

verfasst von | submitted by

Tobias Herzog BSc MSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Doctor of Philosophy (PhD)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 794 685 437

Dissertationsgebiet lt. Studienblatt | Field of
study as it appears on the student record
sheet:

Biologie

Betreut von | Supervisor:

Assoz. Prof. Mag. Dr. Martin Fieder Privatdoz.
Mag. Dr. Helmut Schaschl Privatdoz.

*... and we will share stories,
together at the fire.*

Table of Contents

Acknowledgements	4
Abstract	5
Zusammenfassung	6
Preamble	7
The Term ‘Negrito’	7
Introduction and Scope.....	7
The Maniq	9
Peopling of Southeast Asia.....	10
Adaptation and Immunogenetics	11
Outlook.....	12
Chapter I.....	13
Unveiling the Genetic History of the Maniq, a Primary Hunter-Gatherer Society	13
Supplementary Information.....	26
Chapter II.....	50
Natural Selection and Adaptive Traits in the Maniq, a Nomadic Hunter-Gatherer Society from Mainland Southeast Asia.....	50
Supplementary Information.....	60
Chapter III	62
HLA Diversity and Signatures of Selection in the Maniq, a Nomadic Hunter-Gatherer Population in Southern Thailand	62
Supplementary Information.....	76
Concluding Discussion.....	84
The Maniq as an Isolated Population	84
Adapting to a Hunter-Gatherer Lifestyle and the Tropical Rainforest	86
Immunogenetic Patterns of the Maniq	88
The Future of the Maniq.....	89
References	90
AI Statement.....	92

Acknowledgements

First and foremost, I want to thank the Maniq not only for participating in our research but also for their positive impact on my life. The stark contrasts between hunter-gatherers and our society serve as a constant reminder to think critically about what we attribute to biology and what to culture.

The journey to this thesis started with my supervisor Martin Fieder who inspired me to pursue my master's degree and then continued to support me during my PhD, for this I want to thank him. To Helmut Schaschl, my other supervisor, I want to extend my heartfelt gratitude for believing in our project and in me, for his support during the more challenging phases and for his overall positive spirit, which inspired me throughout the project. Many thanks go out to all our collaborators over the years; there are too many to name each one of you! Special thanks to Maximilian Larena who was invaluable to the success of the project and who taught me so much. A heartfelt thank you to Wibhu Kutanan, who did not hesitate to help us with our project. I want to thank Helmut Lukas for his tireless work with the Maniq, his never-ending stories and his trust in us. Thanks to Pacchira Chindaritha for enabling our discussions with the Maniq and for her hospitality. I also want to thank Khaled Hakami for opening my eyes towards hunter-gatherer research and broadening my view of the world. I want to thank the CHAGS community and especially the student organizing committee for CHAGS XI (2015) in Vienna. I deeply thank Mona Röhm, Bettina Ludwig, Alena Brunner and Reinhard Blumauer for endless discussions and their continued support and friendship.

Thanks to my colleagues of Statistics Austria for their interest and support. A heartfelt thank you goes out to all my friends who supported and encouraged me throughout this time, especially Lukas Liebl. Finally, I want to thank my family, especially my parents, for enabling me to pursue this endeavour and their tireless support.

Lastly, my deepest gratitude goes to my wife Johanna for her endless patience and love, our daughter Mona for brightening every day, and our future child, who motivated me to finally finish this thesis.

Abstract

The Maniq of southern Thailand are one of the last remaining practicing hunter-gatherer communities in the world, yet our knowledge of their genomics is largely limited. This thesis presents novel insights into their evolutionary history, population genetics, genetic adaptations, and immunogenetics. Our analyses reveal the Maniq to be closely related to the Semang populations of Malaysia, who altogether can be modelled as an ancestry mixture of Andamanese and East Asians. The Maniq exhibit the highest values of Andamanese-related ancestry when compared with other Semang and show signs of more recent admixture with East Asian-related ancestry. 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. Their isolation is also apparent in their comparatively low HLA diversity, with only a few alleles occurring at higher frequencies. We investigated positive and balancing selection in the Maniq and found evidence pointing towards adaptations to their lifestyle and environment. Adaptation to a rainforest environment includes enriched pathways in extracellular matrix organization and collagen synthesis, which are linked to structural support and mobility, as well as positive selection acting upon adrenergic signaling in cardiomyocytes pathway, probably to increase stress resilience and physical performance. Balancing selection was primarily detected in immune-related genes, particularly within the HLA region, underscoring the critical role of genetic diversity in surviving pathogen-rich environments. Interestingly, especially regions associated with responses to hepatitis B virus were under selection, which may have contributed to recent HLA evolution in the Maniq. This thesis provides a comprehensive genomic analysis of the Maniq.

Zusammenfassung

Die Maniq im Süden Thailands sind eine der letzten noch existierenden Jäger- und Sammlergemeinschaften der Welt, jedoch ist unser Wissen über ihre Genomik noch sehr gering. Diese Dissertation präsentiert neue Erkenntnisse über ihre Evolutionsgeschichte, Populationsgenetik, genetischen Anpassungen und Immunogenetik. Unsere Analysen zeigen, dass die Maniq nah mit den Semang Gruppen Malaysias verwandt sind, die insgesamt als eine Mischung aus Andamanesen und Ostasiaten modelliert werden können. Die Maniq weisen im Vergleich zu anderen Semang die höchsten Werte für andamanische Abstammung auf und zeigen Anzeichen Vermischung in jüngerer Zeit mit ostasiatischer Abstammung. Die Maniq weisen einen der höchsten Grade an genetischer Differenzierung unter den lebenden menschlichen Populationen auf, was auf ihre geringe Populationsgröße und ihre historische Praxis der Endogamie hindeutet. Ihre Isolation zeigt sich auch in ihrer vergleichsweise geringen HLA-Vielfalt, wobei nur wenige Allele mit höherer Häufigkeit vorkommen. Wir untersuchten directionale und stabilisierende Selektion bei den Maniq und fanden Hinweise auf Anpassungen an Lebensstil und Umwelt. Die Anpassung an den Regenwald umfasst angereicherte Signalwege in der extrazellulären Matrixorganisation und Kollagensynthese, die mit struktureller Unterstützung und Mobilität in Verbindung stehen, sowie directionale Selektion, die auf den adrenergen Signalweg in Kardiomyozyten wirkt, wahrscheinlich um die Stressresistenz und körperliche Leistungsfähigkeit zu erhöhen. Stabilisierende Selektion wurde vor allem in Genen mit Immunbezug festgestellt, insbesondere innerhalb der HLA-Region, was die entscheidende Rolle der genetischen Vielfalt für das Überleben in pathogenreichen Umgebungen unterstreicht. Interessanterweise standen insbesondere Regionen, die mit der Reaktion auf das Hepatitis-B-Virus in Verbindung stehen, unter Selektion, was möglicherweise zur jüngeren HLA-Evolution bei den Maniq beigetragen hat. Diese Dissertation liefert eine umfassende genomische Analyse der Maniq.

Preamble

The Term ‘Negrito’

In this thesis and the respective papers, we used the term ‘negrito’ as it is commonly used in the literature. However, during the process of this work over the last years, we have come to realize that the use of this term should have been avoided. From a biological viewpoint, this term does not bear any scientific meaning. It is used to describe a supposed phenotype, yet the same phenotype can be found in prominently in Africa, where it was called ‘pygmy’. If the term is used to describe the first people to come to Southeast Asia, then there is no reason to use a term that has questionable historical roots and does not describe the grouping. Instead, we would suggest the term ‘First Sundaland People’ coined by Jinam et al. (2017). Even though we were aware of this publication, we failed to incorporate their suggestion into our publications.

Viewing the term through a cultural and social anthropology lens, the term can have some usefulness, especially when working with older publications or historical records. The term is partially being reclaimed by populations of the Philippines to unite under a term for a political agenda, which is, of course, their right and own decision. We underscore only very limited use of the term in human genetics. Lastly, as biological anthropologists, we should always strive to use the respective endonyms of the populations we work with. We are certain that ‘Maniq’ is the correct endonym for them, yet in discussions we realized that we sometimes reproduced exonyms for other populations where we do not have any contact and only know them via the scientific literature.

Introduction and Scope

The oldest fossils of anatomically modern humans are found in Africa and date at least as far back as 200 000 years ago (Stringer 2016). Even though there are older fossils dating back up to 300 000 years ago (Hublin et al. 2017; Richter et al. 2017), there is debate as to whether these can be classified as anatomically modern. The discussion of how long humans have existed on Earth is steadily developing as new evidence, also outside of Africa, is being evaluated (see Bergström et al. 2021 for a review

on this discussion). However, the advent of agriculture started only 15 000 to 10 000 years ago (Diamond and Bellwood 2003). This means that hunting and gathering was the dominant mode of subsistence for around 90–95% of the time that anatomically modern humans have existed. Populations that still practice hunting and gathering as their main source of subsistence are rare among present populations, as there are only very few full-time and -nomadic hunter-gatherers left (Kelly 2013). For a long time, studies, driven by diverse scientific disciplines, focused mainly on African contemporary hunter-gatherer societies. In contrast, hunter-gatherer populations from Asia have received less attention. Southeast Asia is inhabited by various indigenous human populations that show great morphological, anthropological, and linguistic diversity. In 2012, the journal *Human Biology* released a special issue (Volume 85, Issue 1; Revisiting the ‘Negrito’ Hypothesis) of hunting and gathering societies in Southeast Asia, acknowledging the research conducted in this geographic area. One of the main research topics concerns the ‘negritos’ (Spanish diminutive for ‘black’) living in the rainforests of Southeast Asia. The ‘negritos’ are indigenous to this area and are mainly studied in Malaysia. Although Malaysian ‘negritos’ are genetically closest to East Asian populations, they are nevertheless substantially distinct (Aghakhanian et al. 2015). This distinct genetic pattern may have resulted from genetic drift, longer periods of isolation from other populations or potentially due to genetic adaptation to specific environmental conditions.

Part of this indigenous population in Southeast Asia are the Maniq, who live in southern Thailand, near the Malaysian border. The Maniq are of particular interest because they are one of the few remaining practicing primary hunter-gatherer societies, and they have not participated in many published studies. We cooperate with Helmut Lukas (Austrian Academy of Sciences)¹ who conducts long-term participant observation with the Maniq. Together, we aim to uncover the genetic ancestry, diversity and adaptations of the Maniq.

¹ Khaled Hakami (University of Vienna) was crucial in forming this cooperation and contributed to the initial ideas of the study design, yet he could not work with us on this project directly.

The Maniq

The Maniq are hunter-gatherers living in the rainforests of southern Thailand (provinces Satun, Phatthalung and Trang). They represent the northernmost Semang, who are mainly found in Malaysia.

They are fully nomadic and travel in groups of 10 to 55 people. Because there are no signs that the Maniq ever practiced agriculture, they are described as primary hunter-gatherers (Lukas 2004). However, they maintain marginal trade relations with their sedentary neighbours.

The term ‘Maniq’ is an endonym meaning ‘us’. Their subsistence activities (hunting, gathering, and fishing) require small groups (‘bands’) that wander through the rainforest. Social stratification is low, with a strong emphasis on reciprocity. Their camps consist of leaf shelters (windcreens, ‘hapoi’ in the Maniq language) with multiple fireplaces. A recent phenomenon is that the Maniq sometimes work for their surrounding Thai neighbours. The Maniq operate an economy where they receive immediate yield for their work and property rights are very minimal. For an ‘immediate-return system’ these aspects and egalitarianism are essential (Woodburn 1982). Overall, the Maniq represent an ideal type of immediate-return hunter-gatherers. Features of their society include: low population density, flexible social groups with constantly changing composition (‘flux’), a nomadic life style, lack of dependence on stored foods, an immediate return system, no ownership of resources, reciprocity and an egalitarian sharing ethic, almost no specialization in work (everybody can produce everything) and minimal social stratification (Lukas 2004).

Our work would not have been possible without the long-established relationship of trust between the Maniq and Helmut Lukas. He regularly visits the Maniq with different cooperation partners to conduct various research projects (e.g. Kricheff & Lukas 2015). The Maniq are highly endangered by the deforestation for rubber plantations and by poachers (Lukas 2004).

‘Hamiq’ means ‘them’ and refers to everyone who does not live the Maniq way of life. The Maniq classify living beings – not only humans – into those who live like them and those who do not. For instance, the Maniq classify the tiger as being more closely related to them, than a farmer living nearby. Due to rapid deforestation, contact with other humans is becoming more frequent and inevitable. Recently, the phenomenon of ‘transitioning’ from their traditional way of life to a sedentary and agricultural lifestyle is occurring more often. Maniq who transition to sedentarism are often cast out of the

community, and are seldom welcomed back, as settling down and planting crops are strict taboos in their culture. From 2010 onwards such transitions have occurred more frequently, and today the Maniq use two new terms, ‘Maniq-Hamiq’ and ‘Hamiq-Maniq’, to describe individuals transitioning between ways of life. Ethnicity and even species play a minor role in their worldview.

Peopling of Southeast Asia

To address the question of how anatomically modern humans populated Southeast Asia, data from many different populations must be carefully examined. The Maniq are one population that can contribute to elucidating this question. The earliest theory regarding the indigenous people of the Malay Peninsula is the ‘Wellentheorie’ (German for ‘wave theory’, commonly known as the ‘layer-cake theory’ in English literature). In 1927, the Austrian anthropologist Schebesta proposed that the Semang were the oldest inhabitants, followed by the Senoi, and finally by the Aboriginal Malay (Schebesta 1927). In contrast, Rambo proposed that 6 000 years ago the Semang and Senoi shared common ancestors and subsequently adapted physically to different ecological niches (Rambo 1988).

Using mtDNA analysis, Hill et al. (2006) provided evidence for a deep genetic root (dated to ~60 kya) in ‘negrito’ hunter-gatherers (Semang and Senoi), as well as in Aboriginal Malay, who exhibit a different phenotype. They identified four different colonisation events for these three indigenous groups and concluded that, while those groups share a genetic root, they experienced different local evolutionary trajectories combined with varying immigration patterns. Later, an ‘Early Train’ migration (~30 kya to 10 kya) from Indochina or southern China was proposed based on analyses of mtDNA and single nucleotide polymorphisms (SNPs) from numerous Southeast Asian populations (Jinam et al. 2012). Furthermore, phylogenetic tree analyses suggest that ‘negritos’ are basal to other East and Southeast Asians, and may have diverged from West Eurasians at least 38 kya (Jinam et al. 2017). Another study, using data from over two million SNPs, suggested multiple waves of migration to and through Southeast Asia, and stressed the genetic uniqueness of its populations (Aghakhanian et al. 2015). Genetic analyses of the Andamanese, in contrast, suggest a single origin and an Out-of-Africa expansion with significant genomic contributions from an extinct hominin population (Mondal et al. 2016). However, this conclusion triggered some controversy (see Skoglund et al. 2018). Although

Malaysian ‘negritos’ are genetically closest to East Asian populations, they remain substantially distinct (Aghakhanian et al. 2015). This distinct genetic pattern may result from genetic drift, long periods of isolation, or genetic adaptation to specific environmental conditions.

The ‘negrito hypothesis’ predicts that a shared phenotype among various contemporary groups of hunter-gatherers in Southeast Asia – dark skin, short stature and tight curly hair – is due to common descent from a region-wide, pre-Neolithic substrate population. The alternative hypothesis suggests that their phenotype results from convergent evolution due to adaptation to similar eco-environmental conditions (Endicott 2013; Bergey et al. 2018). Nevertheless, despite sharing small adult stature, ‘negritos’ show considerable variation in body dimensions, suggesting the absence of a generic, morphologically integrated phenotype typical of small-bodied foragers (Stock 2013). We aim to contextualize whether and how the Maniq can contribute to these discussions, while maintaining a focus on their specific genetic history.

Adaptation and Immunogenetics

Over time, *Homo sapiens* evolved many adaptations to different local environments. While most human populations today are genetically quite similar, variation persists in adaptations to diet, altitude and pathogens. Well-known examples of human adaptations include lactase persistence, malaria resistance or genes involved in the adaptation of life at high altitudes (Rees et al. 2020). Positive selection increases the frequency of a particular allele in a population, which can sometimes lead to fixation of the allele. In contrast, balancing selection favours multiple alleles and prevents fixation. Both forces leave detectable genomic signals, that can be identified using various methods and algorithms.

The major histocompatibility complex (MHC) on chromosome 6 contains the highly polymorphic human leukocyte antigen (HLA) system. HLA genes are a crucial component of the immune system and can be divided into MHC class Ia and IIa genes (also called ‘classical’; highly polymorphic) and the class Ib and IIb genes (‘non-classical’; less polymorphic). The high diversity of functionally distinct HLA alleles makes it a prime region to study disease risk among individuals and/or populations. In recent years, genome-wide association studies (GWAS) have identified a large number of diseases which are associated with different variants in the MHC region (Kennedy et al. 2017).

We aim to examine adaptations that occurred within the Maniq, including immunogenetic adaptations. While some Orang Asli populations have participated in such studies (e.g. Jinam et al. 2010, 2022; Tasnim et al. 2016), to our knowledge the Maniq have never been included. Our primary interest lies in adaptations to their tropical rainforest environment, including pathogen exposure, as well as adaptations related to their hunter-gatherer lifestyle and diet.

Outlook

In the following chapters, we will examine at the genetic and evolutionary history of the Maniq, their genetic adaptations to their habitat and their immunogenetic signatures. The evidence presented contribute to and deepen our understanding of the Maniq in particular and primary hunter-gatherer populations in general.

Chapter I

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

Authors: Tobias Göllner, Maximilian Larena, Wibhu Kutanan, Helmut Lukas, Martin Fieder, Helmut Schaschl

Reference: Göllner, T., Larena, M., Kutanan, W., Lukas, H., Fieder, M., & Schaschl, H. (2022). Unveiling the Genetic History of the Maniq, a primary hunter-gatherer society. *Genome Biology and Evolution*, 14(4), evac021. <https://doi.org/10.1093/gbe/evac021>

Status: Published in Genome Biology and Evolution.

My contribution: I conceived and designed this study. I drafted and finalized the pipelines to create the various analysis datasets. I performed all statistical analyses of all programmes used. I created all figures and tables. I wrote the draft manuscript and edited the final manuscript. Afterwards I drafted the revised version and edited the final revision.

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 Hòabinhian 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 Hòabinhian 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 Hòabinhian 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 arboreal hills of southern Thailand (Krichauff 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

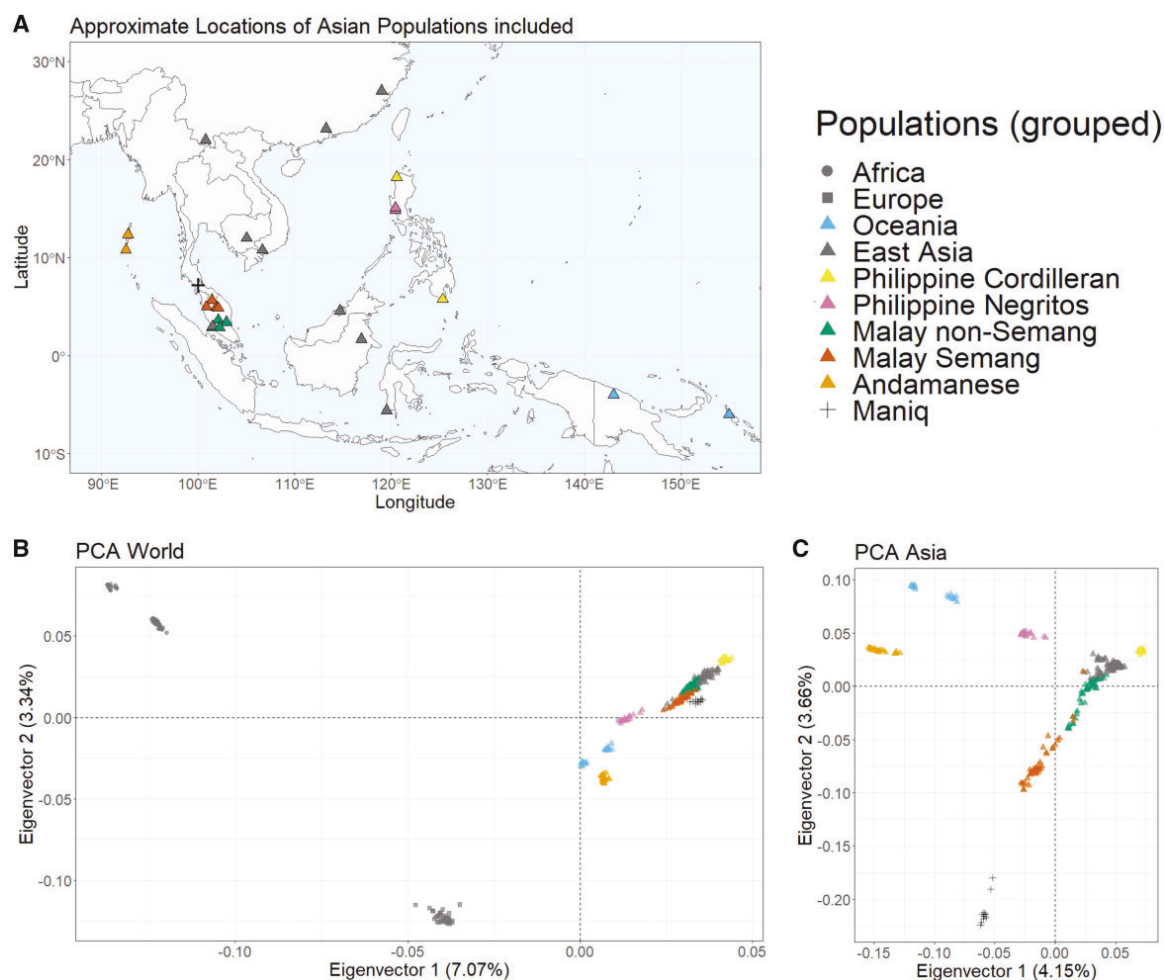


Fig. 1.—Maniq form a genetic cluster close to Malay Semang populations. (A) Map with approximate locations of all Asian populations included in the study. (B) Eigenvector 1 and 2 of a PCA on a worldwide set of populations, number of SNPs used = 320,032. (C) Eigenvector 1 and 2 of a PCA restricted to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

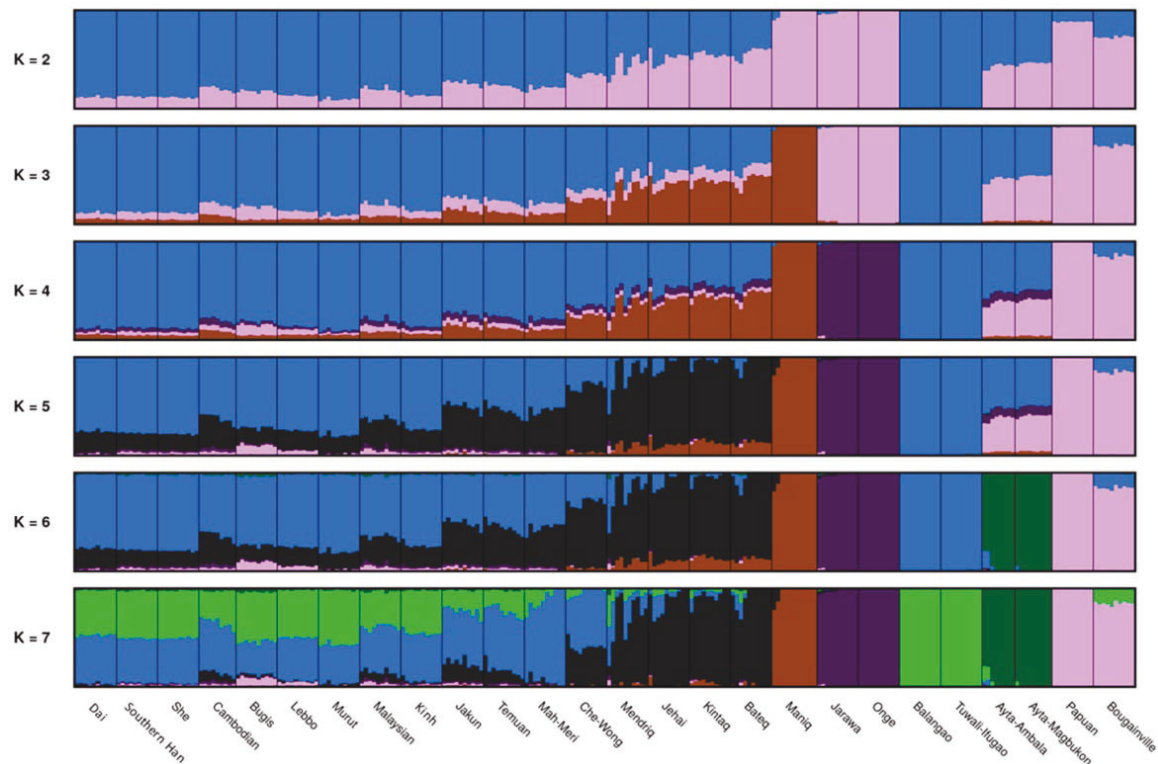


Fig. 2.—Admixture analysis for Asian populations. admixture plots for $K = 2$ – 7 with 170,513 SNPs used ($K = 7$ has the lowest cross-validation error).

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 Hòabinhian-Related Populations of MSEA

It was previously argued that the ancient hunter-gatherers of the Hòabinhian cultural complex are regarded as the ancestors of present-day hunter-gatherers of MSEA (McColl et al. 2018). This was recently supported by an ancient DNA data, where among the ancient individuals, Onge and Jarawa were shown to be most genetically affiliated with an 8,000-year-old Hòabinhian individual from Pha Faen, Laos and a 4,000-year-old Hòabinhian individual from Gua Cha, Malaysia (McColl 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 Hòabinhian, 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 Hòabinhian-related and East Asian-related ancestries for the Maniq. We then formally tested the presence of Hòabinhian-related ancestry in the Maniq, other Semang, and Andamanese populations with $f_4(\text{Mbuti}, \text{Lao Hòabinhian}; \text{Liangdao 2}, X)$ and $f_4(\text{Mbuti}, \text{Lao Hòabinhian}; \text{Balangao}, X)$, where we examine which X population share alleles with the Lao Hòabinhian relative to a least admixed East Asian individual/population, Liangdao 2 or Cordilleran Balangao (Larena,

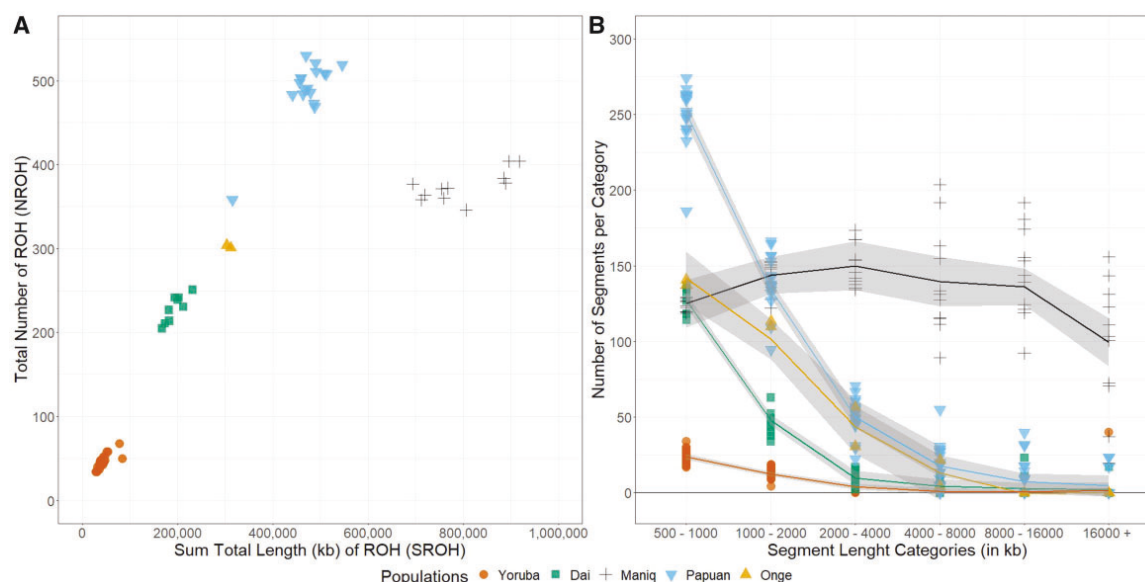


Fig. 3.—High levels of genetic differentiation in the Maniq. (A) Total length of ROH versus total number of ROH per individual and (B) Number of ROH per segment length category covering the Yoruba, Dai, Maniq, Papuan, and Onge populations (for additional populations see [supplementary figs. 4A and B](#),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

[Sanchez-Quinto, et al. 2021](#)). As expected and consistent with previous observations ([McColl et al. 2018](#)), Onge and Jehai displayed significant levels of Hòabínhian-related ancestry. Looking at the remaining Semang populations, the Maniq displayed also high levels of Hòabínhian-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 ([Aghakhani 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

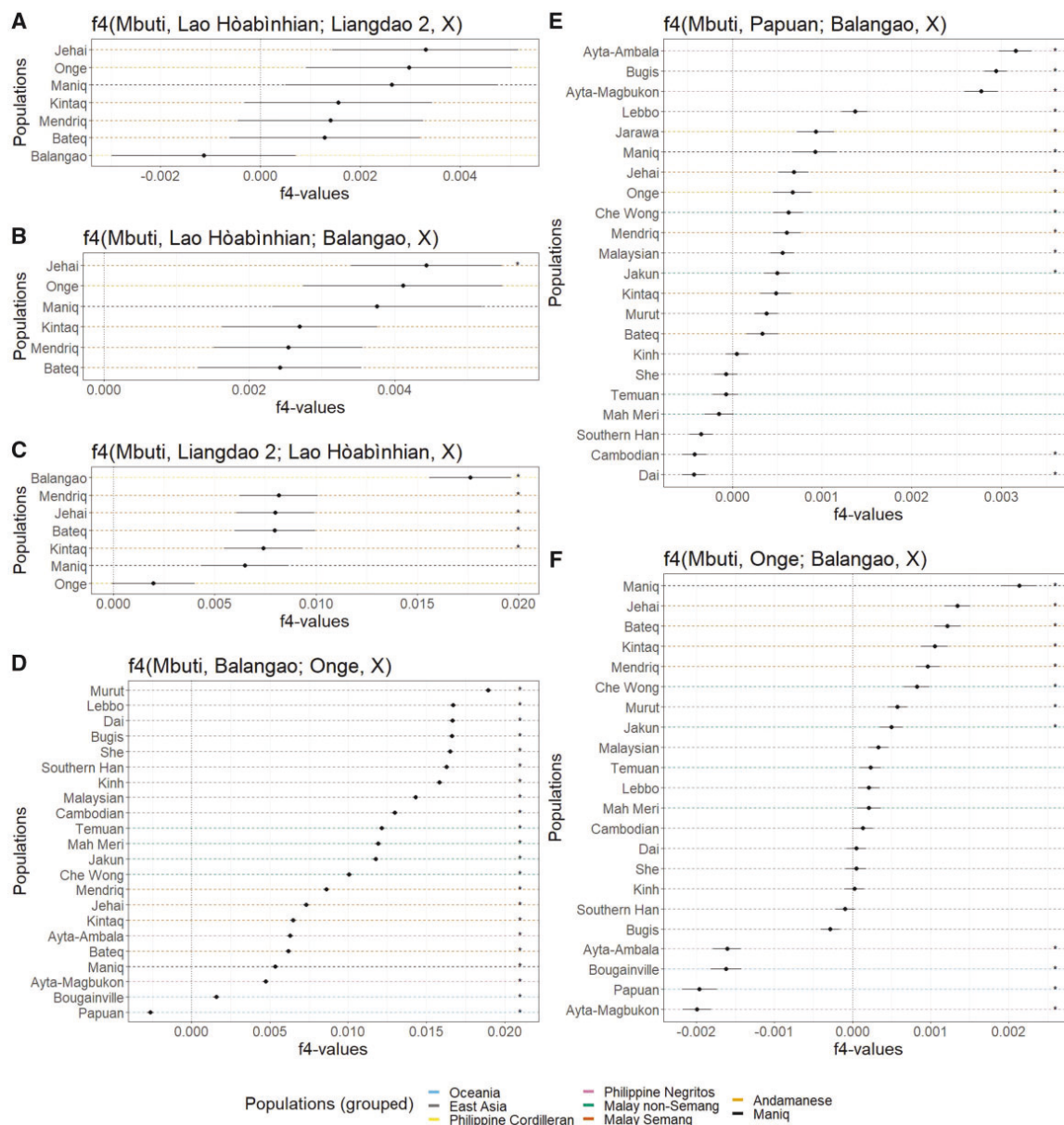


Fig. 4.—Different ancestry components of the Maniq. Results of f_4 statistics including standard errors for (A) $f_4(\text{Mbuti, Lao Hòabinhian; Liangdao 2, X})$, (B) $f_4(\text{Mbuti, Lao Hòabinhian; Balangao, X})$, (C) $f_4(\text{Mbuti, Liangdao 2; Lao Hòabinhian, X})$, and (D) $f_4(\text{Mbuti, Balangao; Onge, X})$, (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 Hòabìnhian, 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: $f_4(\text{Neanderthal(Altai)}, \text{Chimp}; X, \text{Yoruba})/f_4(\text{Neanderthal(Altai)}, \text{Chimp}; \text{Neanderthal(Vindija)}, \text{Yoruba})$, 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 $f_4(\text{Chimp}, \text{Neanderthal}; \text{Mbuti}, X)$ ([supplementary table 8](#), [Supplementary Material online](#)). Using $f_4(\text{Chimp}, \text{Denisovan}; \text{European}, X)$ and the f_4 -ratio test $f_4(\text{Chimp}, \text{Neanderthal}; \text{Southern Han}, X)/f_4(\text{Chimp}, \text{Neanderthal}; \text{Southern Han}, \text{Denisovan})$, 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 Australopapuan (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 Hòabìnhian hunter-gatherers. Hence, like the Malay Semang, we find the Maniq to possess strong genetic affiliation with the 8,000-year-old Lao Hòabìnhian individual. Historically, populations with high levels of Hòabìnhian-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 Hòabìnhian-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 Hòabìnhian-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 Hòabìnhian-related population who arrived in MSEA and remained largely distinct, and who later received limited admixture with neighboring populations carrying East Asian ancestry.

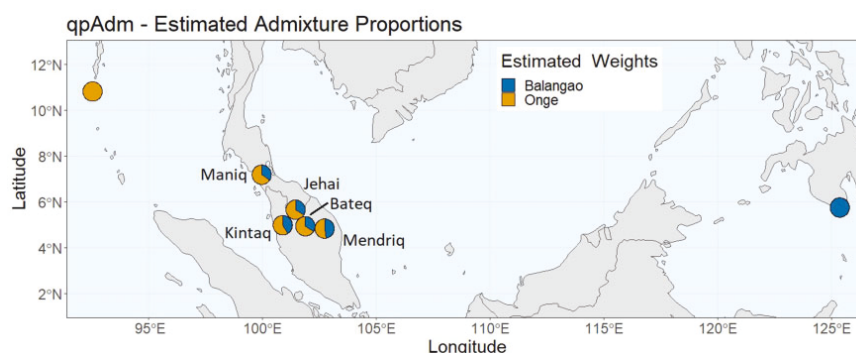


Fig. 5.—Estimated admixture proportions of Andamanese-related and East Asian-related ancestries of the Semang Populations of MSEA. Pie chart results of qpAdm with 319,386 SNPs used. The position of the pie charts corresponds to the approximate location of their respective populations (Aghakhani et al.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

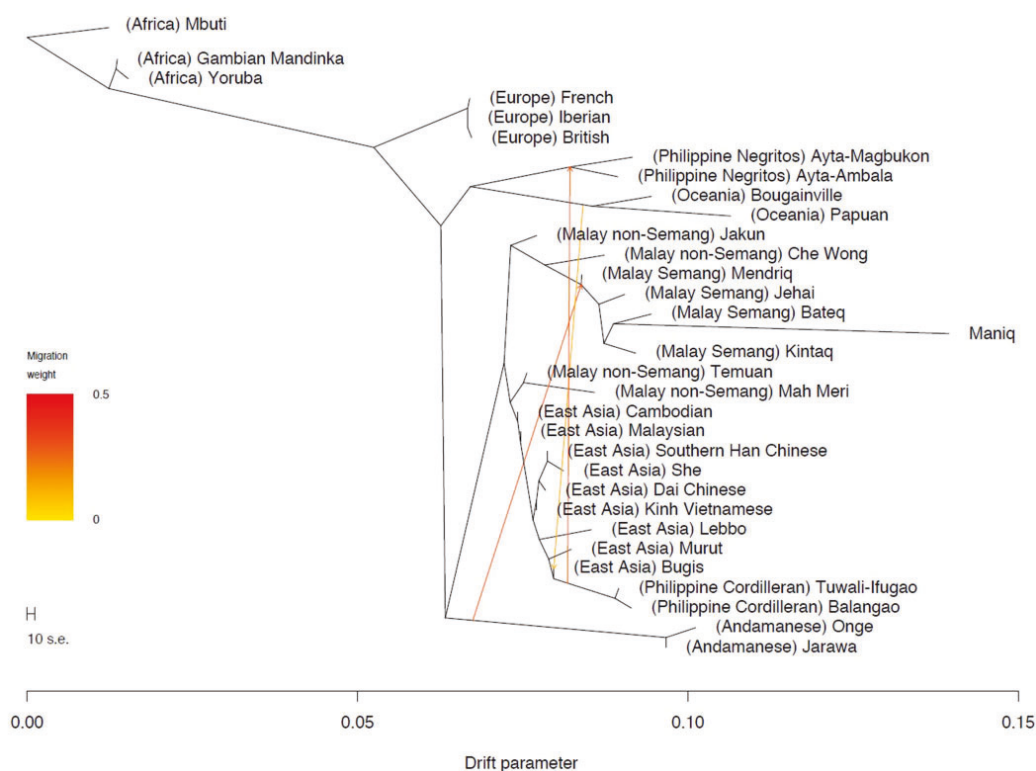


Fig. 6.—Genetic relationship of the Maniq relative to worldwide set of populations. TreeMix analysis with three added migration edges (over 99% explained variance). Number of SNPs used = 190,283.

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

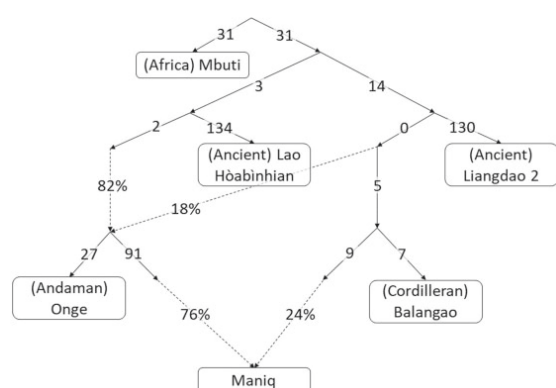


Fig. 7.—Inferred admixture graph model for the Maniq. A possible qpGraph scenario to explain the topology around the Maniq. Worst fitting $Z = 0.788$, number of SNPs used = 56,590.

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 Australopapuan 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 Hòabinhian 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 Australopapuan 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 (snpgdsIBDMLE(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 $-\text{homozyg}$ and $-\text{het}$ 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/uqrmaie1/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 f3 statistics (qp3Pop), and to formally test presence of admixture in populations with f4 and f4-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), blgsize (0.05), diag (0.0001), lsqmode (YES), and allsnps (YES). For an initial assessment of shared drift, we calculated f3 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 Hòabínhian (“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 Hòabínhian-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, Tingsabadh 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

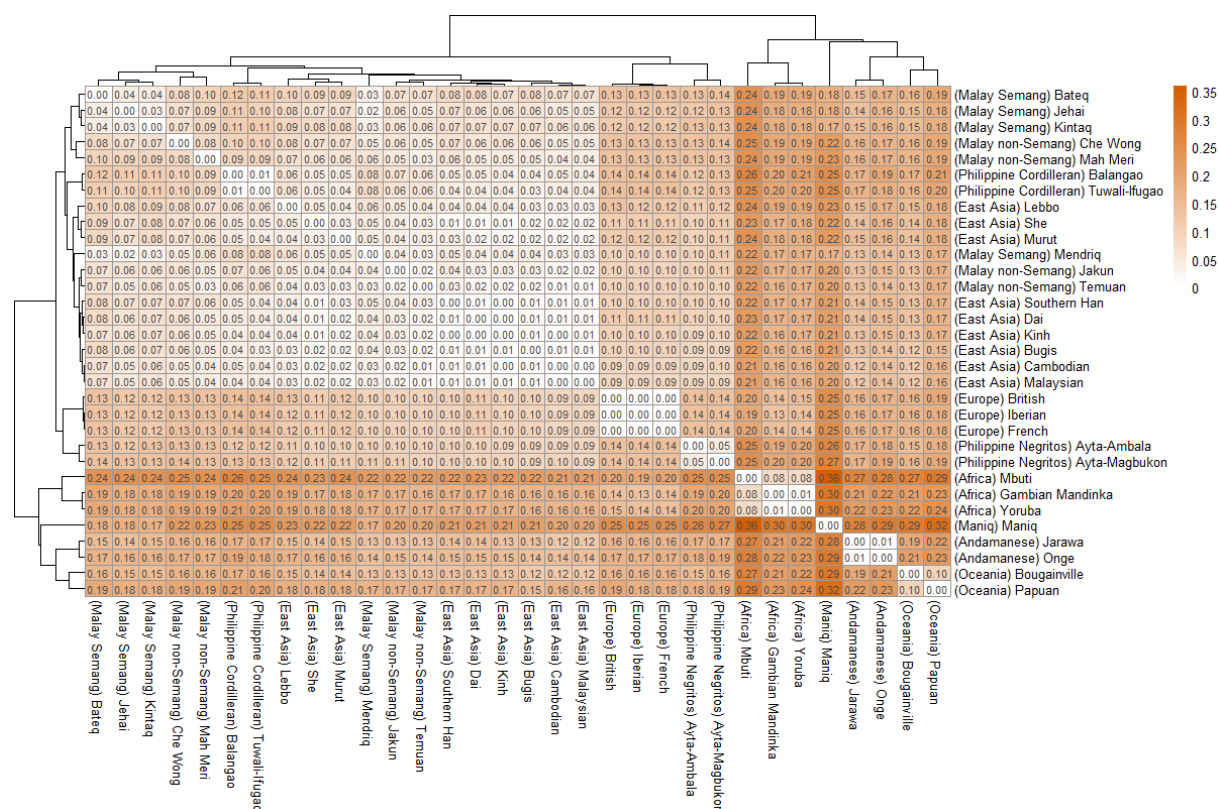
- Aghakhani F, et al. 2015. Unravelling the genetic history of Negritos and indigenous populations of Southeast Asia. *Genome Biol Evol.* 7(5):1206–1215.
- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–1664.
- Auton 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 Ayta 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.

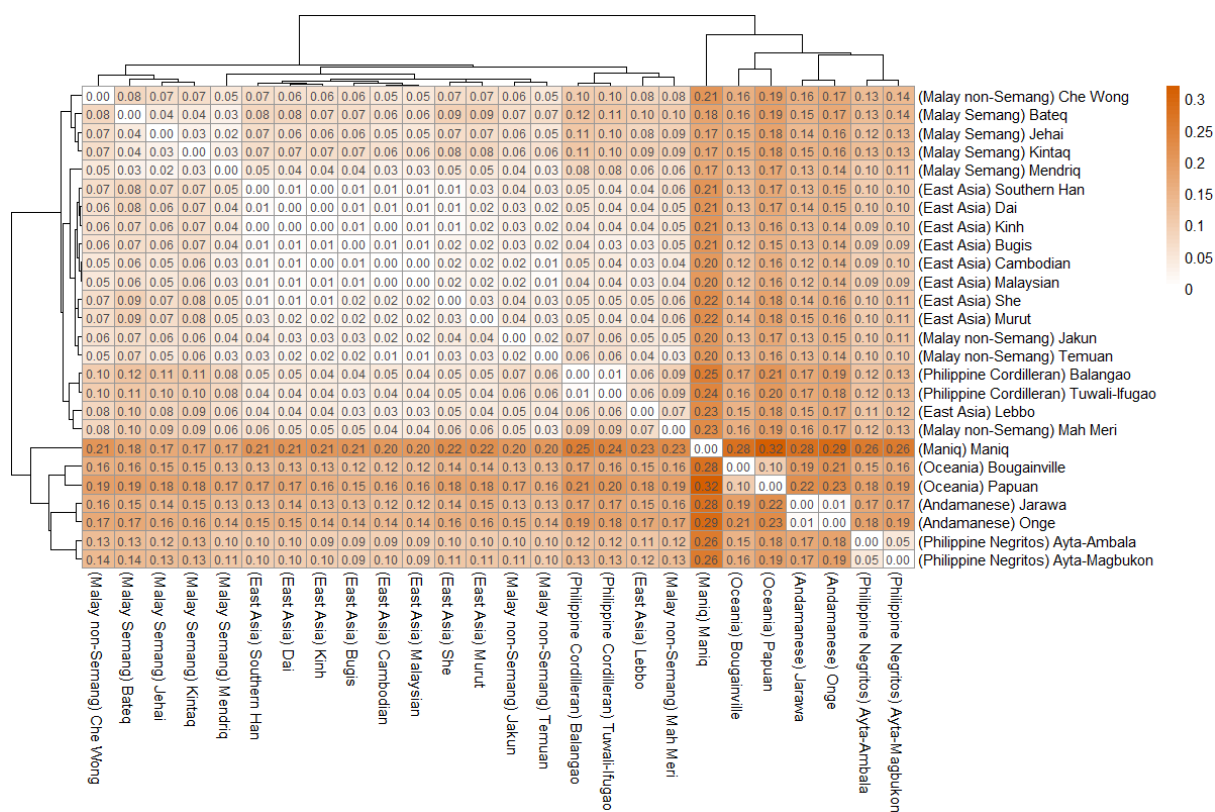
Associate editor: David Enard

Supplementary Information

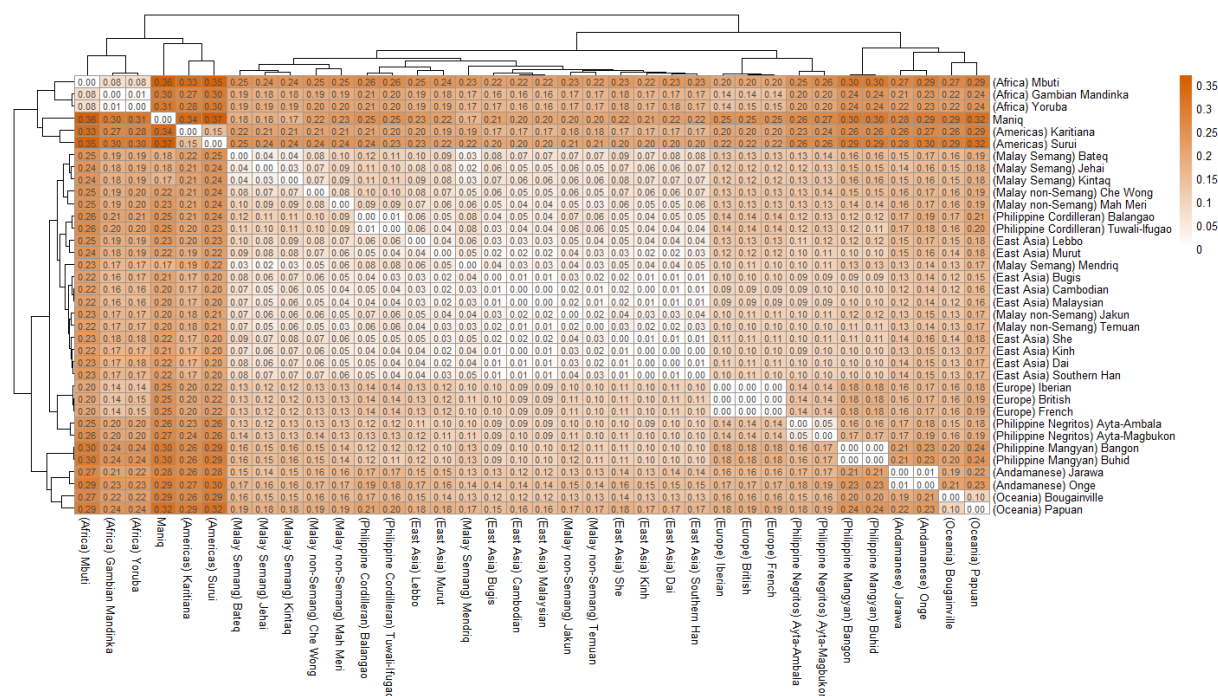
Supplementary Figure 1: F_{ST} analysis with the populations used in the PCA World (main text Figure 1B). Number of SNPs used = 195,542.



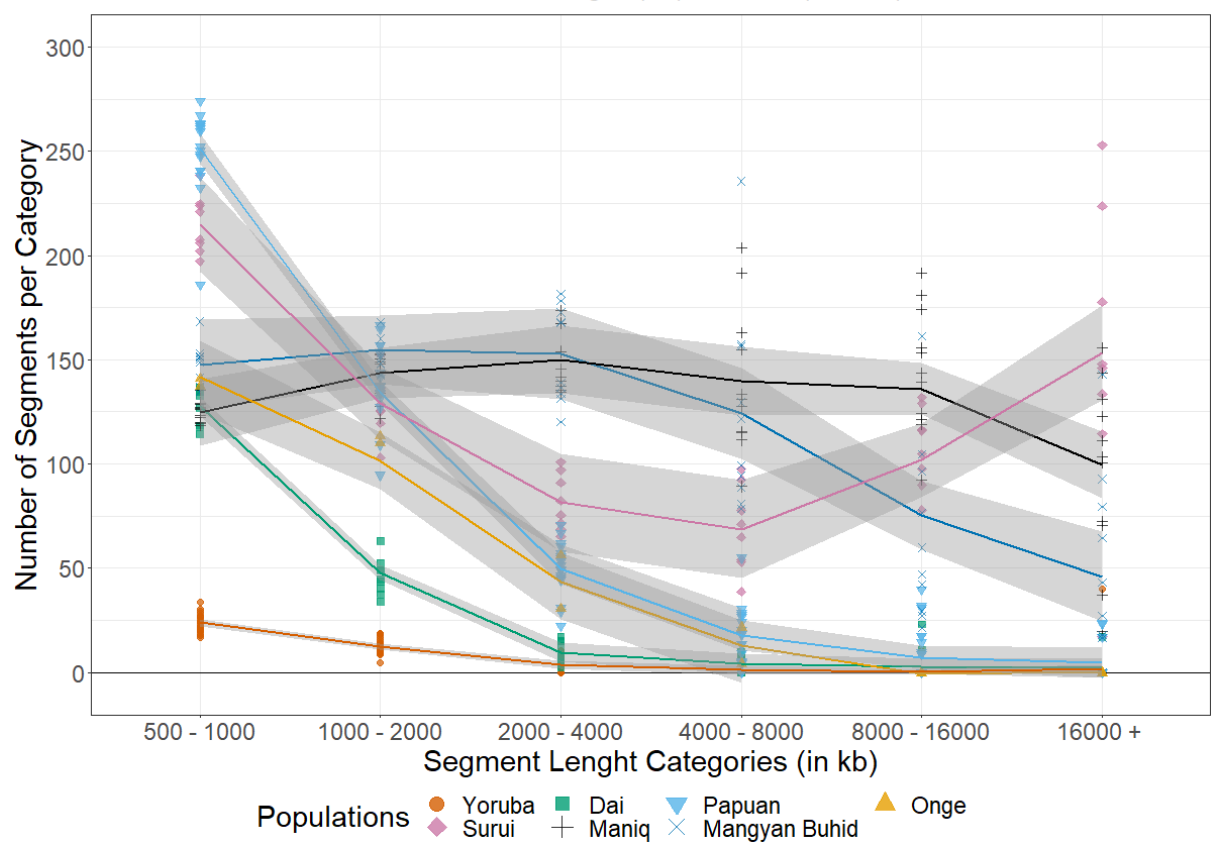
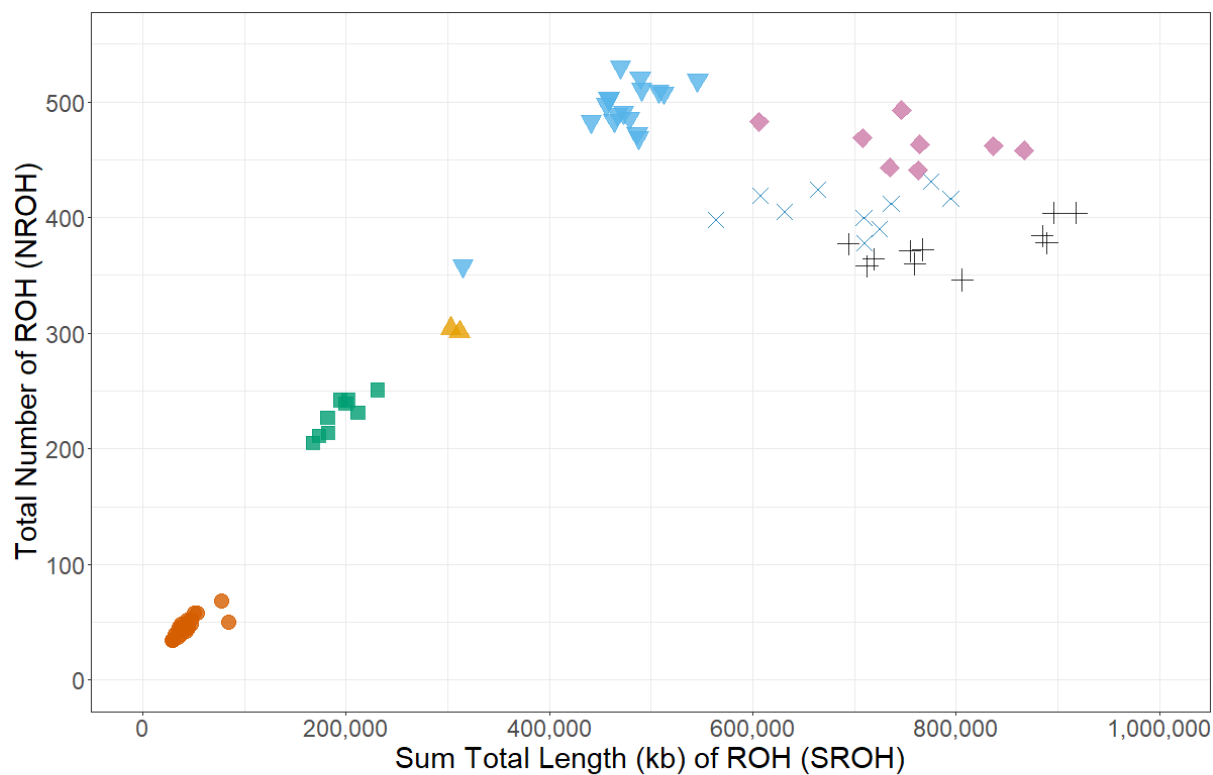
1C). Number of SNPs used = 170,513.



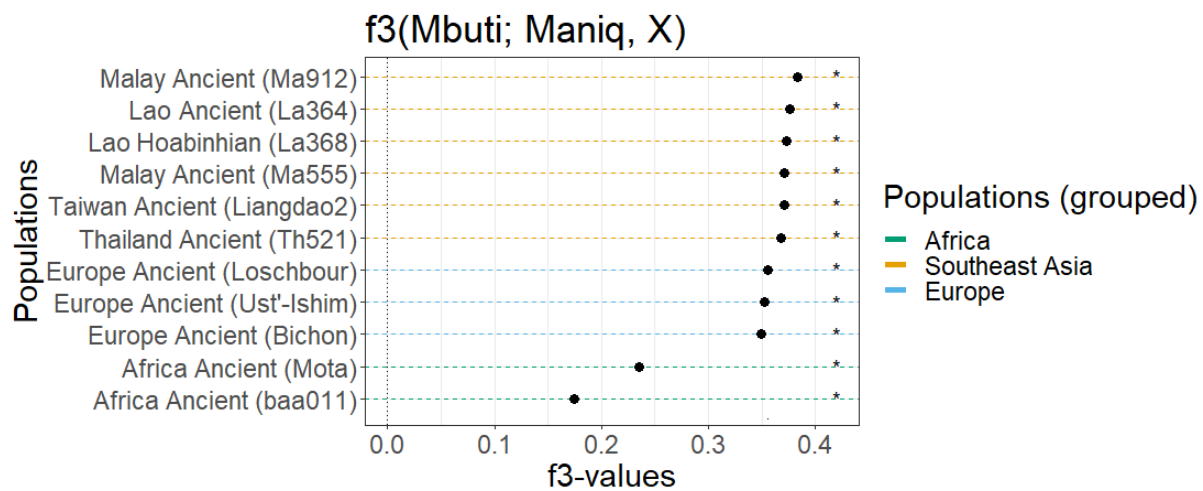
of Homozygosity analysis. Number of SNPs used = 187,955.



Supplementary Figure 4A & B: Extended ROH plots using additional populations.

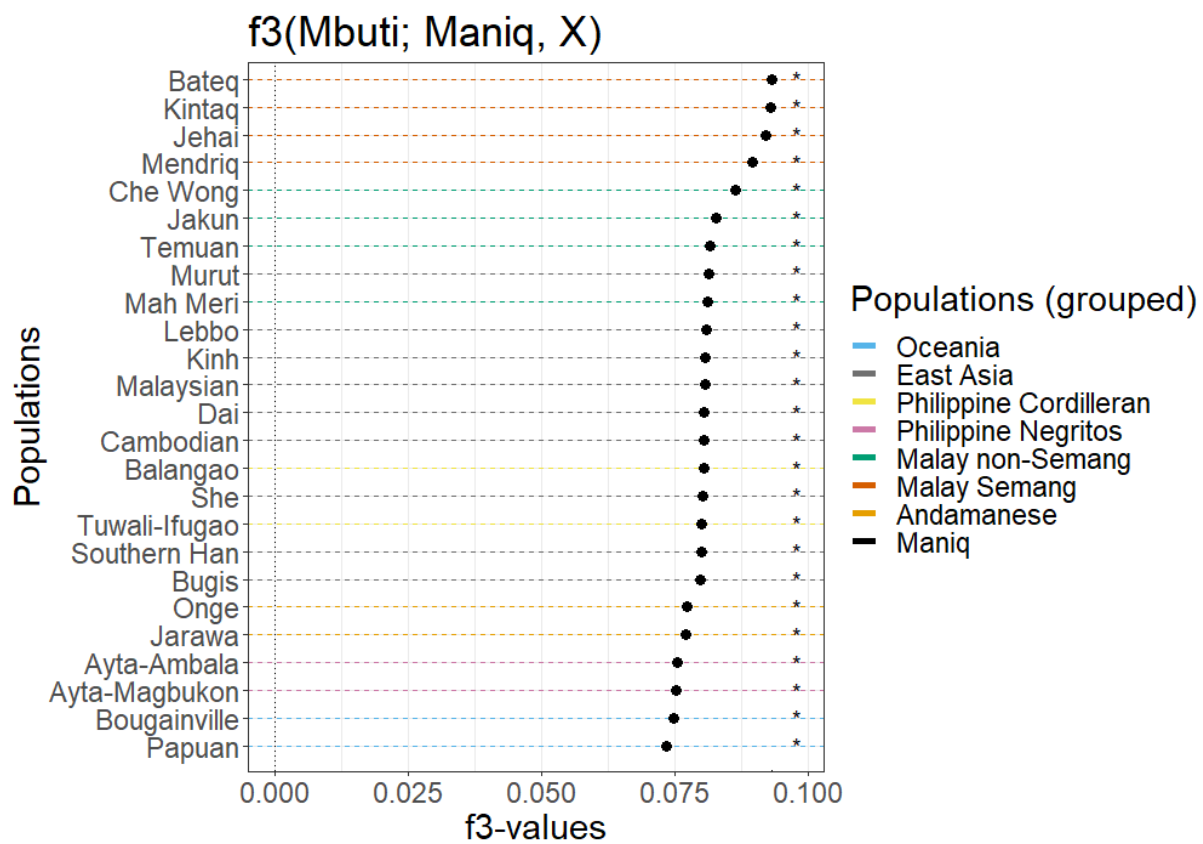


Supplementary Figure 5: F3-tests with ancient populations.

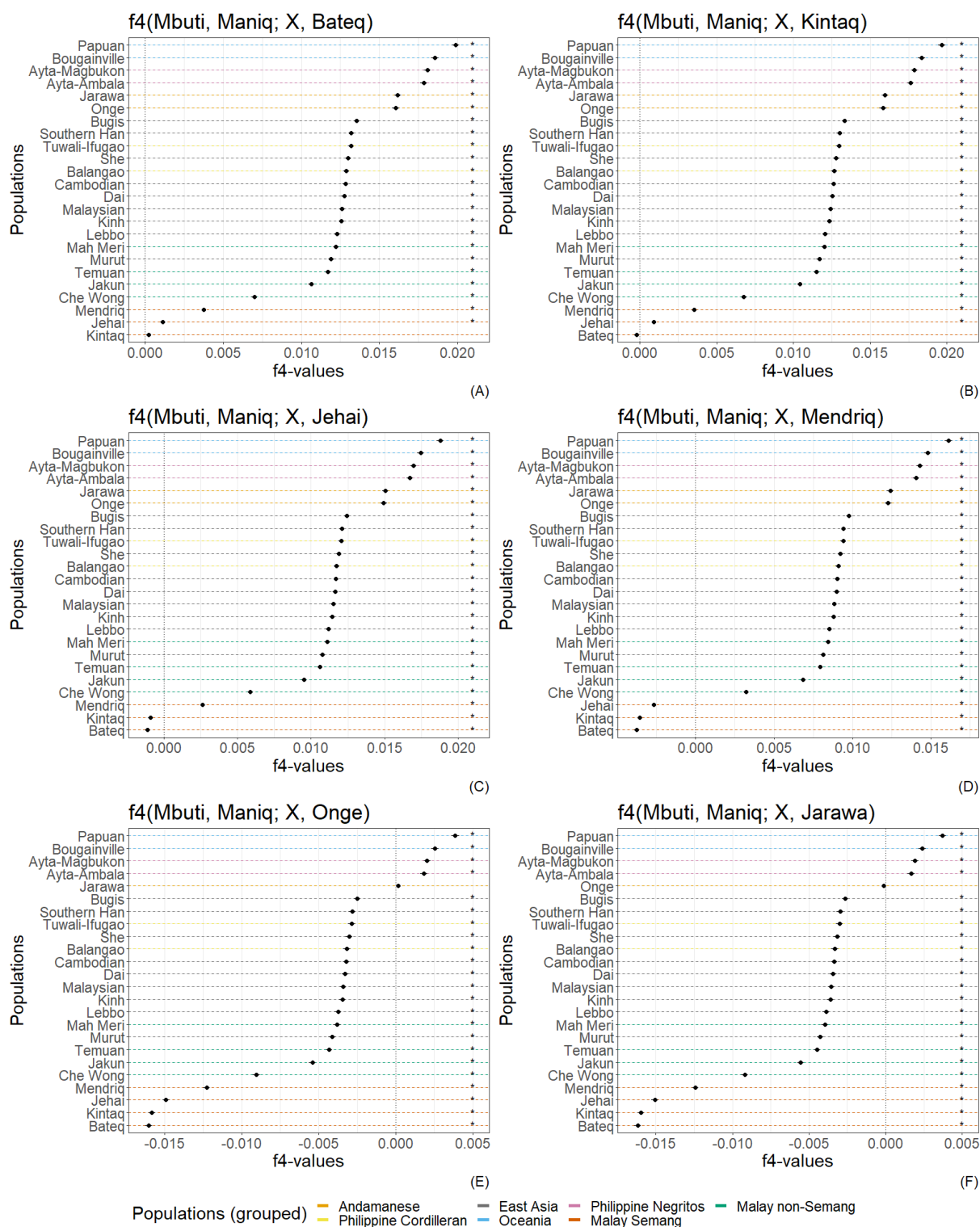


For number of SNPs used see Supplementary Table 3.

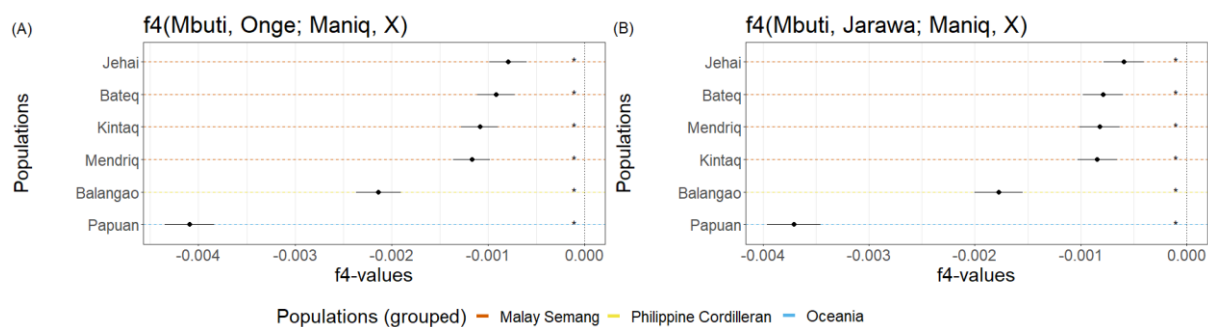
Supplementary Figure 6: F3-tests with Asian populations.



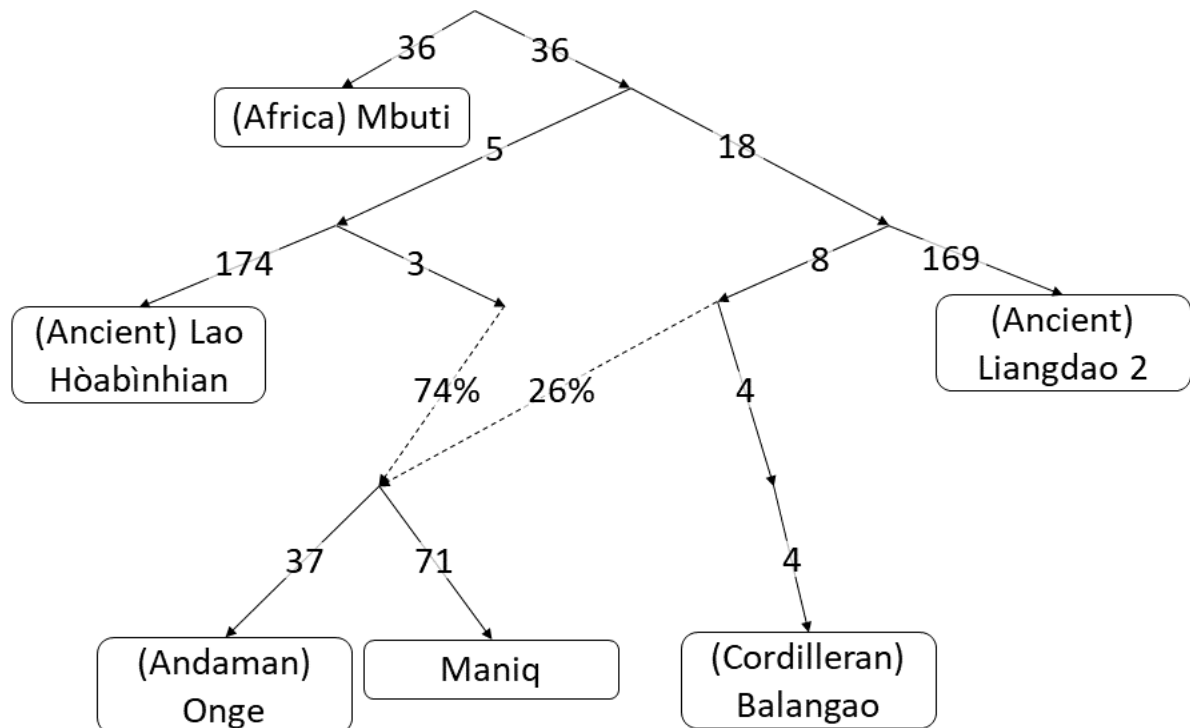
Supplementary Figure 7A-F: F4 statistics for (Mbuti, Maniq, X, Malay or Andamanese Negritos).



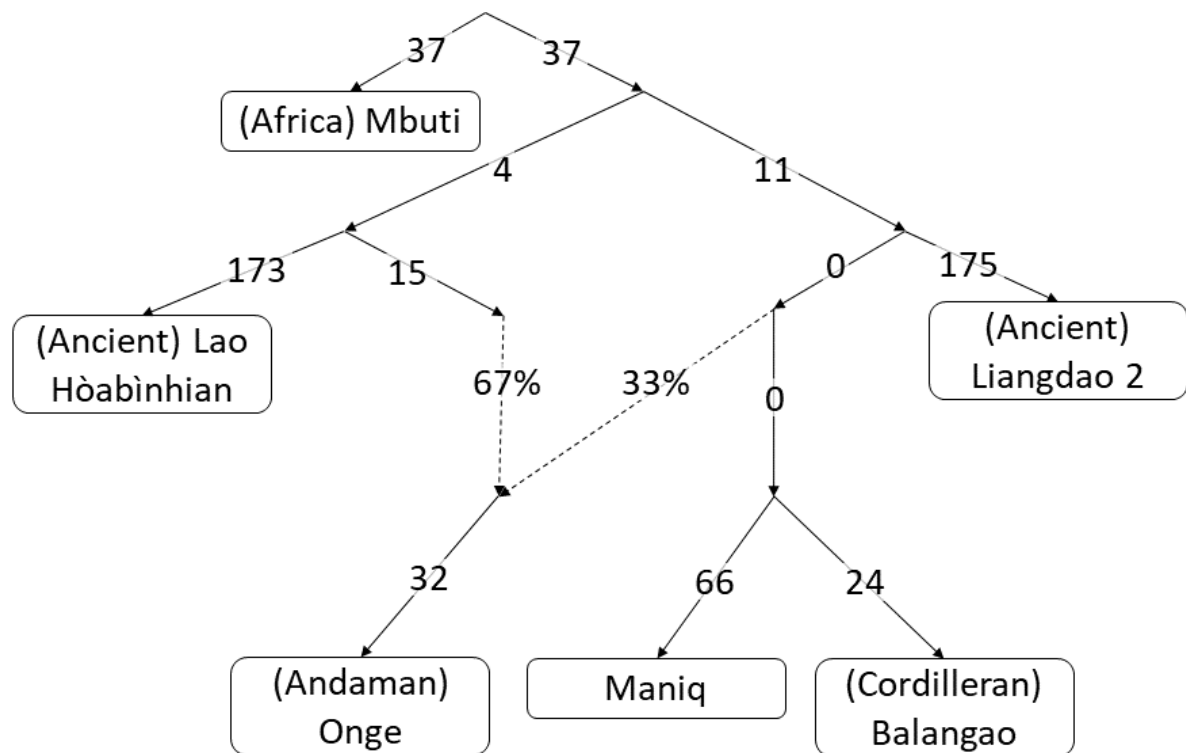
Supplementary Figure 8: Results of f_4 statistics including standard errors for (A): $f_4(\text{Mbuti, Onge; Maniq, X})$ and (B): $f_4(\text{Mbuti, Jarawa; Maniq, X})$ sorted from highest to lowest f_4 values. A line marked with an asterisk indicates an absolute Z value of greater than 3. Number of SNPs used = 319,386.



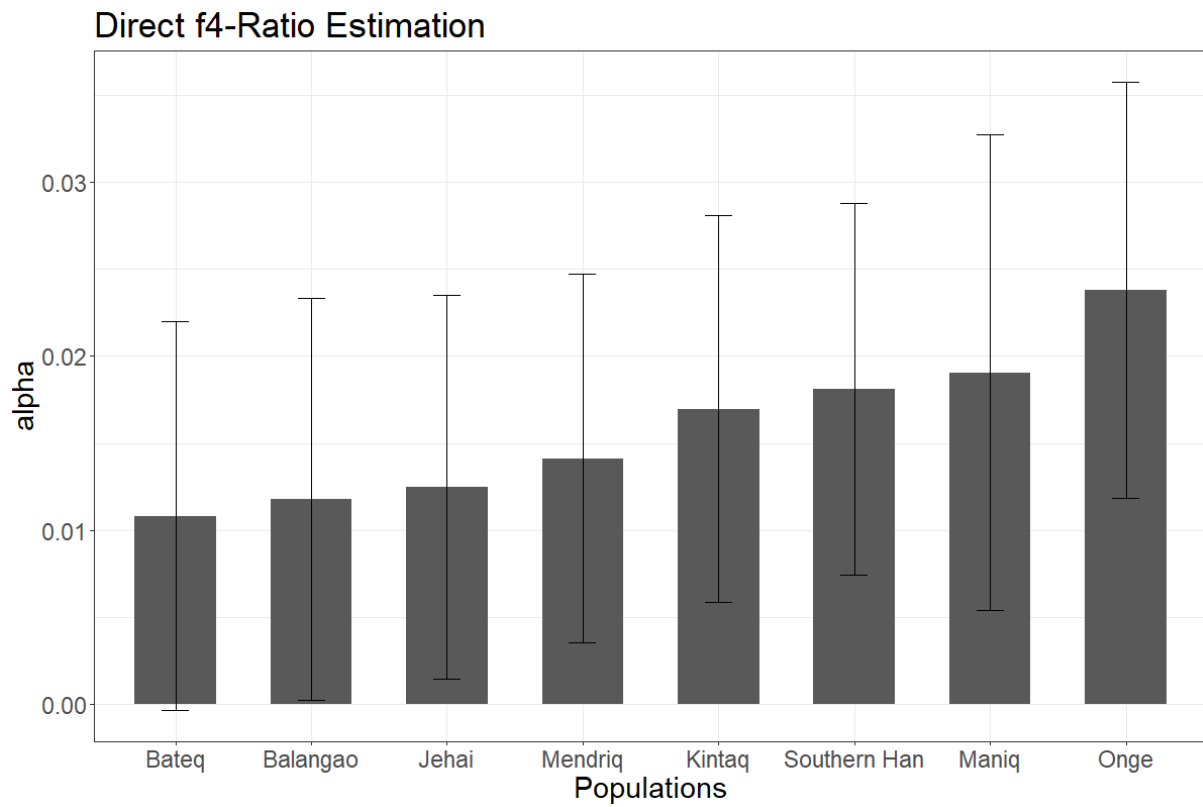
Supplementary Figure 9: Maniq with sole Hòabìnhian-related ancestry is rejected (worst fitting Z of 8.3743). Number of SNPs used = 67,094.



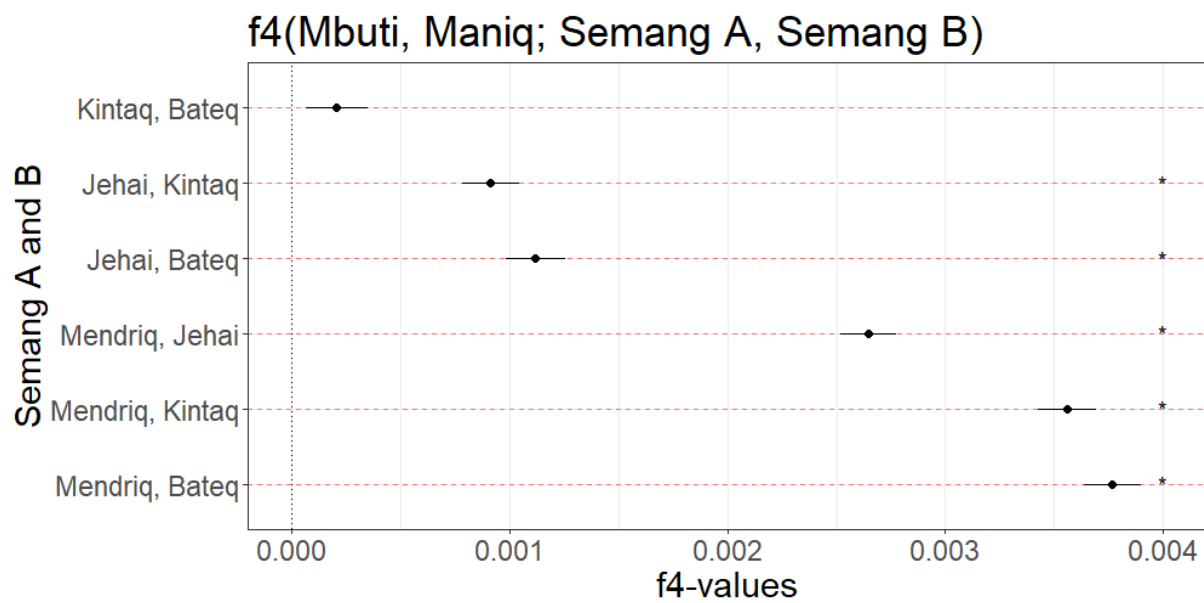
Supplementary Figure 10: Maniq with sole East Asian-related ancestry is rejected (worst fitting Z of 16.0049). Number of SNPs used = 67,094.



Supplementary Figure 11: Neanderthal Introgression f4-Ratio Estimation (Direct Estimation) for selected populations. Number of SNPs used = 36,403 (transversions only).



Supplementary Figure 12: F4 statistics for (Mbuti, Maniq; Semang A, Semang B).



Supplementary Table 1A: Number of samples from populations by analysis.

Populations	PCA	FST	Tree mix	Admix- ture	ROH	Admixtools (Present)	Admixtools (Archaic)	MAL DER
Gambian Mandinka	15	15	10	0	0	10	0	0
Dai	10	10	10	10	9	10	0	15
Southern Han	10	10	10	10	0	10	0	54
British	15	15	10	0	0	10	0	0
Iberian	15	15	10	0	0	10	0	0
Mbuti	13	13	10	0	0	10	13	0
Yoruba	15	15	10	0	22	10	0	0
She	10	10	10	10	0	10	0	10
French	15	15	10	0	0	10	0	0
Cambodian	9	9	9	9	0	9	0	9
Bougainville	10	10	10	10	0	10	0	11
Bugis	10	10	10	10	0	10	0	27
Lebbo	10	10	10	10	0	10	0	18
Murut	10	10	10	10	0	10	0	23
Bateq	10	10	10	10	0	10	12	0
Che Wong	10	10	10	10	0	10	0	0
Jakun	10	10	10	10	0	10	0	11
Jehai	10	10	10	10	0	10	22	0
Kintaq	10	10	10	10	0	10	15	0
Mah Meri	10	10	10	10	0	10	0	24
Malaysian	10	10	10	10	0	10	0	25
Mendriq	10	10	10	10	0	10	13	0
Temuan	10	10	10	10	0	10	0	13
Maniq	11	11	11	11	11	11	11	11
Kinh	10	10	10	10	0	10	0	16
Papuan	10	10	10	10	17	10	0	15
Balangao	10	10	10	10	0	10	10	10
Tuwali- Ifugao	10	10	10	10	0	10	0	10
Ayta- Ambala	8	8	8	8	0	8	0	8
Ayta- Magbukon	9	9	9	9	0	9	0	9
Jarawa	10	10	10	10	0	10	0	19
Onge	10	10	10	10	2	10	7	17
Surui	0	0	0	0	8	0	0	0
Karitiana	0	0	0	0	0	0	0	0
Mangyan Buhid	0	0	0	0	10	0	0	0
Lao	0	0	0	0	0	0	1	0

Hòabìnhian								
Liangdao2	0	0	0	0	0	0	1	0
Denisovan	0	0	0	0	0	0	1	0
Neanderthal (Altai)	0	0	0	0	0	0	1	0
Neanderthal (Vindija)	0	0	0	0	0	0	1	0

Supplementary Table 1B: Number of SNPs by analysis, with reference to the figures and tables, as well as additional information.

Analysis	Number of SNPs	References	Info
<i>PCA</i>			
PCA World	320032	Figure 1B	
PCA Asia	319845	Figure 1C	
<i>FST</i>			
FST World	195542	Supplementary Figure 1	pruned
FST Asia	170513	Supplementary Figure 2	pruned
FST World Extended	187955	Supplementary Figure 3	pruned
Treemix	190283	Figure 6	pruned
Admixture	170513	Figure 2	pruned
ROH, Inbreeding Coefficient	779956	Figure 3 A&B, Supplementary Table 2	pruned
MALDER	320032	Supplementary Table 6	
<i>Admixtools (Present)</i>			
f3-statistics	319386	Supplementary Figure 6	maxmiss = 0 (equivalent to allsnps = YES)
f4-statistics	319386	Figure 4 D-F, Supplementary Figure 7A-F, Supplementary Figure 8, Supplementary Figure 12, Supplementary Table 4	maxmiss = 0 (equivalent to allsnps = YES)
qpAdm	319386	Figure 5, Supplementary Table 5	maxmiss = 0 (equivalent to allsnps = YES)
<i>Admixtools (Archaic)</i>			
f4-statistics	16428	Figure 4 A-C	maxmiss = 0 (equivalent to allsnps = YES)
qpgraph	67094	Figure 7, Supplementary Figure 9 and 10	maxmiss = 0 (equivalent to allsnps = YES)
qpf4ratio, f4-statistics	36403	Supplementary Table 7 to 11, Supplementary Figure 11	maxmiss = 0 (equivalent to allsnps = YES)

Supplementary Table 2: Inbreeding Coefficient for various populations.

Population (Grouping)	N	Mean F	Minimum F	Median F	Maximum F
Maniq	11	0.379	0.332	0.375	0.412
Surui (Americas)	8	0.339	0.299	0.340	0.383
Mangyan Buhid (Philippines)	10	0.308	0.265	0.319	0.340
Karitiana (Americas)	12	0.302	0.254	0.296	0.367
Yoruba (Africa)	22	-0.292	-0.299	-0.292	-0.284
Mangyan Bangon (Philippines)	10	0.292	0.240	0.299	0.334
Papuan (Oceania)	17	0.273	0.179	0.278	0.305
Onge (Andamanese)	2	0.219	0.213	0.219	0.224
Dai (East Asia)	9	0.108	0.093	0.112	0.122

Supplementary Table 3: Detailed results for F3-tests with ancient populations using f3(Mbuti; Ma-
niq, X).

Target Population (Group)	Target Population	f3	SE	Z	Number of SNPs
Southeast Asia	Malay Ancient (Ma912)	0.38318	0.00426	89.889	581
Southeast Asia	Lao Ancient (La364)	0.37670	0.00413	91.186	537
Southeast Asia	Lao Hòabinhian (La368)	0.37260	0.00533	69.865	237
Southeast Asia	Malay Ancient (Ma555)	0.37127	0.00480	77.332	315
Southeast Asia	Taiwan Ancient (Liang-dao2)	0.37122	0.00388	95.793	623
Southeast Asia	Thailand Ancient (Th521)	0.36769	0.00517	71.093	246
Europe	Europe Ancient (Loschbour)	0.35594	0.00403	88.252	670
Europe	Europe Ancient (Ust'-Ishim)	0.35209	0.00424	83.016	695
Europe	Europe Ancient (Bichon)	0.34927	0.00405	86.260	651
Africa	Africa Ancient (Mota)	0.23498	0.00309	76.080	703
Africa	Africa Ancient (baa011)	0.17454	0.00275	63.584	721

Supplementary Table 4: Exact Results of f4(Mbuti, Balangao; Onge, X).

Target Population (Group)	Target Population	f4 (est)	SE	Z
East Asia	Murut	0.01897	0.00019	98.017
East Asia	Lebbo	0.01671	0.00020	85.146
East Asia	Dai	0.01668	0.00019	87.985
East Asia	Bugis	0.01665	0.00019	89.784
East Asia	She	0.01651	0.00019	85.572
East Asia	Southern Han	0.01629	0.00019	86.063
East Asia	Kinh	0.01585	0.00019	84.489
East Asia	Malaysian	0.01430	0.00018	78.819
East Asia	Cambodian	0.01299	0.00018	71.028
Malay non-Semang	Temuan	0.01216	0.00018	66.301
Malay non-Semang	Mah Meri	0.01190	0.00020	60.613
Malay non-Semang	Jakun	0.01175	0.00018	63.575
Malay non-Semang	Che Wong	0.01006	0.00019	51.624
Malay Semang	Mendriq	0.00860	0.00018	47.143
Malay Semang	Jehai	0.00731	0.00019	38.707
Malay Semang	Kintaq	0.00648	0.00019	34.122
Philippine Negritos	Ayta-Ambala	0.00631	0.00020	30.986
Malay Semang	Bateq	0.00619	0.00019	32.424
Maniq	Maniq	0.00533	0.00024	22.184
Philippine Negritos	Ayta-Magbukon	0.00473	0.00020	23.160
Oceania	Bougainville	0.00156	0.00021	7.402
Oceania	Papuan	-0.00263	0.00022	-11.861

Supplementary Table 5: Detailed qpAdm Results for the Maniq and Malay Semang.

Popula- tion	Balanga o (Weight Estima- tion)	Balanga o (Min. Weight Estima- tion)	Balanga o (Max. Weight Estima- tion)	Balanga o (Z)	Onge (Weight Estima- tion)	Onge (Min. Weight Estima- tion)	Onge (Max. Weight Estima- tion)	Onge (Z)
Maniq	0.34646	0.31739	0.37553	11.916	0.65354	0.62447	0.68261	22.478
Bateq	0.33936	0.31450	0.36421	13.652	0.66064	0.63579	0.68550	26.578
Jehai	0.41875	0.39842	0.43909	20.591	0.58125	0.56091	0.60158	28.581
Kintaq	0.34369	0.32156	0.36582	15.529	0.65631	0.63418	0.67844	29.654
Mendriq	0.47802	0.45824	0.49781	24.164	0.52198	0.50219	0.54176	26.386

Supplementary Table 6: MALDER Results with Maniq as Admixed Population. MALDER gives only one time estimation for all possible combinations. Time = 28.378 generations, Time +/- = 9.391, Time Z-value = 3.022; when assuming 25 years as generation time then this equates to estimated 709.453 years (minimum = 474.687 years, maximum = 944.218 years).

Reference Population 1	Reference Population 2	Amplitude	Amplitude +/-	Amplitude Z-values
Papuan	Mah-Meri	0.0000782	0.0000219	3.572
She	Papuan	0.0000616	0.0000142	4.348
Bougainville	Mah-Meri	0.0000600	0.0000190	3.154
Cambodian	Papuan	0.0000534	0.0000126	4.223
She	Bougainville	0.0000485	0.0000134	3.628
Papuan	Murut	0.0000484	0.0000106	4.578
Bougainville	Kinh	0.0000393	0.0000127	3.087
Temuan	Onge	0.0000307	0.0000082	3.748
Bugis	Jakun	0.0000123	0.0000047	2.639
Southern Han	Jakun	0.0000122	0.0000057	2.140

Supplementary Table 7: Neanderthal Introgression f4-Ratio Estimation (Direct Estimation).

f4(Neanderthal (Altai), Chimp; X, Yoruba)/f4(Neanderthal (Altai), Chimp; Neanderthal (Vindija), Yoruba)

Target Population (Group)	Target Population	Alpha	SE	Z
Malay Semang	Bateq	0.01082	0.01116	0.969
Philippine Cordilleran	Tuwali-Ifugao	0.01131	0.01138	0.994
Philippine Cordilleran	Balangao	0.01179	0.01155	1.02
Malay Semang	Jehai	0.01249	0.01102	1.134
Malay Semang	Mendriq	0.01413	0.01061	1.332
East Asia	Cambodian	0.01449	0.01037	1.398
East Asia	Lebbo	0.01607	0.01116	1.44
East Asia	Bugis	0.01647	0.01049	1.571
East Asia	Murut	0.01663	0.01094	1.52
Malay Semang	Kintaq	0.01697	0.01111	1.528
East Asia	She	0.01711	0.01086	1.576
East Asia	Southern Han	0.01811	0.0107	1.693
Malay non-Semang	Jakun	0.01853	0.01066	1.739
East Asia	Dai	0.01858	0.01072	1.733
Maniq	Maniq	0.01906	0.01365	1.397
East Asia	Malaysian	0.01931	0.01043	1.852
Malay non-Semang	Mah Meri	0.01986	0.01109	1.791
Malay non-Semang	Temuan	0.02183	0.01043	2.093
East Asia	Kinh	0.02293	0.01055	2.173
Andamanese	Onge	0.0238	0.01194	1.993
Malay non-Semang	Che Wong	0.02394	0.01122	2.134
Andamanese	Jarawa	0.02396	0.01164	2.059
Philippine Negritos	Ayta-Magbukon	0.03016	0.01127	2.676
Oceania	Bougainville	0.03279	0.01193	2.749
Philippine Negritos	Ayta-Ambala	0.03582	0.01129	3.174
Oceania	Papuan	0.04257	0.0123	3.462

Supplementary Table 8: Neanderthal Introgression f4-test. f4(Chimp, Neanderthal; Mbuti, X)

Target Population (Group)	Target Population	f4 (est)	SE	Z	p
Philippine Cordilleran	Tuwali-Ifugao	0.000048	0.000764	0.063	0.95
Malay Semang	Bateq	0.000087	0.000749	0.116	0.908
Philippine Cordilleran	Balangao	0.000126	0.000771	0.163	0.871
Malay Semang	Jehai	0.000262	0.000744	0.352	0.725
Malay Semang	Mendriq	0.000358	0.000722	0.495	0.62
East Asia	Cambodian	0.00036	0.000708	0.509	0.611
East Asia	Bugis	0.000504	0.00071	0.71	0.478
East Asia	Lebbo	0.000508	0.000751	0.676	0.499
East Asia	She	0.000551	0.000731	0.753	0.452
East Asia	Murut	0.000561	0.00074	0.758	0.448
Malay Semang	Kintaq	0.000583	0.000747	0.781	0.435
Malay non-Semang	Jakun	0.000638	0.000726	0.879	0.38
East Asia	Southern Han	0.000688	0.000722	0.953	0.341
East Asia	Malaysian	0.000705	0.000706	0.998	0.318
East Asia	Dai	0.000713	0.000729	0.979	0.328
Maniq	Maniq	0.000766	0.000889	0.862	0.389
Malay non-Semang	Mah Meri	0.000813	0.000745	1.091	0.275
Malay non-Semang	Temuan	0.000962	0.000709	1.357	0.175
East Asia	Kinh	0.001048	0.000717	1.461	0.144
Andamanese	Onge	0.001083	0.000792	1.367	0.171
Andamanese	Jarawa	0.001087	0.000782	1.39	0.164
Malay non-Semang	Che Wong	0.001116	0.000752	1.484	0.138
Philippine Negritos	Ayta-Magbukon	0.001611	0.000759	2.122	0.034
Oceania	Bougainville	0.001795	0.00079	2.272	0.023
Philippine Negritos	Ayta-Ambala	0.002133	0.00076	2.807	0.005
Oceania	Papuan	0.002552	0.000815	3.129	0.002

Supplementary Table 9: Denisovan Introgression f4-Ratio Estimation. f4(Chimp, Neanderthal; Southern Han, X)/f4(Chimp, Neanderthal; Southern Han, Denisovan)

Target Population (Group)	Target Population	Alpha	SE	Z
Philippine Cordilleran	Tuwali-Ifugao	-0.02591	0.0227	-1.141
Malay Semang	Bateq	-0.02421	0.02929	-0.827
Philippine Cordilleran	Balangao	-0.02278	0.0236	-0.965
Malay Semang	Jehai	-0.01966	0.02752	-0.714
East Asia	Cambodian	-0.01105	0.01964	-0.562
Malay Semang	Mendriq	-0.0094	0.02496	-0.377
East Asia	Lebbo	-0.00647	0.02329	-0.278
East Asia	Bugis	-0.00629	0.01966	-0.32
East Asia	She	-0.00345	0.01876	-0.184
Malay Semang	Kintaq	-0.0031	0.02798	-0.111
East Asia	Murut	-0.00242	0.02133	-0.113
Maniq	Maniq	0.0028	0.04069	0.069
Malay non-Semang	Jakun	0.00301	0.02314	0.13
East Asia	Dai	0.00354	0.01859	0.191
East Asia	Malaysian	0.00429	0.01963	0.218
Malay non-Semang	Mah Meri	0.00862	0.02535	0.34
Malay non-Semang	Temuan	0.01498	0.02159	0.694
East Asia	Kinh	0.01824	0.01773	1.028
Andamanese	Jarawa	0.0186	0.03388	0.549
Andamanese	Onge	0.01903	0.0353	0.539
Malay non-Semang	Che Wong	0.0229	0.02655	0.862
Philippine Negritos	Ayta-Magbukon	0.04455	0.03019	1.475
Oceania	Bougainville	0.05573	0.03327	1.675
Philippine Negritos	Ayta-Ambala	0.06444	0.03023	2.132
Oceania	Papuan	0.08359	0.0371	2.253

Supplementary Table 10: Denisovan Introgression f4-test. f4(Chimp, Denisovan; French, X)

Target Population (Group)	Target Population	f4 (est)	SE	Z	p
Philippine Cordilleran	Tuwali-Ifugao	-0.001281	0.000578	-2.218	0.027
East Asia	Southern Han	-0.001199	0.000532	-2.252	0.024
East Asia	Cambodian	-0.001192	0.000499	-2.389	0.017
Malay non-Semang	Jakun	-0.001168	0.000526	-2.221	0.026
Malay Semang	Jehai	-0.001143	0.000553	-2.065	0.039
Malay Semang	Bateq	-0.001101	0.000558	-1.973	0.048
East Asia	Lebbo	-0.001085	0.000559	-1.94	0.052
East Asia	Malaysian	-0.001053	0.000498	-2.112	0.035
Maniq	Maniq	-0.001017	0.000727	-1.399	0.162
East Asia	Bugis	-0.00094	0.00052	-1.809	0.07
Malay Semang	Mendriq	-0.000906	0.000519	-1.745	0.081
Philippine Cordilleran	Balangao	-0.000894	0.000587	-1.525	0.127
East Asia	She	-0.000845	0.000531	-1.592	0.111
Malay non-Semang	Mah Meri	-0.000824	0.000564	-1.46	0.144
Malay Semang	Kintaq	-0.0008	0.000551	-1.451	0.147
East Asia	Murut	-0.000791	0.000541	-1.463	0.143
East Asia	Dai	-0.000702	0.00053	-1.325	0.185
Malay non-Semang	Temuan	-0.000631	0.000507	-1.246	0.213
Malay non-Semang	Che Wong	-0.00041	0.000561	-0.731	0.465
East Asia	Kinh	-0.000358	0.000517	-0.693	0.489
Andamanese	Onge	-0.0003	0.000614	-0.488	0.625
Andamanese	Jarawa	-0.000297	0.000601	-0.495	0.621
Oceania	Bougainville	0.001119	0.000608	1.84	0.066
Philippine Negritos	Ayta-Magbukon	0.00117	0.000582	2.01	0.044
Philippine Negritos	Ayta-Ambala	0.00131	0.000584	2.241	0.025
Oceania	Papuan	0.001775	0.000655	2.709	0.007

Chapter II

Natural Selection and Adaptive Traits in the Maniq, a Nomadic Hunter-Gatherer Society from Mainland Southeast Asia

Authors: Tobias Herzog, Maximilian Larena, Wibhu Kutanan, Helmut Lukas, Martin Fieder, Helmut Schaschl

Reference: Herzog, T., Larena, M., Kutanan, W., Lukas, H., Fieder, M., & Schaschl, H. (2025). Natural selection and adaptive traits in the Maniq, a nomadic hunter-gatherer society from Mainland Southeast Asia. *Scientific Reports*, 15(1), 4809. <https://doi.org/10.1038/s41598-024-83657-0>

Status: Published in Scientific Reports.

My contribution I conceived and designed this study. I worked on the pipelines to create various analysis datasets. I worked on the analysis pipelines for all results. I created all figures and tables (except Table 3 and Supplementary Figure 1). I drafted the results, discussion and contributed to the draft of all other sections. I edited the final manuscript. Afterwards I contributed to drafts of the revision and edited the final version.



OPEN Natural selection and adaptive traits in the Maniq, a nomadic hunter-gatherer society from Mainland Southeast Asia

Tobias Herzog¹✉, Maximilian Larena², Wibhu Kutanan³, Helmut Lukas⁴, Martin Fieder¹ & Helmut Schaschl¹✉

Asia is home to diverse hunter-gatherer populations characterized by significant morphological, anthropological, cultural, and linguistic diversity. Despite their importance in understanding ancestral human subsistence, little is known about the essential genetic adaptations of these groups. This study investigates the evolutionary pressures shaping the genome of the Maniq population, a nomadic hunter-gatherer group inhabiting the rainforests of southern Thailand. Using genome-wide approaches, including iHS, xp-EHH, PBE, and beta statistics, we identified signatures of positive and balancing selection. Genes under positive selection were enriched in pathways related to immunity, metabolic regulation, structural adaptation, cardiovascular performance, and neuromodulatory traits. Several genes associated with the Southeast Asian 'negrito-like' phenotype were also under positive selection. Balancing selection was primarily detected in immune-related genes, particularly within the HLA region, underscoring the critical role of genetic diversity in surviving pathogen-rich environments. Additionally, balancing selection in olfactory receptor genes highlights their importance in environmental sensing and adaptation. These results reveal the intricate interplay of positive and balancing selection in shaping the genetic landscape of the Maniq population and highlight their adaptations to the ecological and lifestyle challenges of life in the rainforest. This study contributes to our understanding of human evolutionary processes in tropical environments and hunter-gatherer societies.

Keywords Maniq people, Hunter-Gatherer, Genetic adaptation, Positive selection, Balancing selection

The indigenous people of the Thai-Malay Peninsula in Southeast Asia are known as the 'Orang Asli', who are traditionally divided into three groups: the Semang residing in the north, the Senoi in the central region, and the Malay indigenous people, sometimes referred to as 'Proto-Malay', residing in the south. The Semang are phenotypically characterized by small body size, dark skin, and tightly curled hair on average, a phenotype which is called 'negrito' and also found outside of Malaysia¹. We recently demonstrated a close genetic relationship between the Maniq, nomadic hunter-gatherers inhabiting the rainforests of southern Thailand, and the Semang². They speak an aboriginal Mon-Khmer language belonging to the Austroasiatic language family. Their society is marked by flexible social groups with regularly changing composition, a lack of reliance on stored food, a system of immediate resource consumption, absence of resource ownership, reciprocal relationships, an egalitarian ethos of sharing, and minimal social hierarchy^{3,4}. An estimated 350 individuals of the Maniq society still pursue a nomadic hunter-gatherer lifestyle. Due to the drastic deforestation in Thailand, however, some Maniq have been compelled to alter their settlement patterns, leading to increased sedentarization and departure from their traditional way of life⁵. This transition highlights the importance of understanding their genetic adaptive traits and potential adverse effects associated with transitioning to a modern lifestyle in an industrialized environment.

In this study, we analyze the signatures of natural selection within the Maniq genome, aiming to elucidate their genetic adaptation to the rainforest environment and hunter-gatherer lifestyle. This involved genome-wide

¹Department of Evolutionary Anthropology, Faculty of Life Sciences, University of Vienna, Djerassiplatz 1, Vienna 1030, Austria. ²Human Evolution, Department of Organismal Biology, Uppsala University, Norbyvägen 18C, Uppsala 75236, Sweden. ³Department of Biology, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand. ⁴Institute for Social Anthropology, Austrian Academy of Sciences, Georg-Coch-Platz 2, Vienna 1010, Austria. ✉email: tobias.goellner@univie.ac.at; helmut.schaschl@univie.ac.at

searches for signals of both positive directional selection, known as selective sweeps, and balancing selection. Positive selection is the process by which the frequency of a favorable allele increases in a population. This leaves a discernable signal in genomic regions characterized by reduced diversity, longer haplotypes, or pronounced genetic differentiation, often leading to localized genetic adaptation⁶. Previous research has identified genes linked to various traits such as skin pigmentation, nutritional adaptation, olfactory receptors, brain size, and immune function as targets of positive selection in the human genome (see review in⁶). Moreover, the relatively diminutive stature of many rainforest hunter-gatherers is thought to serve as an adaptation to the demanding rainforest habitat. It potentially confers evolutionary advantages by reducing metabolic demands in an environment with limited caloric resources, facilitating thermoregulation, enhancing mobility in dense undergrowth, and aiding in tree climbing^{7–11}. One study posited that the small adult body size phenotype across diverse hunter-gatherer populations prompts compensatory changes in cardiac pathways¹⁰. This suggests convergent adaptations influenced by similar rainforest ecology. Furthermore, signatures of positive selection were found in different African rainforest hunter-gatherers across a set of genes associated with bone synthesis, immunity, reproduction, cell signaling, and energy metabolism^{8,12,13}.

In contrast to positive directional selection, which drives the prevalence of specific alleles within a population, balancing selection leads to the maintenance of diversity at a genetic locus in the presence of selection pressure, thereby maintaining multiple variants at intermediate frequencies^{14,15}. The mechanisms facilitating this diversity include heterozygote advantage, negative frequency-dependent selection, and spatial or temporal habitat heterogeneity. Notably, genes within the human leukocyte antigen (HLA) system, also known as the major histocompatibility complex (MHC), represent well-documented targets of long-term balancing selection in many vertebrate species including humans. To detect recent signatures of positive directional selection and (long-term) balancing selection in the Maniq genome, we employed the integrated haplotype score (iHS) test¹⁶, the cross-population extended haplotype homozygosity (xp-EHH) test¹⁷, the Population Branch Excess (PBE) statistics¹⁸ and the summary statistic $\beta^{19,20}$, respectively. Our analysis revealed that positive selection was enriched in different biological pathways related to immunity, metabolic regulation, structural integrity, cardiovascular function and neuromodulatory traits, suggesting adaptations to environmental and dietary challenges, pathogen pressure, physical demands, and to the lifestyle as hunter-gatherer. Furthermore, we found signals of balancing selection in genes involved in immune response and immunological signaling pathways. Together with our previous study (see²), this study thus provides new insights into the complex evolutionary genetic history and adaptive genetics of the Maniq people, one of the last remaining primary hunter-gatherer societies in contemporary times.

Results

Genome-wide signature of positive selection

We identified signatures of positive selection in the Maniq population using the iHS, xp-EHH and PBE statistics. We combined the genome-wide rank of all three statistics for each SNP to compute a Fisher score called F_{SC} ²¹. In subsequent analyses, we defined the highest 1% of F_{SC} as outlier SNPs and targets of positive selection (Supplementary Table S1). Some of the identified candidates genes (see Supplementary Table S2 for the full list) were previously documented as being under positive selection, among other populations, in East Asians (*SLC35F3*²², *COL11A2*²³), in Orang Asli populations of the Malay Peninsula (*LRP2*, *RUNX1*, *SYN3*, *STK39*)^{24,25} and specifically in Bateq and Mendriq of the Malay Peninsula (*FAM78B*, *TADA1*, *POGK*, *DUSP27*, *CNTNAP2*, *CDH11*, *RBFOX3*)²⁵. Another group of the positively selected genes (*PTPRN2*, *NRXN3*, *AMZ1*, *FRAS1*) have been linked to the ‘negrito-like’ phenotypes observed in some Southeast Asians, characterized by short stature, dark skin and hair color, and tightly curled hair²⁶. Over-representation analysis (ORA), using WebGestalt²⁷ to perform pathway analyses using Reactome²⁸ and KEGG (Kyoto Encyclopedia of Genes and Genomes)²⁹ databases, revealed that the positively selected genes are enriched in biological pathways related to human diseases such as asthma or type I diabetes mellitus. The other enriched pathways range from cell adhesion molecules, over muscle contraction to extracellular matrix organization (Table 1 and Supplementary Table S3).

Figure 1 shows the highest fraction of putative positive selected genes that overlap with different GWAS traits, including body height, type II diabetes mellitus, different blood protein parameters, BMI-adjusted waist circumference, a specific type of scoliosis, and gut microbiome composition (Supplementary Table S4).

Pathway database	Gene set	Description	P-Value	False discovery rate (FDR)
KEGG	hsa05310	Asthma	0.000009	0.002177
KEGG	hsa04940	Type I diabetes mellitus	0.000001	0.000331
KEGG	hsa04750	Inflammatory mediator regulation of TRP channels	0.000001	0.000331
KEGG	hsa04514	Cell adhesion molecules	< 0.000001	0.000063
KEGG	hsa04921	Oxytocin signaling pathway	< 0.000001	0.000151
KEGG	hsa04261	Adrenergic signaling in cardiomyocytes	0.000002	0.000475
Reactome	R-HSA-397014	Muscle contraction	< 0.000001	0.000056
KEGG	hsa04015	Rap1 signaling pathway	0.000081	0.009808
Reactome	R-HSA-1474244	Extracellular matrix organization	0.000072	0.009445
Reactome	R-HSA-112316	Neuronal System	0.000014	0.002581

Table 1. Gene sets under positive selection in the Maniq enriched in Reactome or KEGG pathways.

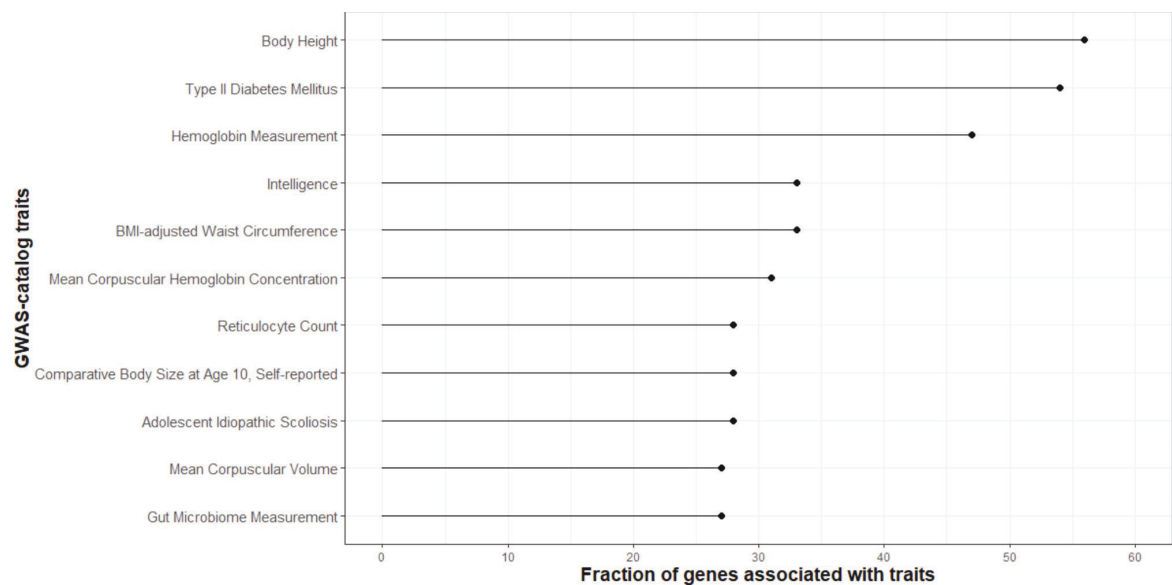


Fig. 1. Positive selection on genes that overlap with different GWAS traits.

Pathway database	Gene set	Description	P-value	False discovery rate (FDR)
Reactome	R-HSA-202430	Translocation of ZAP-70 to Immunological synapse	0.000014	0.005565
Reactome	R-HSA-202427	Phosphorylation of CD3 and TCR zeta chains	0.000023	0.006733
Reactome	R-HSA-389948	PD-1 signaling	0.000026	0.006733
KEGG	hsa05330	Allograft rejection	0.000002	0.00238
KEGG	hsa05332	Graft-versus-host disease	0.000004	0.00238
KEGG	hsa04940	Type I diabetes mellitus	0.000004	0.00238
KEGG	hsa05320	Autoimmune thyroid disease	0.000009	0.004401
KEGG	hsa05416	Viral myocarditis	0.000024	0.006733
Reactome	R-HSA-877300	Interferon gamma signaling	0.000004	0.00238

Table 2. Gene sets under balancing selection in the Maniq enriched in Reactome or KEGG pathways.

Genome-wide signature of balancing selection

A signature of balancing selection was detected in 41 genes (Supplementary Table S5). Notably, the HLA region stands out with the highest $\beta(1)$ values such as the genes *HLA-C* (rs1130838), *HLA-DQB1* (rs1049056), *HLA-DPA1* (rs1126769), and *HLA-DPB1* (rs1126719). The Reactome and KEGG pathway analysis revealed that only immunity-related or immunological-related biological pathways were enriched (Table 2 and Supplementary Table S6).

Several other genes involved in innate immunity, such as *IL1RL1*, *GNLY*, *TNFRSF10D*, *TRIM22*, *LILRB1*, *CCCL4*, and *DEFB127*, show a signature of balancing selection (Table 3).

In addition, the *MUC19*, which shows a signature of balancing selection, encodes mucin, a major component of mucus, which is part of the innate immune responses. Genes with a function similar to *OR51B2*, *OR51Q1*, and *OR1N2* that code for olfactory receptor molecules also show evidence of balancing selection.

Discussion

Our analysis of the Maniq autosomal genome provides valuable insights into the interplay between positive and balancing selection in shaping their unique genetic makeup. By identifying signals of positive selection and balancing selection, our study elucidates the genetic underpinnings of traits that may have evolved in response to their distinct hunter-gatherer lifestyle and rainforest environment. One of the key findings is the identification of genes under positive selection that are enriched in biological pathways indicative of adaptations related to immunity and metabolism, structural and locomotor adaptations, adaptations related to cardiovascular functions, and neuromodulatory traits (Table 1, Supplementary Table S3). For instance, the enrichment of positive selected genes in the type I diabetes mellitus pathway suggests a connection between immunity and metabolic regulation, likely driven by the ecological and dietary challenges of the rainforest. The enrichment of genes in the oxytocin signaling pathway may reflect adaptations to physiological processes

Chr	Gene	Gene description
2q12.1	<i>IL1RL1</i>	Interleukin 1 receptor like 1
2p11.2	<i>GNLY</i>	Granulysin
6p211.33	<i>HLA-C</i>	Major histocompatibility complex, class I, C
6p211.32	<i>HLA-DPB1, -DQB1, -DPA1</i>	Major histocompatibility complex, class II
8p21.3	<i>TNFRSF10D</i>	Tumor necrosis factor receptor superfamily member 10D
11p15.4	<i>TRIM22</i>	Tripartite Motif Containing 22
12q12	<i>MUC19</i>	Mucin 19, oligomeric
12p13.31	<i>CLEC4D</i>	C-type lectin domain family 4 member D
14q11.2	<i>TRAV14DV4</i>	T cell receptor alpha variable 14/delta variable 4
19q13.42	<i>LILRB1</i>	Leukocyte immunoglobulin like receptor B1
17q12	<i>CCL4</i>	C-C motif chemokine ligand 4
20p13	<i>DEFB127</i>	Defensin beta 127

Table 3. Immune genes under balancing selection in the Maniq.

such as labour (uterine contractility) and lactation, but could also be related to the Maniq's cooperative social structure, stress modulation, and parenting behaviors, which would be critical in their egalitarian and group-dependent hunter-gatherer lifestyle. The adrenergic signaling in cardiomyocytes pathway highlights potential cardiovascular adaptations, such as enhanced resilience to stress and improved physical performance, which are vital for survival in resource-scarce and physically demanding rainforest habitats. Similarly, the extracellular matrix organization pathway, enriched with collagen-related genes (*COL11A2*, *COL5A1*, *COL4A2*, *COL18A1*), highlights adaptations for structural support and mobility, potentially critical adaptations for life in the humid, dense rainforest. Notably, *COL11A2* has been repeatedly identified as a target of positive selection in other populations, including East Asians, the Orang Asli²⁶, indigenous Mesoamericans³³, and the Luhya in Webuye, Kenya²³, suggesting its importance in fibrillogenesis and structural adaptation.

We also identified several genes, including *SLC35F3*, *LRP2*, *RUNX1*, *SYN3*, *STK39*, *FAM78B*, *TADA1*, *POGK*, *DUSP27*, *CNTNAP2*, *CDH11*, *RBFOX3*, *PTPRN2*, *NRXN3*, *AMZ1*, and *FRAS1*, which have previously been identified as undergoing positive selection in various Asian populations (Supplementary Table S2). Specifically, several of the genes mentioned above (*PTPRN2*, *NRXN3*, *AMZ1*, *FRAS1*) are under positive selection in different Orang Asli populations and have been linked to the so-called Southeast Asian 'negrito-like' phenotype (i.e., small body height, curled black hair color and hair shape)²⁶. While these results clearly add evidence for shared genotypes among these populations, the specific genes responsible for the claimed 'negrito-like' phenotype remain uncertain.

Height, a highly polygenic trait with over 700 associated loci³⁴, reflects a complex interplay of genetic and environmental factors. Our analysis identified, amongst others, the gene *DNM3* as being under positive selection in the Maniq population, similar to a finding about the Batak in the Philippines⁷. The short stature of tropical hunter-gatherers has been hypothesized to reflect genetic adaptation to ecological conditions including thermoregulation, reduced food energy requirements, agile tree climbing, and rapid movement in dense rainforests^{8,10–12,30}. An alternative hypothesis proposes that short stature in tropical hunter-gatherers is an outcome of a life history trade-off between accelerated growth and early sexual maturation in high-mortality environment, where early sexual maturation and reproduction is favored (fertility benefits)^{7,31} (but see³²). Nonetheless, we did not find genes or gene-networks related with reproduction under positive selection. Our results therefore support the theory that short stature among rainforest hunter-gatherers is primarily the result of genetic adaptation to similar ecological conditions experienced by these populations. Historically the height of 'negritos' (including Maniq, called 'Tonga' in this publication) was measured around 150 cm³³. From anecdotal observations of our fieldwork we would estimate of the average of the main Maniq group we were with, to be more around 160 cm, with some (young) Maniq being around 175 cm tall. We assume that this is mainly due to increased access to rice and other nutrient rich foods. While this change in diet, can lead to increased height, it could also have adverse effects. Several genes putatively under positive selection in the Maniq are associated with GWAS traits like type II diabetes mellitus, body height, or BMI-adjusted waist circumferences (Fig. 1) suggesting the possibility of polygenic selection. If polygenic selection has indeed contributed to the so-called 'thrifty genotype', as proposed by Neel in 1962³⁴, it could confer advantages to the Maniq population by efficient food sequestration and optimal fat storage during times of food abundance. However, the transition to a sedentary lifestyle in an agricultural or industrialized setting may lead to health issues associated with certain GWAS traits, such as obesity and its consequences, as observed in other populations³⁵.

The Maniq inhabit a tropical, humid environment in the rainforests of southern Thailand. Pathogen pressure is therefore expected to be an important selective force acting on HLA genes as well as on genes of the innate immune system. Reactome analysis identified the immunity pathway 'phosphorylation of CD3 and TCR zeta chains' to be enriched with genes under positive selection as well as balancing selection (Table 2, Supplementary Tables S3 and S6). Among these genes are the HLA genes, which encode molecules crucial for presenting antigens on cell surfaces to T cells. HLA molecules play a pivotal role in mediating the adaptive immune response, and numerous studies have linked HLA diversity to susceptibility to major infectious diseases³⁶. It has been suggested that HLA diversity is mainly influenced by pathogen-driven selection³⁷. The *PTPRC* gene, encoding the CD45 molecule essential for antigen recognition by T and B lymphocytes in viral-host interactions³⁸, was also found

to be under positive selection in the present study, suggesting potential adaptation to pathogen pressure in the tropical rainforest environment inhabited by the Maniq population. This gene was also under positive selection in indigenous individuals living in the low Amazon tropical forest³⁹, reinforce the role of pathogen pressure in shaping immune system diversity in rainforest-dwelling groups.

We also found evidence for balancing selection, particularly among several immune-related genes. The HLA genes show the highest $\beta(1)$ scores, confirming other studies identifying the HLA genomic region as a prime region for balancing selection⁴⁰. Several of the genes involved in immune response under balancing selection in the Maniq were already found to be under balancing selection in previous studies. These include *IL1RL1*⁴¹, *TRIM22*^{42,43} and *DEFB127*⁴⁴. Balancing selection was also found in the *MUC19* gene, which may play a role in mucosal immunity, such as against human metapneumovirus infection⁴⁵. This gene is also under positive selection in North American indigenous populations⁴⁶, underscoring its potential importance in immune defence among diverse populations. The high prevalence of *Streptococcus*, *Neisseria*, and *Haemophilus* in the upper respiratory tract of the Orang Asli⁴⁷ further supports the role of *MUC19* in bacterial clearance, given its expression in the salivary glands, where it contributes to saliva's gel-forming mucin⁴⁸.

Lastly, we found signatures of balancing selection for genes encoding olfactory receptor molecules. Prior research has suggested that the diversity of certain human olfactory receptors may be maintained by balancing selection, possibly through mechanisms such as overdominance⁴⁹. In the latter study the gene *OR1N2*, which is under balancing selection in the Maniq, was also under balancing selection in the Yoruba population in Ibadan, Nigeria (1000 Genomes population). These findings support the hypothesis that genetic diversity in certain human olfactory receptor genes has been shaped by balancing selection, potentially over extended periods through mechanisms such as overdominance⁵⁰.

In our previous study², we provide evidence that the Maniq always have been a small population with a long history of isolation and endogamy. Current estimates on the population size of the Maniq range from 300 to 350 people, estimations before 1960 ranged from 100 to 300 people³³. The extremely high genetic drift of the Maniq could, together with the small sample size, influence our results, although it is unclear in which direction. However, the detection of positive selection in this study was strengthened by employing multiple complementary approaches, including iHS, xp-EHH, and PBE statistics. By integrating these methods and calculating a Fisher score, we ensured a robust identification of genomic regions under positive selection. As with many genetic studies, a higher sample size as well as data from full genome sequencing, would allow for more comprehensive conclusions.

In summary, our study provides novel insights into the genetic adaptations of the Maniq population, highlighting the interplay of positive and balancing selection in shaping traits critical for survival in their rainforest environment. Genes under positive selection were enriched in pathways related to immunity, metabolism, structural integrity, cardiovascular performance, and neuromodulatory signaling, reflecting adaptations to environmental challenges, pathogen pressure, and the demands of a hunter-gatherer lifestyle. Evidence of balancing selection in immune-related genes highlights the critical role of maintaining genetic diversity in a pathogen-rich environment, ensuring robust defenses against a wide array of infectious threats. These findings enhance our understanding of the evolutionary pressures faced by rainforest hunter-gatherers.

Materials and methods

Genomic data preparation and phasing

This study is based on SNP genotype data of 21 Maniq individuals obtained in our previous study using the Infinium® Omni2.5-8 v1.3 BeadChip (Illumina Inc., San Diego, CA). This SNP chip is enriched with an additional 200k exonic SNPs, yielding over 2.5 million markers from autosomes, sex chromosomes, and mitochondrial DNA. The SNP data were processed using PLINK 1.9 software for quality checking and filtering⁵¹. Mitochondrial and sex chromosomal SNPs were removed along with variants that are non-biallelic or duplicates. We included only SNPs with a genotyping rate of 90% ($--geno\ 0.1$), and all SNPs that deviated from Hardy-Weinberg equilibrium were filtered out using the $--hwe\ midp$ threshold filter (with a p -value $< 1e-6$). The Michigan Imputation Server (MIS) (<https://imputationserver.sph.umich.edu/index.html>) was used for phasing and imputation of the autosomal genotype data applying the Minimac4 algorithm⁵². We followed the recommended MIS data preparation guidelines (<https://imputationserver.readthedocs.io/en/latest/prepare-your-data/>) for pre-imputation checks using the imputation preparation tools available at <https://www.well.ox.ac.uk/~wrayner/tools/> (the Perl script HRC-1000G-check-bim-v4.2.13). As reference panel we used the ASIA genome reference legend (available at <https://www.well.ox.ac.uk/~wrayner/tools/ASIA.Genome.Reference.legend.zip>); we specified as population Southeast Asia (SEA) using the provided Perl script *HRC-1000G-check-bim.pl -b <bim file> -f <Frequency file> -r ASIA.Genome.Reference.legend -g -p SEA*; we did not apply any SNP frequency threshold. The Asian reference panel derived from the GenomeAsia 100 K Project⁵³, which also includes genome sequences of the Semang (Malay Negrito) and Onge (Andamanese). The provided VcfCooker software (<https://genome.sph.umich.edu/wiki/VcfCooker>) was used to convert PLINK (binary) format to VCF format. The autosomal SNP genotype files were aligned to the human reference sequence (GRCh37/hg19: <http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/>). For post-imputation control, MIS provides an imputation quality (*MACH-Rsq*) metrics file (*.info* files). This also includes *Rsq* as a quality metric presenting the estimated value of the squared correlation between imputed genotypes and true, unobserved genotypes. We excluded all SNPs with a *Rsq* threshold < 0.8 . The final imputed dataset, after filtering, contained 3,098,600 autosomal SNPs. In addition, we included in this study autosomal SNP genotype data from the 1000 Genomes project (phase 3; <http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>)⁵⁴. These data had already been prepared in a previous study⁵⁵. We selected 25 individuals representing East Asian ancestry (EAS), South Asian ancestry (SAS), European ancestry (EUR), and African ancestry (AFR) from the following populations: CHB (Han Chinese in Beijing, China), BEB

(Bengali from Bangladesh), GBR (British in England and Scotland), and LWK (Luhya in Webuye, Kenya). The data were finally merged with the Maniq genotype data using PLINK (--bmerge).

Identifying signatures of positive directional selection

To detect positive selection (selective sweeps) in the phased autosomal chromosomes, we employed the integrated haplotype score (iHS) approach¹⁶, the cross-population extended haplotype homozygosity (xp-EHH)¹⁷ approach, and the Population Branch Excess (PBE) statistics¹⁸ and combined all three methods by calculating a Fisher score. The iHS method captures the ratio of extended haplotype homozygosity (EHH) for the haplotypes carrying the derived iHHD and ancestral allele iHHA at candidate SNP sites. The required genetic maps were obtained from HapMap phase II b37⁵⁶. We used the software selscan, version 1.2.0a (<https://github.com/szpiech/selscan>)⁵⁷, to calculate iHS scores for non-overlapping windows of 100 kilobase pairs (kb); iHS scores for core sites with a minor allele frequency (MAF) < 0.05 are not calculated by the program; the unstandardized iHS scores were normalized in default frequency bins (--bins 100) across the entire genome using the script norm provided by the selscan programme.

The xp-EHH algorithm compares the extended haplotype homozygosity (EHH) at a given locus in one population to that in a reference population to detect differences in the decay of haplotype homozygosity between the two populations. A positive xp-EHH score indicates longer haplotype homozygosity in the test population, which suggests that positive selection has acted more strongly on the focal population relative to the reference population. Conversely, a negative xp-EHH score implies stronger positive selection in the reference population. We computed xp-EHH scores between the Maniq population and as reference population, the 1000 Genomes East Asian population, Kinhin Ho Chi Minh City (KHV), from Vietnam, using selscan software, version 1.2.0a⁵⁷. The xp-EHH scores were calculated for non-overlapping windows of 100 kilobase pairs (kb) across the autosomal chromosomes. Unstandardized xp-EHH values were normalized using default parameters (--bins 100) to account for allele frequency differences between populations. Pairwise F_{ST} (Maniq and 1000 Genomes populations) were calculated using Weir & Cockerham F_{ST} calculation implemented in VCFtools^{58,59}. Negative F_{ST} values were set to zero. The PBE method¹⁸, is a modified approach of the original Population Branch Statistics (PBS)⁶⁰. PBS is calculated using F_{ST} values between three populations: the focal population (A), and two reference populations (B and C). The PBE approach considers the excess branch length that is specific to the focal population by incorporating additional genomic context and population structure information. We implemented the PBE calculation using custom scripts in R (version 4.1.0)⁶¹ using the formula as described by Yassin in 2016¹⁸.

$$PBE = PBS_{obs} - PBS_{exp} = PBS - [T^{BC} \times (PBS_{med}/T_{med}^{BC})].$$

PBS values are calculated using F_{ST} values between three populations: the focal population A = Maniq population, and two reference populations B = CHB (East Asian ancestry) and C = LWK (African ancestry) from the 1000 Genomes database (see above). The PBS for the focal population A is given by: $PBS_A = (T_{AB} + T_{AC} + T_{BC})/2$ where T_{AB} , T_{AC} , and T_{BC} are the branch lengths derived from the F_{ST} values between focal and reference populations. PBE is calculated using the following formula: $PBE = PBS_{obs} - PBS_{exp}$. Here, PBS_{obs} represents the observed PBS for the focal population, and PBS_{exp} is the expected PBS, calculated as given in the formulae above whereby presents T^{BC} the branch length between the two reference populations B and C and serves to scale locus-specific expectations for expectations for genetic differentiation and should therefore be centred around zero (as observed by the original study¹⁸ and in our data—see Supplementary Figure S1); PBS_{med} is the median PBS value across the genome; T_{med}^{BC} is the median branch length between the reference populations B and C across the genome. The above formula adjusts thus the PBS by subtracting the expected value, which is derived from the PBS_{med} and branch length values, thus highlighting regions with excess differentiation specific to the focal population.

We calculated the genome-wide rank for each of the three statistics independently, using 'rank(-test_statistic, ties.method = "average")' in R. The absolute values of iHS, and only the positive values of xp-EHH and PBE were used to assign the respective ranks. The Fisher score F_{SC} was computed for each SNP by summing $-\log_{10}(\text{rank of the statistic/number of SNPs})$ over all three tests²¹. In this study, we considered SNPs as outliers and candidates under positive selection if they were among the highest 1%. The Ensembl Variant Effect Predictor programme package (<https://github.com/Ensembl/ensembl-vep>)⁶² was used to map genetic information such as gene symbol and consequence type to the SNP data.

Identifying signatures of balancing selection

We used BetaScan v1.0 software (<https://github.com/ksiewert/BetaScan>) to detect the signature of long-term balancing selection¹⁹. In addition, the toolkit of the glactools programme (<https://grenaud.github.io/glactools/>)⁶³ was used to convert the phased chromosomes to folded site frequency spectra format in accordance with the requirements of the BetaScan software (*glactools vcfm2acf vcfm2acf --onlyGT --fai human_g1k_v37.fasta.fai*) to *acf* format. We used as window size 1kbp around the core SNP. This software calculates the allele frequency correlation summary statistic $\beta(1)$ to detect balancing selection using polymorphism data. A characteristic of balancing selection is high polymorphism and an excess of alleles at intermediate frequencies. The β statistic scans for an excess of genomically proximate SNPs that have a very similar allele frequency surrounding a core balanced SNP and then weights the sum of SNP counts around the core SNP higher if they have very similar frequencies to the core SNP. These SNP clusters of SNPs at highly similar intermediate frequency are consistent with the action of balancing selection and are summarized as the $\beta(1)$ statistic, where high values indicate balancing selection. Balancing selection is expected to decrease F_{ST} among populations⁶⁴. We calculated in R genome-wide empirical *p*-values using 100 permutations. Because of the size of the data and number of permutations, we ran this test separately for each chromosome and then calculated the average significance value across all autosomes. We thus report $\beta(1) \geq 6.3$ as statistically significant values ($p < 0.01$) and report only

functional SNP polymorphism, i.e., missense, synonymous variants, and SNPs that are located in the 3' UTR or 5' UTR.

Pathway and gene genontology (GO) analysis

The web-based gene set analysis toolkit (WebGestalt)²⁷ was used for an overrepresentation analysis (ORA) to test whether genes under selection are enriched in certain biological pathways. In our WebGestalt analyses we included pathways from the functional databases Reactome²⁸ and KEGG²⁹. WebGestalt employs the Benjamini-Hochberg (BH) procedure to take multiple testing into account with a false discovery rate (FDR) < 0.01. If more than 10 pathways were found to be enriched, we used the provided weighted set cover method to reduce redundancy in results. We accessed genome-wide association data (GWAS) from the NHGRI-EBI GWAS catalog (<https://www.ebi.ac.uk/gwas/>)⁶⁵ (accessed between October and November 2024) to obtain data on human traits associated with genes under selection. In addition, we used the web-server application snpXplorer (<https://snpxplorer.net/>)⁶⁶ to explore association statistics from published GWAS. The approach examines the proportion of genes associated with the input SNPs for which a previous association was reported in the GWAS catalog. Further information on genes, like gene type or description was obtained using BioMart of Ensembl⁶⁷ (<https://www.ensembl.org/biomart/martview/>).

Data availability

The genomic data can be obtained from 1000 Genomes database. The data from the Maniq generated in this study are not openly available due to reasons of sensitivity and will be made available to researchers upon reasonable request after review and approval by the Data Access Committee led by the corresponding authors.

Received: 17 April 2024; Accepted: 16 December 2024

Published online: 09 February 2025

References

- Fix, A. G. Malayan paleosociology—implications for patterns of genetic-variation among the orang-asli. *Am. Anthropol.* **97**, 313–323 (1995).
- Goellner, T. et al. Unveiling the genetic history of the Maniq, a primary Hunter-Gatherer Society. *Genome Biol. Evol.* **14**, evac021 (2022).
- Lukas, H. Can they save us, the foragers? Indonesian and Thai Hunter-Gatherer cultures under threat from outside (2004).
- Kricheff, D. A. & Lukas, H. Being Maniq: practice and identity in the forests of Southern Thailand. *Hunt. Gatherer Res.* **1**, 139–155 (2015).
- Khunweechuay, N., Roongtawanreongsri, S. & Hatta, K. Cultural forest ecosystem services of the Maniq Indigenous People in Southern Thailand. *Hum. Ecol.* **50**, 559–576 (2022).
- Rees, J. S., Castellano, S. & Andrés, A. M. The genomics of human local adaptation. *Trends Genet.* **36**, 415–428 (2020).
- Migliano, A. B. et al. Evolution of the pygmy phenotype: evidence of positive selection from genome-wide scans in African, Asian, and melanesian pygmies. *Hum. Biol.* **85**, 251–284 (2013).
- Hsieh, P. et al. Whole-genome sequence analyses of western Central African pygmy hunter-gatherers reveal a complex demographic history and identify candidate genes under positive natural selection. *Genome Res.* **26**, 279–290 (2016).
- Perry, G. H. & Verdu, P. Genomic perspectives on the history and evolutionary ecology of tropical rainforest occupation by humans. *Quat. Int.* **448**, 150–157 (2017).
- Bergey, C. M. et al. Polygenic adaptation and convergent evolution on growth and cardiac genetic pathways in African and Asian rainforest hunter-gatherers. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E11256–E11263 (2018).
- Little, M. A. Evolutionary strategies for body size. *Front. Endocrinol.* **11**, 107 (2020).
- Lopez, M. et al. Genomic evidence for local adaptation of hunter-gatherers to the African rainforest. *Curr. Biol.* **29**, 2926–2935 (2019).
- Scheinfeldt, L. B. et al. Genomic evidence for shared common ancestry of East African hunting-gathering populations and insights into local adaptation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 4166–4175 (2019).
- Hedrick, P. W. Balancing selection. *Curr. Biol.* **17**, R230–R231 (2007).
- Bitarello, B. D., Brandt, D. Y. C., Meyer, D. & Andrés, A. M. Inferring balancing selection from genome-scale data. *Genome Biol. Evol.* **15**, 785 (2023).
- Voight, B. F., Kudaravalli, S., Wen, X. Q. & Pritchard, J. K. A map of recent positive selection in the human genome. *Plos Biol.* **4**, 659–659 (2006).
- The International HapMap Consortium. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
- Yassin, A. et al. Recurrent specialization on a toxic fruit in an island Drosophila population. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4771–4776 (2016).
- Siewert, K. M. & Voight, B. F. Detecting long-term balancing selection using allele frequency correlation. *Mol. Biol. Evol.* **34**, 2996–3005 (2017).
- Siewert, K. M. & Voight, B. F. BetaScan2: standardized statistics to detect balancing selection utilizing substitution data. *Genome Biol. Evol.* **12**, 3873–3877 (2020).
- Patin, E. et al. Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. *Science* **356**, 543–546 (2017).
- Ma, X. X. & Xu, S. H. Archaic introgression contributed to the pre-agriculture adaptation of vitamin B1 metabolism in East Asia. *Iscience* **25**, 12 (2022).
- Liu, X. Y. et al. Detecting and characterizing genomic signatures of positive selection in global populations. *Am. J. Hum. Genet.* **92**, 866–881 (2013).
- Deng, L. et al. The population genomic landscape of human genetic structure, admixture history and local adaptation in Peninsular Malaysia. *Hum. Genet.* **133**, 1169–1185 (2014).
- Liu, X. Y. et al. Differential positive selection of malaria resistance genes in three indigenous populations of Peninsular Malaysia (134, pg 375, 2015). *Hum. Genet.* **134**, 677–677 (2015).
- Deng, L. A. et al. Genetic connections and convergent evolution of tropical indigenous peoples in Asia. *Mol. Biol. Evol.* **39**, 456 (2022).
- Liao, Y. X., Wang, J., Jaehnig, E. J., Shi, Z. A. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).

28. Milacic, M. et al. The reactome pathway knowledgebase 2024. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkad1025> (2023).
29. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
30. H Perry, G. et al. Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. *Proc. Natl. Acad. Sci.* **111**, E3596–E3603 (2014).
31. Miglino, A. B., Vinicius, L. & Lahr, M. M. Life history trade-offs explain the evolution of human pygmies. *Proc. Natl. Acad. Sci. U S A.* **104**, 20216–20219 (2007).
32. S. A. Becker, N., Verdu, P., Hewlett, B. & Pavard, S. Can life history trade-offs explain the evolution of short stature in human pygmies? A response to Miglino et al. (2007). *Hum. Biol.* **82**, 17–27 (2010).
33. Brandt, J. H. The negrito of peninsular Thailand. *J. Siam Soc.* **49**, 123–158 (1961).
34. Neel, J. V. Diabetes mellitus—a thrifty genotype rendered detrimental by progress. *Am. J. Hum. Genet.* **14**, 353 (1962).
35. Pressler, M. et al. Dietary transitions and health outcomes in four populations—systematic review. *Front. Nutr.* **9**, 748305 (2022).
36. Alicia, S. M. A review of HLA allele and SNP associations with highly prevalent infectious diseases in human populations. *Swiss Med. Wkly.* **150**, 20214 (2020).
37. Prugnolle, F. et al. Pathogen-driven selection and worldwide HLA class I diversity. *Curr. Biol.* **15**, 1022–1027 (2005).
38. Saunders, A. E. & Johnson, P. Modulation of immune cell signalling by the leukocyte common tyrosine phosphatase, CD45. *Cell. Signal.* **22**, 339–348 (2010).
39. Borda, V. et al. The genetic structure and adaptation of Andean highlanders and amazonians are influenced by the interplay between geography and culture. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 32557–32565 (2020).
40. Barreiro, L. B. & Quintana-Murci, L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat. Rev. Genet.* **11**, 17–30 (2010).
41. Meyer, D., Aguiar, V. R. C., Bitarello, B. D., Brandt, D. Y. C. & Nunes, K. A genomic perspective on HLA evolution. *Immunogenetics* **70**, 5–27 (2018).
42. Andrés, A. M. et al. Targets of balancing selection in the human genome. *Mol. Biol. Evol.* **26**, 2755–2764 (2009).
43. Kelly, J. N., Woods, M. W., Xhiku, S. & Barr, S. D. Ancient and recent adaptive evolution in the antiviral TRIM22 gene: identification of a single-nucleotide polymorphism that impacts TRIM22 function. *Hum. Mutat.* **35**, 1072–1081 (2014).
44. Hollox, E. J. & Armour, J. A. L. Directional and balancing selection in human beta-defensins. *Bmc Evol. Biol.* **8**, 1–14 (2008).
45. McBride, K., Banos-Lara, M. D., Cheemarla, N. R. & Guerrero-Plata, A. Human metapneumovirus induces mucin 19 which contributes to viral pathogenesis. *Pathogens* **9**, 726 (2020).
46. Reynolds, A. W. et al. Comparing signals of natural selection between three indigenous north American populations. *Proc. Natl. Acad. Sci. U S A* **116**, 9312–9317 (2019).
47. Cleary, D. W. et al. The upper respiratory tract microbiome of indigenous Orang Asli in north-eastern Peninsular Malaysia. *Npj Biofilms Microbiomes* **7**, 1 (2021).
48. Culp, D. J. et al. Salivary mucin 19 glycoproteins innate immune functions in streptococcus mutans-induced caries in mice and evidence for expression in human saliva. *J. Biol. Chem.* **290**, 2993–3008 (2015).
49. Alonso, S., López, S., Izaguirre, N. & de la Rúa, C. Overdominance in the human genome and olfactory receptor activity. *Mol. Biol. Evol.* **25**, 997–1001 (2008).
50. Olender, T. et al. Personal receptor repertoires: olfaction as a model. *Bmc Genom* **13**, 1–16 (2012).
51. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 13742 (2015).
52. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
53. Wall, J. D. et al. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature* **576**, 106 (2019).
54. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
55. Schaschl, H., Gollner, T. & Morris, D. L. Positive selection acts on regulatory genetic variants in populations of European ancestry that affect ALDH2 gene expression. *Sci. Rep.* **12**, 4563 (2022).
56. Altshuler, D. M. et al. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
57. Szpiech, Z. A. & Hernandez, R. D. Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* **31**, 2824–2827 (2014).
58. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population-structure. *Evolution* **38**, 1358–1370 (1984).
59. Danecsek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
60. Yi, X. et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010).
61. Team, R. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/> (2024).
62. McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 1–14 (2016).
63. Renaud, G. Glactools: a command-line toolset for the management of genotype likelihoods and allele counts. *Bioinformatics* **34**, 1398–1400 (2018).
64. Nielsen, R. Molecular signatures of natural selection. *Annu. Rev. Genet.* **39**, 197–218 (2005).
65. Bunio, A. et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
66. Tesi, N., van der Lee, S., Hulsman, M., Holstege, H. & Reinders, M. J. T. snpXplorer: a web application to explore human SNP-associations and annotate SNP-sets. *Nucleic Acids Res.* **49**, W603–W612 (2021).
67. Harrison, P. W. et al. Ensembl 2024. *Nucleic Acids Res.* **52**, D891–D899 (2024).

Acknowledgements

We thank our Ethno-Linguist Pacchira Chindaritha (Bangkok) and Tingsabadh Charit (Chulalongkorn University) for their continued support. We especially thank the Maniq for their interest and participation. Research reported in this publication was supported via travel grants 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). We thank Michael Stachowitsch from the University of Vienna for valuable comments on the manuscript. M.L. was supported by the Swedish Research Council (2020-04789) and the European Commission under the Horizon 2020 Marie Skłodowska-Curie Research and Innovation Staff Exchange program (MSCA-RISE-2019, Project Number 873207). W. K. was supported by the Global and Frontier Research University Fund, Naresuan University (Grant number: R2566C051).

Author contributions

T.H. and H.S. conceived and designed this study, T.H. and H.S. performed statistical analyses. T.H. and H.S. wrote the draft manuscript. All authors contributed to the results, read, edited and approved the final manuscript.

Funding

Open access funding provided by University of Vienna. ASEAN - European Academic University Network (ASEA-UNINET) Grants to Helmut Schaschl: ASEA 2019/University of Vienna/4, ASEA 2022–2023/University of Vienna/9 and ASEA-UNINET Bernd Rode Award 2022 (<https://asea-uninet.org/scholarships-grants/bra-lau-reates-2022/>).

Competing interests

The authors declare no competing interests.

Ethics declarations

This study was approved by the Ethics Committee of the University of Vienna (reference number 00444) and the Human Research Ethics Committee of Khon Kaen University (reference number HE622223).

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-83657-0>.

Correspondence and requests for materials should be addressed to T.H. or H.S.

Reprints and permissions information is available at www.nature.com/reprints.

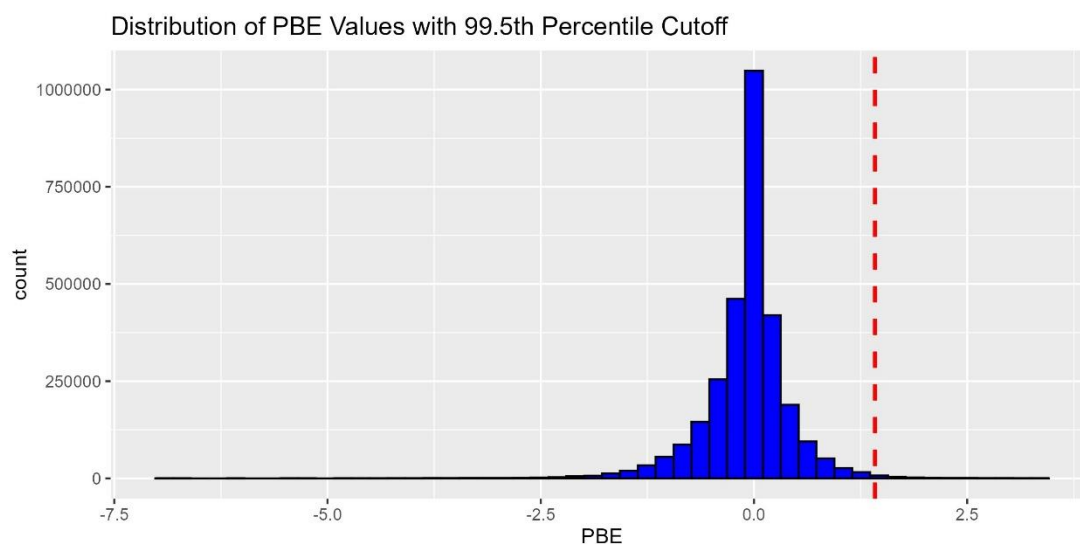
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

Supplementary Information

Supplementary Figure S1. Calculated distribution of PBE values of the Maniq population with 99.5th percentile cutoff. The expectation is that PBE centered around zero; it will be positive to the degree that PBS exceeds its predicted value (population A), and it will be negative if there is elevated genetic differentiation specific to population B or C.



Due to spacing issues only direct links to the supplementary tables can be provided.

Supplementary Table S1. Candidate SNPs for positive selection in the Maniq.

https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-024-83657-0/MediaObjects/41598_2024_83657_MOESM7_ESM.xlsx

Supplementary Table S2. Genes containing candidate SNPs for positive selection in the Maniq.

https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-024-83657-0/MediaObjects/41598_2024_83657_MOESM3_ESM.xlsx

Supplementary Table S3. Details of gene sets under positive selection in the Maniq enriched in Reactome or KEGG pathways.

https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-024-83657-0/MediaObjects/41598_2024_83657_MOESM4_ESM.xlsx

Supplementary Table S4. GWAS traits of putative positive selected genes in the Maniq.

https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-024-83657-0/MediaObjects/41598_2024_83657_MOESM2_ESM.xlsx

Supplementary Table S5. Genes containing candidate SNPs under balancing selection in the Maniq.

https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-024-83657-0/MediaObjects/41598_2024_83657_MOESM5_ESM.xlsx

Supplementary Table S6. Details of gene sets under balancing selection in the Maniq enriched in Reactome or KEGG pathways.

https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-024-83657-0/MediaObjects/41598_2024_83657_MOESM6_ESM.xlsx

Chapter III

HLA Diversity and Signatures of Selection in the Maniq, a Nomadic Hunter-Gatherer Population in Southern Thailand

Authors: Helmut Schaschl, Tobias Herzog, Victoria Oberreiter, Wibhu Kutanan, Mattias Jakobsson, Maximilian Larena

Reference: Schaschl, H., Herzog, T., Oberreiter, V., Kutanan, W., Jakobsson, M., & Larena, M. (2025). HLA diversity and signatures of selection in the Maniq, a nomadic hunter-gatherer population in Southern Thailand. *Immunogenetics*, 77(1), 1-13. <https://doi.org/10.1007/s00251-025-01380-0>

Status: Published in Immunogenetics.

My contribution: I contributed to the design of the study. I contributed to the pipelines to create the various analysis datasets. I contributed to the bioinformatic and statistical analyses. I contributed to the manuscript and edited it.



HLA diversity and signatures of selection in the Maniq, a nomadic hunter-gatherer population in Southern Thailand

Helmut Schaschl¹ · Tobias Herzog¹ · Victoria Oberreiter^{1,2,3} · Wibhu Kutanan⁴ · Mattias Jakobsson⁵ · Maximilian Larena⁵

Received: 9 February 2025 / Accepted: 26 May 2025
© The Author(s) 2025

Abstract

The Maniq, a small and isolated nomadic hunter-gatherer population from the rainforests of Southern Thailand, offer a unique context for investigating how demographic history, genetic drift, and pathogen-driven selection shape human leucocyte antigen (HLA) diversity. Using high-coverage whole-genome data from 21 individuals (12 unrelated), we genotyped HLA alleles with HLA-HD and T1K, identifying 32 alleles in classical and 14 in non-classical HLA genes. Although overall HLA diversity was comparatively low, a few alleles at each locus occurred at high frequency, mirroring patterns observed in other small, isolated populations. Principal-component analysis clustered the Maniq with other Austroasiatic-speaking Semang hunter-gatherers (Jehai, Kintaq) on the Malay Peninsula and, intriguingly, with the Austronesian-speaking Tao of Taiwan, indicating shared immunogenetic features across linguistic boundaries. Despite reduced diversity, multiple loci bore signatures of both long-term balancing and recent positive selection. Several SNPs under selection were in complete linkage disequilibrium with eQTLs known to influence responses to hepatitis B virus (*HBV*) and other pathogens, suggesting that pathogen-driven pressure—in particular *HBV*—may have contributed to recent HLA evolution in the Maniq. These findings provide critical insights into how demographic constraints and pathogen landscapes converge to shape HLA diversity and evolution. In light of increasing infectious disease burdens in indigenous communities, our results underscore the importance of studying small, isolated populations to better understand the adaptive significance of HLA genes.

Keywords Maniq people · Hunter-gatherer · HLA diversity · Positive selection · Balancing selection

Introduction

The immunogenetic diversity of indigenous populations remains poorly understood despite its critical importance in uncovering how human immune systems have adapted to different lifestyles and diverse environmental pressures. This gap in understanding is particularly significant for isolated and small hunter-gatherer populations, whose unique evolutionary trajectories can offer key insights into the dynamics of immune function and genetic adaptation. The Maniq people, residing in the rainforests of southern Thailand, are one of the few remaining nomadic hunter-gatherer groups in Southeast Asia (Kricheff and Lukas 2015). We estimate that there are only about 350 Maniq individuals who still pursue a nomadic hunter-gatherer lifestyle (Göllner et al. 2022). Their genetic isolation and traditional lifestyle, combined with exposure to diverse pathogens in their environment, provide a unique context to explore how natural selection has shaped their immune-related genetic diversity.

✉ Helmut Schaschl
helmut.schaschl@univie.ac.at

¹ Department of Evolutionary Anthropology, Faculty of Life Sciences, University of Vienna, Djerassiplatz 1, 1030 Vienna, Austria

² Human Evolution and Archeological Sciences (HEAS), University of Vienna, Djerassiplatz 1, 1030 Vienna, Austria

³ Konrad Lorenz Institute of Ethology, University of Veterinary Medicine Vienna, Savoyenstraße 1 A, 1160 Vienna, Austria

⁴ Department of Biology, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand

⁵ Human Evolution, Department of Organismal Biology, Uppsala University, Norbyvägen 18 C, 75236 Uppsala, Sweden

The human leukocyte antigen (HLA) system is a fundamental component of the immune response, known for its extreme polymorphism, which is crucial for recognizing and presenting antigens to T cells. The HLA genes are located within the major histocompatibility complex (MHC) on chromosome 6 and include both classical, highly polymorphic class Ia and class IIa genes, as well as the limited polymorphic non-classical HLA class Ib and class IIb genes. Due to its extreme diversity of functionally different HLA alleles, the human MHC region has become one of the most important genomic regions for inter-individual and population-related variations in disease risk, especially for infectious and autoimmune diseases. This significance is highlighted by hundreds of notable associations identified through genome-wide association studies (GWAS) (Kennedy et al. 2017). HLA molecules are also crucial in directing and shaping the repertoire of T cell receptors (TCRs) during T cell maturation in the thymus, a process known as MHC restriction of TCRs. This process ensures that TCRs do not recognize HLA-presenting self-antigens, thereby promoting tolerance and preventing autoimmune responses. Classical class Ia HLA genes are broadly expressed on nucleated cells, allowing CD8+ T cells to recognize and eliminate cells infected with intracellular pathogens, such as viruses. In contrast, classical class IIa genes are expressed only on antigen-presenting cells (dendritic cells, B lymphocytes, and macrophages), presenting antigens to CD4+ T cells, which activate immune responses to target extracellular pathogens. Non-classical class Ib HLA molecules, primarily found in immune and endothelial cells, modulate immune responses by interacting with specific activating or inhibitory receptors; class IIb molecules, expressed on professional antigen-presenting cells, are essential for selecting peptides subsequently presented by classical class IIa molecules, thus fine-tuning the immune response (Beltrami et al. 2023).

Multiple studies across vertebrate species indicate that pathogen-driven selection is a key mechanism in maintaining MHC diversity (Hill et al. 1991; Prugnolle et al. 2005; Sanchez-Mazas et al. 2017). It is widely believed that different forms of balancing selection such as heterozygote advantage, frequency-dependent selection (rare allele advantage), or selection varying in space and time, primarily maintain the polymorphism of MHC genes in humans and other species (Hedrick 2007; Solberg et al. 2008; Key et al. 2014). Furthermore, ancient balancing selection can result in trans-species polymorphism, where many alleles appear to be older than the species in which they are found (Klein et al. 1998). A characteristic of long-term balancing selection is high polymorphism and an excess of alleles with intermediate frequencies close to the balanced variant (Siewert and Voight 2017, 2020; Bitarello et al. 2023). In contrast to balancing selection, positive directional selection results in adaptively important genetic variants increasing in

frequency, leading to fixation or near fixation, resulting in the occurrence of a selective sweep and reduced variability in the area near the selected locus (Hedrick 2007). While most studies support balancing selection as the main evolutionary mechanism maintaining diversity at MHC genes, there is also evidence suggesting that positive directional selection may operate on some MHC loci (Sanchez-Mazas et al. 2017; Meyer et al. 2018; Harris and DeGiorgio 2020; Caro-Consuegra et al. 2022).

In a recent study, we showed that the Maniq are closely related to other Semang populations on the Thai-Malay Peninsula and share their ancestry with the ancient Hôabinhian hunter-gatherers of mainland Southeast Asia (Göllner et al. 2022). The Semang, traditionally hunter-gatherers also known for their nomadic lifestyle, are part of the Orang Asli groups, meaning “original people.” The rainforest and their subsistence strategies likely expose them to diverse pathogens. However, HLA diversity among Southeast Asian hunter-gatherers remains underexplored. To date, only a few studies have investigated HLA variation in Orang Asli (including few Semang) populations in Malaysia (Jinam et al. 2010, 2022; Tasnim et al. 2016), underscoring the need for further research in this region. Although we are not aware of any specific health data published for the Maniq, studies have shown that closely related Orang Asli groups in Malaysia exhibit high prevalence rates of various parasites and infectious diseases (Mahmud et al. 2022). Given their close genetic relationships and geographic proximity, it is reasonable to infer that the Maniq face similar pathogen pressures. In this study, we use whole-genome sequencing (WGS) data to investigate the immunogenetic landscape of the Maniq by examining HLA diversity and pinpointing signatures of balancing and recent positive selection.

Material 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). Saliva samples were collected from Maniq individuals ($n = 21$) who provided informed consent, using the Oragene DNA (OG-500) collection kit (DNA Genotek Inc., Canada). All study participants identified themselves as members of the Maniq people and were at least 18 years old. We visited the Maniq people several times to explain this and our previous study (Göllner et al. 2022). The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments.

Whole-genome sequencing and variant calling

DNA isolation was previously performed as part of a prior study (Göllner et al. 2022; Herzog et al. 2025). In this study, whole-genome sequencing (WGS) with high coverage was conducted. The sequencing libraries were prepared using 100 ng of DNA and the TruSeq Nano DNA sample preparation kit (Illumina Inc.), incorporating unique dual indexes from Illumina. Library preparation followed the manufacturer's guidelines, and next generation sequencing was carried out on the NovaSeq 6000 platform with an S4 flow cell and v1.5 chemistry, producing paired-end reads of 150 base pairs in length. The Genome Analysis Toolkit (GATK) pipeline's best practices were employed for data pre-processing and variant calling (DePristo et al. 2011; Van der Auwera et al. 2013). We employed Variant Quality Score Recalibration (VQSR) with GATK recommended thresholds (truth sensitivity cutoff of 99.0% for single-nucleotide polymorphism (SNP) and 99.9% for indels). Additionally, variants were filtered for Hardy–Weinberg equilibrium ($p < 1 \times 10^{-6}$), and we assessed individual and variant-level missingness using standard QC measures (genotype call rates $> 95\%$). The reads were aligned to the GRCh38 human reference genome using the “bwa mem” algorithm from the Burrows–Wheeler Aligner v0.7.17 (Li and Durbin 2010). Following alignment, duplicate reads were identified and marked with the MarkDuplicates tool, and base quality scores were recalibrated using the BaseRecalibrator and ApplyBQSR tools from GATK v4.1.4.1. Variants were initially called on a per-sample basis using HaplotypeCaller in GVCF mode (Poplin et al. 2018), and the individual GVCF files were then combined into a multi-sample gVCF using the CombineGVCFs tool. Joint genotyping was performed with GenotypeGVCFs, and variant filtering was performed with Variant Quality Score Recalibration, utilizing the VariantRecalibrator and ApplyRecalibration tools from GATK v4.1.4.1, resulting in a final multi-sample VCF file. To ensure unrelated samples, we used the R package SNPRelate (Zheng et al. 2012), first pruning SNPs in linkage disequilibrium (LD; $\text{ld.threshold} = 0.5$), then estimating identity-by-descent using a maximum likelihood approach (snpgdsIBDML, $\text{maf} = 0.01$, $\text{missing rate} = 0.05$). Applying a kinship cutoff of 0.125, we retained 12 unrelated Maniq individuals including a trio for downstream analyses.

HLA genotyping from WGS data

To ensure high accuracy in HLA genotyping, we employed two recently published bioinformatic approaches — HLA-HD v1.7.0 (Kawaguchi et al. 2017) and T1 K v1.0.8 (Song et al. 2023) — which have both been validated for high accuracy in determining HLA alleles from WGS data (Dashti et al. 2024; Lai et al. 2024). Paired-end FASTQ

data were analyzed to obtain HLA genotypes for class Ia (*A*, *B*, *C*), class IIa (*DRA*, *DRB1*, *DQA1*, *DQB1*, *DPA1*, *DPB1*), as well as non-classical class Ib (*E*, *F*, *G*) and class IIb (*DMA*, *DMB*, *DOA*, *DOB*) genes. HLA-HD also requires the software bowtie2 v2.5.4 (Langmead and Salzberg 2012), Samtools v1.2.0 (Danecek et al. 2021), and the Picard software (Picard Toolkit 2019. Broad Institute, release 3.2.0; <https://broadinstitute.github.io/picard/>) to filter reads and to extract mapped reads from the WGS data. T1 K first extracts candidate reads from the FASTQ files and computes the abundance of all the input alleles simultaneously using the weighted expectation–maximization (EM) algorithm to maximize the likelihood of read alignments to the reference HLA alleles. The reads were compared to a reference panel from the IPD-IMGT/HLA database Release 3.57 (Robinson et al. 2020).

HLA allele-based analysis

We analyzed HLA allele and haplotype frequencies using PyPop v1.1.0 (Lancaster et al. 2024). Hardy–Weinberg proportions were evaluated with Guo and Thompson's Monte Carlo exact test and neutrality with the Ewens–Watterson (EW) homozygosity test using Slatkin's implementation. Following PyPop's two-tailed test, $p < 0.025$ indicates balancing selection, whereas $p > 0.975$ suggests directional selection or drift. Population structure was explored with principal-component analysis (PCA) in R v4.4.0 (R Core Team 2024). Besides the Maniq, we included published HLA frequency datasets for two Malaysian Semang groups—Jehai (JEH) and Kintaq (KIN) (Jinam et al. 2010, 2022; Tasnim et al. 2016)—along with East Asian (EAS) populations from the 1000 Genomes Project (Supplementary Table 1), the Tao of Taiwan, Papuans (Goroka Asaro and Madang), Māori, and Aboriginal Australians (Kimberley and Cape York). Source frequencies were taken from the 1000 Genomes HLA panel (Gourraud et al. 2014) and the Allele Frequency Net Database (AFND; Gonzalez-Galarza et al. 2020). PCA was performed based on allele frequencies from five classical HLA loci (*HLA-A*, *HLA-B*, *HLA-C*, *HLA-DQB1*, and *HLA-DRB1*). For most populations, frequencies for all five loci were available; however, for the Tao population, *DQB1* data were not available, and PCA was conducted using the available loci only. Allele frequencies for each locus were pivoted to wide format (missing values set to zero), then centered and scaled prior to PCA via the `prcomp` function (base R stats package). Principal components were merged with population metadata, and the proportion of variance explained by each PC was recorded. PCA results were visualized using `ggplot2` (Wickham 2016), plotting PC1 against PC2.

SNP-based analyses in the MHC region

For the subsequent natural selection and population differentiation analyses, we focused on single-nucleotide polymorphism (SNP) genotype data from chromosome 6, covering the entire MHC region.

Quality control, phasing, and annotation

Variants were filtered to remove multi-allelic sites and those deviating from Hardy–Weinberg equilibrium (HWE) ($p < 1 \times 10^{-6}$) using BCFtools v1.2.0 and VCFtools v0.1.16 (Danecek et al. 2011, 2021). Genotypes were phased with SHAPEIT5 (Hofmeister et al. 2023) using the *phase_common* algorithm for unrelated samples with default parameters. Genetic maps aligned to the GRCh38 reference genome were obtained from the SHAPEIT repository (<https://github.com/odelaneau/shapeit4/tree/master/maps>). Phasing utilized an Asian genetic ancestry reference panel (Supplementary Table 1) created from pre-phased, high-coverage ($30 \times$) genotype data from the 1000 Genomes Project (GRCh38) (Byrska-Bishop et al. 2022). We used Ensembl Variant Effect Predictor (McLaren et al. 2016) to annotate SNPs with gene symbols, biotypes, and consequence types based on GRCh38.p14. Additionally, expression quantitative trait loci (eQTLs) were accessed from GTEx Portal V8 (dbGaP Accession phs000424.v8.p2) (Ardlie et al. 2015) to investigate whether SNPs associated with selected HLA genes function as eQTLs.

Detection of positive selection

To detect positive selection, we applied the integrated haplotype score (iHS) (Voight et al. 2006) and cross-population extended haplotype homozygosity (xp-EHH) (Sabeti et al. 2007) methods on phased autosomal chromosomes, using selscan v1.2.0a (Szpiech and Hernandez 2014). The iHS method assesses extended haplotype homozygosity (EHH) around derived and ancestral alleles at candidate SNP sites, with alleles under positive selection showing unusually long-range linkage disequilibrium (LD) relative to allele frequency. Significant iHS values (≤ -2.0 for derived alleles; ≥ 2.0 for ancestral alleles) indicate positive selection. We calculated iHS in non-overlapping 100-kb windows, normalizing scores with 100 frequency bins across the genome. To compare selection between populations, we used xp-EHH, which examines EHH decay differences at a locus between the test and reference populations. A positive xp-EHH score (> 2.0) suggests stronger selection in the test population. We used the Kinh in

Ho Chi Minh City (KHV) population from the 1000 Genomes Project as the reference population, with unstandardized xp-EHH scores normalized using default settings. SNPs in the 99.9 th percentile (absolute iHS > 2.9 and xp-EHH > 2.4) were considered candidates under positive selection.

Detection of balancing selection

We used the BetaScan software (Siewert and Voight 2017) to detect long-term signatures of balancing selection in the MHC region. We employed the BetaScan2 algorithm to calculate standardized Beta2 scores (Beta2_std) (Siewert and Voight 2020), leveraging chimpanzee as an outgroup to infer ancestral alleles. Ancestral allele data were obtained from the all.epo.gz file derived from the Enredo-Pecan-Ortheus (EPO) multi-species alignments, as used in the original BetaScan studies (Siewert and Voight 2017, 2020). This file was downloaded from the BetaScan GitHub repository and is based on alignments to the GRCh37/hg19 reference genome. Since the EPO ancestral alignments are aligned to the GRCh37/hg19 reference genome, we first converted the VCF files to GRCh37 coordinates using Picard's LiftoverVcf tool. We obtained the necessary chain file for liftover (hg38 ToHg19.over.chain) and the GRCh37 human reference genome from the UCSC Genome Browser (Raney et al. 2024). We used glactools (Renaud 2018) to convert phased SNP genotype data from chromosome 6 into the folded site frequency spectrum format required for BetaScan analysis. SNPs with Beta2_std scores above the 99.9 th percentile (Beta2_std > 17.23) were considered outlier loci and thus putative candidates under balancing selection. Although newer EPO alignments are available on GRCh38, we followed the original BetaScan2 setup to maintain consistency with prior studies and ancestral state inference pipelines.

FST analysis

The phased Maniq SNP genotype data were merged with the phased, high-coverage genotype data (aligned to GRCh38) from the 1000 Genomes Project from the study (Byrska-Bishop et al. 2022), consisting of the four super-populations with East Asian (EAS), South Asian (SAS), European (EUR) and African (AFR) genetic ancestry. We calculated genome-wide pairwise (Maniq vs. 1000 Genomes super-populations) *FST* values for each variant using the Weir and Cockerham method (Weir and Cockerham 1984) implemented in VCFtools. Negative *FST* values were set to zero. We then computed global locus-specific *FST* values and standard deviations (sd) in R v4.4.0.

Results

Maniq HLA allele and haplotype diversity

The average sequencing depth for chromosome 6 was $28.3 \times$, ranging from 19.52 to $40.07 \times$. We employed two methods, HLA-HD (Kawaguchi et al. 2017) and T1K (Song et al. 2023), to identify HLA alleles in the Maniq population. Both methods identified the same HLA alleles. In total, we detected 32 HLA alleles across the class Ia loci (*HLA-A*, *HLA-B*, *HLA-C*) and class IIa loci (*HLA-DRA*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1*, *HLA-DPB1*), as well as 14 alleles in the non-classical HLA loci, including class Ib genes (*HLA-E*, *HLA-F*, *HLA-G*) and class IIb genes (*HLA-DMA*, *HLA-DMB*, *HLA-DOA*, *HLA-DOB*) (Tables 1 and 2, respectively). Most of these alleles are also common in several Southeast Asian populations (in accordance with the HLA Allele Frequency Net Database, AFND) (Gonzalez-Galarza et al. 2020). The prevalence of these common alleles suggests shared ancestry with other Southeast Asian populations and potential local adaptation. A literature search indicated that several of these HLA alleles are associated with increased resistance or susceptibility to widespread human infectious diseases (Supplementary Table 2). However, some of the detected HLA alleles, such as *B*27:06:01*, *B*38:02:01*, and *C*07:199:01*, are rare in the broader Asian region (in accordance with AFND). All HLA loci were in Hardy–Weinberg equilibrium (HWE) (Supplementary Table 3). The Ewens–Watterson (EW) test, however, revealed significant deviations from neutrality for loci *DQA1* ($p = 0.9859$), *DMA* ($p = 1.0000$), and *DMB* ($p = 1.0000$), each showing greater observed homozygosity than expected (Supplementary Table 4). These findings suggest directional selection or genetic drift acting on these loci. All other loci conformed to neutral expectations. Table 3 presents the estimated HLA haplotype frequencies. The most common haplotypes in both HLA class I and class II loci are also found in closely related Semang populations, such as the Jehai and Kintaq, indicating most likely shared ancestry or similar pathogen-driven selection. Additionally, the AFND database shows that the most common HLA haplotypes occur at very low frequency and predominately in East Asian populations. The presence of these haplotypes at low frequencies suggests that the Maniq may have retained some ancestral haplotypes that are now less common elsewhere, possibly due to the effects of genetic drift and isolation.

HLA diversity and population structure in the Semang

The most common HLA alleles found across Semang populations (combined Maniq, Jehai, and Kintaq) are shown in Fig. 1. To explore their population structure in

Table 1 HLA class Ia and class IIa allele frequencies in the Maniq population ($n = 12$)

	Allele	Frequency
HLA class Ia gene		
A	<i>A*24:07:01</i>	0.7917
	<i>A*02:01:02</i>	0.1250
	<i>A*11:01:01</i>	0.0417
	<i>A*24:02:01</i>	0.0417
C	<i>C*03:04:01</i>	0.8333
	<i>C*07:199:01</i>	0.1250
	<i>C*07:02:01</i>	0.0417
B	<i>B*13:01:01</i>	0.7917
	<i>B*18:01:01</i>	0.1250
	<i>B*27:06:01</i>	0.0417
	<i>B*38:02:01</i>	0.0417
HLA class IIa gene		
DRA	<i>DRA*01:01:01</i>	0.9167
	<i>DRA*01:02:02</i>	0.0833
DRB1	<i>DRB1*15:01:01</i>	0.7917
	<i>DRB1*09:01:02</i>	0.1250
	<i>DRB1*12:02:01</i>	0.0417
	<i>DRB1*15:02:01</i>	0.0417
DQA1	<i>DQA1*01:02:01</i>	0.7917
	<i>DQA1*03:02</i>	0.0833
	<i>DQA1*01:01:01</i>	0.0417
	<i>DQA1*03:01:01</i>	0.0417
DQB1	<i>DQA1*06:01:01</i>	0.0417
	<i>DQB1*05:02:01</i>	0.7917
	<i>DQB1*03:03:02</i>	0.1250
	<i>DQB1*03:01:01</i>	0.0417
DPA1	<i>DQB1*05:01:01</i>	0.0417
	<i>DPA1*01:03:01</i>	0.8750
	<i>DPA1*02:01:01</i>	0.0833
	<i>DPA1*02:02:02</i>	0.0417
DPB1	<i>DPB1*02:01:02</i>	0.8750
	<i>DPB1*13:01:01</i>	0.0833
	<i>DPB1*05:01:01</i>	0.0417

a broader regional context, we performed principal component analysis (PCA) based on HLA allele frequencies, incorporating additional populations from East Asia (1000 Genomes Project), the Tao from Taiwan, Māori from New Zealand, Aboriginal Australians, and Papuans from Papua New Guinea. The first two principal components (PC1, 18.73%; PC2, 14.67%) captured major axes of differentiation, together explaining approximately one-third (33.4%) of total genetic variance (Fig. 2). Subsequent principal components explained progressively less variance (PC3, 12.46%; PC4, 10.54%; PC5, 7.86%), indicating diminishing contributions to overall population structure.

Table 2 HLA class Ib and class IIb allele frequency in the Maniq population ($n = 12$)

	Allele	Frequency
HLA class Ib gene		
<i>E</i>	<i>E*01:03:02</i>	0.7917
	<i>E*01:03:01</i>	0.2083
<i>F</i>	<i>F*01:01:01</i>	1.0
<i>G</i>	<i>G*01:04:01</i>	0.7917
	<i>G*01:01:01</i>	0.1667
	<i>G*01:01:03</i>	0.0417
HLA class IIb gene		
<i>DMA</i>	<i>DMA*01:01:01</i>	0.9167
	<i>DMA*01:02:01</i>	0.0417
	<i>DMA*01:03:01</i>	0.0417
<i>DMB</i>	<i>DMB*01:01:01</i>	0.9583
	<i>DMB*01:07:01</i>	0.0417
<i>DOA</i>	<i>DOA*01:01:02</i>	0.8750
	<i>DOA*01:01:04</i>	0.1250
<i>DOB</i>	<i>DOB*01:01:01</i>	1.0

MHC under recent positive selection in the Maniq

The iHS analysis revealed that within the MHC, the *HLA-B* and all class IIa genes are candidates under recent positive selection. Additionally, the xp-EHH analysis identified *HLA-DRB1* and the non-classical class IIb locus *HLA-DOB* as candidates under positive selection (Fig. 3). Unlike classical HLA molecules, *HLA-DOB* is primarily expressed in lysosomes within B cells and plays a regulatory role in HLA-DM-mediated peptide loading onto HLA class II molecules. The lead SNPs with the highest iHS and xp-EHH values as

well as locus-specific F_{ST} values are listed in Supplementary Tables 5 and 6, respectively. Notably, the SNPs under selection are functional variants that serve as eQTLs, potentially modulating HLA gene expression and influencing immune responses. Additionally, the positively selected SNPs at *DPA1* and *DPB1* exhibit complete linkage disequilibrium with two 3'UTR variants: rs3077 and rs9277535, respectively. These variants act as strong cis-eQTLs in immunologically relevant tissues, including the liver, whole blood, and spleen, as shown in the GTEx database, and have been strongly associated with chronic hepatitis B virus (*HBV*) infection outcomes in Asian populations (Kamatani et al. 2009; An et al. 2011; Ou et al. 2019, 2021). This suggests that pathogen-driven selective pressures — particularly from *HBV*—may have played a key role in shaping HLA diversity in the Maniq population.

MHC loci under balancing selection in the Maniq

Our analysis identified the MHC region as a distinct outlier (Supplementary Fig. 1), with several HLA loci showing strong evidence of long-term balancing selection, marked by high Beta2_std scores (Fig. 4). The SNPs with the highest Beta2_std scores, all of which also function as expression eQTLs, are given in Supplementary Table 7. This suggests that these variants not only play a role in genetic diversity but may also influence gene expression, further underlining their functional importance. The HLA genes *DPA1* and *DPB1* exhibit the highest Beta2_std scores, positioning them as major candidates for balancing selection in the Maniq population. In addition to the classical loci, we also detected the non-classical gene *DOA* as an outlier.

Table 3 HLA haplotype frequency estimates for class Ia and class IIb genes in the Maniq population

		Frequency
HLA class Ia	<i>A ~ C ~ B</i>	
	<i>24:07:01 ~ 03:04:01 ~ 13:01:01</i>	0.7474
	<i>02:01:01 ~ 07:199:01 ~ 18:01:01</i>	0.0807
	<i>02:01:01 ~ 03:04:01 ~ 13:01:01</i>	0.0443
	<i>24:07:01 ~ 07:199:01 ~ 18:01:01</i>	0.0443
	<i>11:01:01 ~ 03:04:01 ~ 27:06:01</i>	0.0417
	<i>24:02:01 ~ 07:02:01 ~ 38:02:01</i>	0.0417
	<i>DRA ~ DRB1 ~ DQA1 ~ DQB1 ~ DPA1 ~ DPB1</i>	
HLA class IIa	<i>01:01:01 ~ 15:01:01 ~ 01:02:01 ~ 05:02:01 ~ 01:03:01 ~ 02:01:02</i>	0.7917
	<i>01:01:01 ~ 09:01:02 ~ 03:02 ~ 03:03:02 ~ 01:03:01 ~ 02:01:02</i>	0.0833
	<i>01:02:02 ~ 12:02:01 ~ 06:01:01 ~ 03:01:01 ~ 02:01:01 ~ 13:01:01</i>	0.0417
	<i>01:01:01 ~ 09:01:02 ~ 03:01:01 ~ 03:03:02 ~ 02:02:02 ~ 05:01:01</i>	0.0417
	<i>01:02:02 ~ 15:02:01 ~ 01:01:01 ~ 05:01:01 ~ 02:01:01 ~ 13:01:01</i>	0.0417

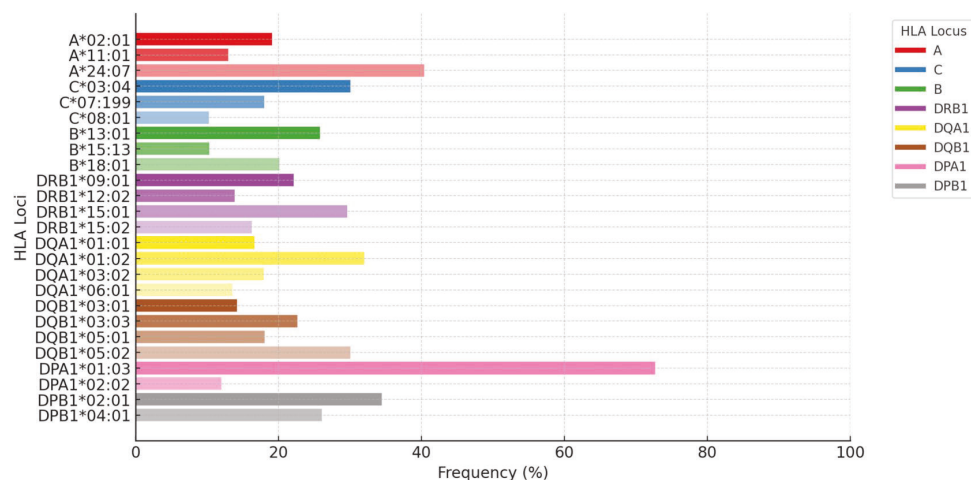
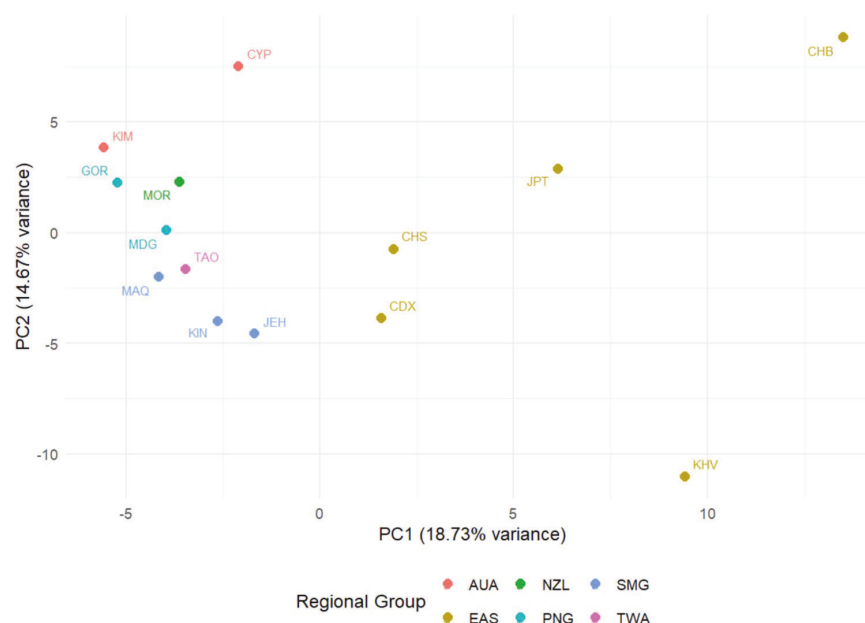


Fig. 1 Most common HLA alleles ($\geq 10\%$) across HLA loci in the Semang ($n = 75$)

Fig. 2 Principal component analysis (PCA) of HLA allele frequencies. PC1 (18.73% of the variance) is plotted against PC2 (14.67%). Point colors indicate regional groupings: Semang populations (SMG) from the Thai–Malay Peninsula—Maniq (MAQ), Jehai (JEH), and Kintaq (KIN); East Asian populations (EAS) from the 1000 Genomes Project (population codes provided in Supplementary Table 1); Aboriginal Australians (AUA) from Kimberley (KIM) and Cape York Peninsula (CYP); Māori (MOR) from New Zealand (NZL); Papuans from Goroka (GOR) and Madang (MAD) in Papua New Guinea (PNG); and the Tao (TAO) from Taiwan (TWA)



Discussion

Our study presents the first analysis of HLA diversity in the Maniq, a small, isolated nomadic hunter-gatherer group inhabiting the rainforests of Southern Thailand. Recent mitochondrial and genome-wide studies (Kutanan et al. 2018; Göllner et al. 2022) have shed light on the Maniq's demographic history, providing important context for interpreting their HLA diversity. Their unique demographic trajectory is characterized by early divergence

from other Southeast Asian groups, limited gene flow, and prolonged isolation. Genome-wide data suggest that the Maniq retain substantial ancient (hunter-gatherer) Hòabinhian-related ancestry, combined with approximately 35% East Asian-related admixture introduced through more recent contact with agriculturalist populations, followed by strong genetic drift and endogamy (Göllner et al. 2022). Such a demographic history has probably contributed to the reduced HLA diversity and the skewed allele-frequency patterns we observe. By identifying 32 alleles

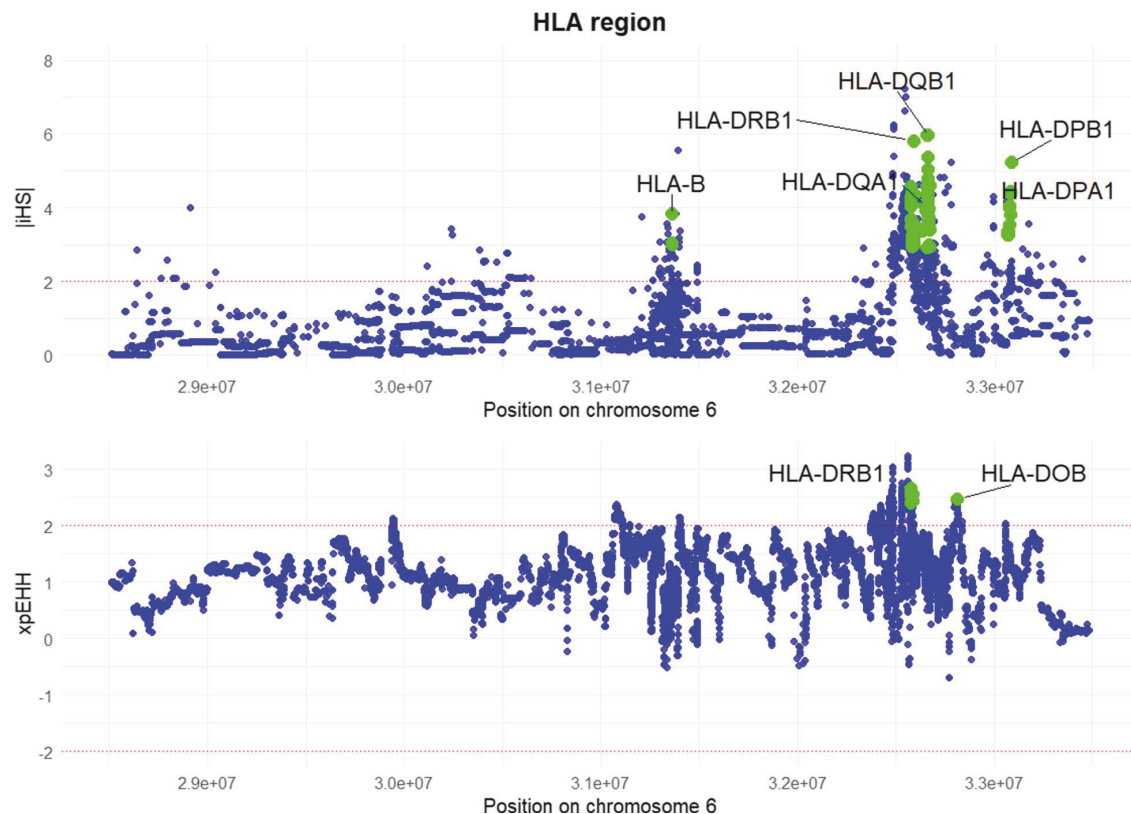


Fig. 3 liHSI and xp-EHH scores plotted across the HLA region. Red dotted line indicates significant iHS and xp-EHH scores >2.0 , and green dots highlight the outlier HLA loci with liHSI values >2.9 and xp-EHH >2.4

from classical HLA genes and 14 from non-classical HLA genes, we reveal patterns of genetic diversity shaped by demographic history, drift, and selection pressures.

We determined the HLA alleles using two recently developed methods, HLA-HD (Kawaguchi et al. 2017) and T1K (Song et al. 2023), applied to WGS data. Both methods yielded highly consistent results, underscoring the robustness of these methods in determining HLA diversity from WGS data. Notably, a few alleles at each HLA locus occur at very high frequencies in the Maniq population. This pattern mirrors findings from the Aché, a small Amerindian population of semi-nomadic hunter-gatherers in Eastern Paraguay (Tsuneto et al. 2003; Single et al. 2020). These parallels suggest that restricted HLA diversity, with a concentration of dominant alleles, may characterize very small, isolated indigenous populations, driven by genetic drift and local adaptation. Our previous research (Göllner et al. 2022) showed that the Maniq have one of the highest levels of genetic drift among living human populations, likely due to their prolonged geographic isolation, small population size, and history of endogamy. Consequently, HLA diversity in the Maniq is also relatively low compared to other Southeast

Asian populations, possibly resulting in skewed HLA allele frequencies.

However, despite reduced diversity, the Maniq share specific HLA alleles with other Southeast Asian populations, indicating shared ancestry and potential common adaptive responses to regional pathogens. Some of these shared alleles have been found to be associated with protective immunity in some Asian populations (Supplementary Table 2). The most common class Ia alleles in the Maniq were *A*24:07:01*, *B*13:01:01*, and *C*03:04:01*; common class IIb alleles included *DRB1*15:01:01*, *DQA1*01:02:01*, *DQB1*05:02:01*, *DPA1*01:03:01*, and *DPB1*02:01:02*. The presence of rare HLA alleles such as *B*27:06:01*, *B*38:02:01*, and *C*07:199:01*, which are uncommon in the broader Asian region, suggests unique evolutionary pressures on the Maniq or the retention of ancestral alleles possibly lost in other populations. Principal component analysis (PCA) of HLA allele frequencies (Fig. 2) revealed distinct patterns of immunogenetic structure among Southeast Asian, Australian, and Oceanian populations. The first two principal components (PC1 and PC2) explained 18.7% and 14.7% of the total variance, respectively. The hunter-gatherer

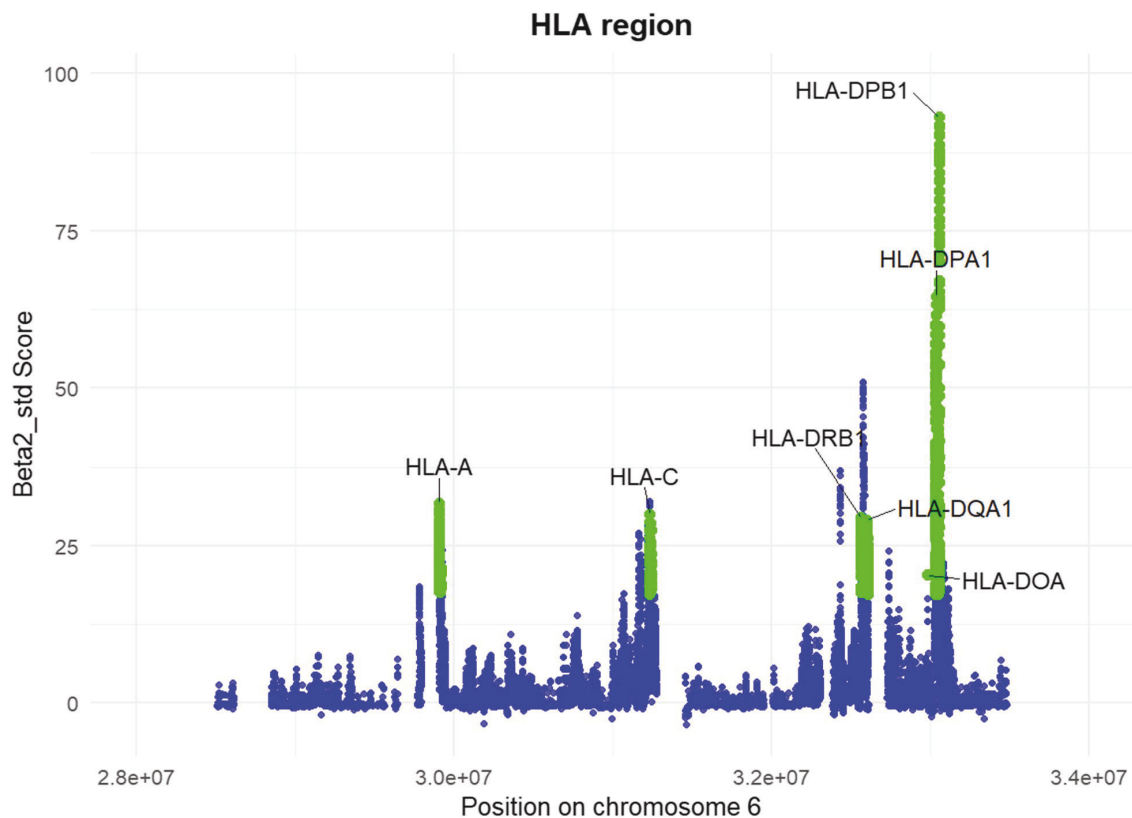


Fig. 4 Standardized Beta2 (Beta2_std) scores plotted across the Maniq HLA region. Green dots highlight outlier HLA loci with Beta2_std scores > 17.23

(Semang) populations Maniq, Jehai, and Kintaq formed a tight cluster, closely aligned with the Tao, an indigenous Austronesian-speaking group of Taiwan, and the indigenous groups from Papua New Guinea (Goroka, Madang). Interestingly, despite linguistic and cultural differences—the Semang being Austroasiatic speakers and the Tao being Austronesian—their clustering may reflect shared immune pressures from similar environments or ancient genetic links across Island and Mainland Southeast Asia (Jinam et al. 2012). In contrast, East Asian populations (Chinese, Japanese, Vietnamese) clustered distinctly, with Aboriginal Australians (Kimberly, Cape York) and Māori forming more differentiated positions along both PCs. This structure mirrors broader continental-scale HLA differentiation patterns, such as those identified by Arrieta-Bolaños et al. (2023), who reported marked HLA discontinuities across Southeast Asia, including along the Wallace Line. These findings suggest that populations, even if geographically isolated, are part of larger immunogenetic ecosystems shaped by migration, drift, and region-specific pathogen pressures.

Some of the HLA alleles such as *C*07:199:01* are fairly common (~ 18%) across Semang groups, suggesting not only shared ancestry but potentially similar selection pressure maintaining specific HLA alleles at high frequency in the hunter-gatherers on the Thai-Malay Peninsula (Fig. 1). Notably, *HLA-C*07:199:01* is nearly identical in sequence to *C*07:04:01* (sequence data from the IPD-IMGT/HLA database (Robinson et al. 2020)), differing only at codon 95 in exon 3, where a phenylalanine to leucine substitution occurs. Due to this subtle difference, earlier studies based on lower-resolution HLA typing may have misclassified or failed to report the allele *C*07:199* in Southeast Asian populations.

Although specific health data for the Maniq are lacking, studies on related Orang Asli groups in Malaysia indicate high infectious disease burdens, including infections with soil-transmitted helminths, protozoan parasites, and viral pathogens such as hepatitis B virus (*HBV*) (Sahlan et al. 2019; Mahmud et al. 2022). Notably, recent research has shown that *HBV* infection rates in some Semang populations

in Malaysia are almost three times higher than the national average (Sahlan et al. 2019). Given their geographic proximity and similar subsistence practices, it is plausible that the Maniq experiences comparable pathogen pressures, which may have influenced the selection of specific HLA alleles associated with immunity to these diseases. Several HLA alleles detected in the Maniq are associated with either protective effects or increased susceptibility to specific pathogens. For instance, *HLA-DPB1*02:01* is linked to protection against chronic *HBV* infection (Nishida et al. 2014; Ou et al. 2021), while *HLA-DPB1*05:01* and *DQB1*05:02* are associated with increased susceptibility (Zhu et al. 2007; Ou et al. 2021). The role of HLA genes in *HBV* infection is further supported by the prevalence of the common allele *HLA-B*13:01* in the Maniq population, which has been associated with enhanced clearance of hepatitis B surface antigen in Asian populations (Miao et al. 2013). Furthermore, a study revealed that *DQB1*03:03*, which is a common allele in the Maniq population, is associated with protection against *Helicobacter pylori* (*Hp*) infection in Asian populations (Wang et al. 2015). Moreover, high prevalence of amoebiasis, caused by *Entamoeba histolytica* infection, has been recorded among Orang Asli (Anuar et al. 2012), and in a study on Bangladeshi children, it has been found that the heterozygous haplotype *DQB1*06:01-DRB1*15:01* had protective effects against this infection (Duggal et al. 2004). Although *DQB1*06:01* was not detected in our study, *DRB1*15:01* has the highest frequency among *DRB1* alleles in the Maniq. This allele has also been found to be associated with a protective role against leishmaniasis (Blackwell et al. 2020). Several of the HLA alleles commonly found in the Maniq, including *HLA-C*03:04*, *DRB1*09:01*, *DRB1*15:01*, and *DQB1*05:02*, are among globally frequent alleles, reinforcing their potential long-term adaptive value (Sanchez-Mazas et al. 2024). Moreover, Arrieta-Bolaños et al. (2023) identified strong genetic barriers in HLA diversity across Southeast Asia, notably along the Wallace Line, suggesting that even isolated populations such as the Maniq are embedded within larger immunogenetic ecosystems shaped by shared histories of migration, drift, and exposure to regional pathogen pressures. These observations show that the HLA profile of the hunter-gatherer groups on the Thai-Malay Peninsula (see PCA in Fig. 2), while unique, reflects broader patterns of selection acting on human populations across time and geography.

We did not find any significant deviation from HWE at the HLA loci (Supplementary Table 3). The absence of HWE deviations despite pathogen-associated alleles may reflect a long-standing equilibrium shaped by past selection events, suggesting we may be observing the genetic legacy of ancient host–pathogen interactions. However, the EW tests of selective neutrality revealed significant ($p > 0.975$) deviations at *DQA1*, *DMA*, and *DMB*, with higher

observed homozygosity than expected under neutrality, indicating potential directional selection or the effects of strong genetic drift at these loci (Supplementary Table 4). For other HLA loci, no significant deviation from neutrality was detected, suggesting more neutral patterns of allele frequency distribution.

The MHC SNP-based analyses revealed variants linked to different HLA genes under both balancing selection and positive selection (Figs. 3 and 4; Supplementary Tables 5, 6 and 7). Balancing selection plays a crucial role in maintaining genetic diversity at immune-related loci, allowing populations to respond to a diverse array of pathogens. Our analysis identified classical HLA class Ia and all classical class IIa loci as candidates under balancing selection. Notably, *DPA1* and *DPB1* exhibited the highest Beta2_std scores (Supplementary Table 7), indicating strong selective pressure to maintain diversity at these loci. Interestingly, the non-classical HLA class IIb gene *DOA* also emerged as a candidate under balancing selection. Beyond that, we observed signals of recent positive selection at multiple classical HLA loci, and xp-EHH pinpointed the non-classical locus *DOB* as under positive selection (Fig. 3). The fact that both classical and non-classical HLA loci show different selective signatures underscores that the entire MHC region may experience varied evolutionary pressures, reflecting the broad pathogen landscape confronting the Maniq. Moreover, the overlapping evidence for long-term balancing and recent positive selection at certain HLA loci highlights a complex interplay wherein populations retain genetic diversity to combat numerous pathogens while also adapting to specific, high-prevalence threats. Notably, *DPA1* and *DPB1* under positive selection in the Maniq have also been reported as positively selected in indigenous Peruvian and Mesoamerican populations (Caro-Consuegra et al. 2022; Garcia et al. 2023). The lead SNPs at *DPA1* and *DPB1* are in complete LD with functional 3'UTR variants (rs3077 and rs9277535) associated with *HBV* infection outcomes in Asian populations (Kamatani et al. 2009; An et al. 2011; Nishida et al. 2014; Mardian et al. 2017). A recent study reported higher levels of *HBV* diversity in Eastern Eurasia compared to Western Eurasia between 5000 and 3000 years ago, as well as a possible transition from non-recombinant *HBV* sub-genotypes to recombinant sub-genotypes (Sun et al. 2024). These historical patterns support the hypothesis that *HBV* may have exerted strong selective pressure favoring protective *HLA-DP* variants in the Maniq, evidenced by their elevated frequencies and high F_{ST} at these SNPs (Supplementary Fig. 2) relative to East Asian populations. These results indicate a central role of HLA genes in host–pathogen co-evolution and the adaptive immune response to viral pathogens like *HBV*.

In conclusion, HLA diversity in the Maniq reflects a dynamic interplay of genetic drift, balancing selection, and recent positive selection—shaped by unique demographic

history and pathogen pressures. These insights underscore the adaptive importance of HLA alleles in small, isolated populations that face diverse pathogenic challenges. The high concordance between FASTQ- and SNP-based genotypes lends confidence to our findings, yet the small sample and the mapping pitfalls of short-read data caution against overinterpretation of locus-specific signals. Future work should combine (i) larger Maniq and neighboring Semang cohorts, (ii) long-read or graph-based assemblies to resolve the complex MHC, and (iii) matched immunological phenotypes. Such integrative studies will clarify how drift, migration, and region-specific pathogens have jointly sculpted HLA evolution in Southeast Asia's remaining hunter-gatherer populations.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00251-025-01380-0>.

Acknowledgements We thank our Ethno-Linguist Pacchira Chindaritha (Bangkok), Helmut Lukas (Austrian Academy of Sciences) and Tingsabath Charit (Chulalongkorn University) for their continued support. We would like to thank Fabian Wieshofer (Vienna, Austria), who supported us in our field research. We especially thank the Maniq for their interest and participation. We would also like to thank the two anonymous reviewers for their work and valuable comments. Research reported in this publication was supported via travel grants 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).

Author contribution Helmut Schaschl and Tobias Herzog conceived and designed this study; the bioinformatic and statistical analyses. Victoria Oberreiter and Wibhu Kutanan supported with the statistical analyses and writing. Mattias Jakobsson and Maximilian Larena lead the genomic analysis and data preparation. All authors contributed to the results, edited, read, and approved the final manuscript.

Funding Open access funding provided by University of Vienna. ASEAN-European Academic University Network (ASEA-UNINET) grants to Helmut Schaschl: ASEA 2019/University of Vienna/4, ASEA 2022–2023/University of Vienna/9 and Bernd Rode Award 2022. Maximilian Larena was supported by the Swedish Research Council (2020–04789) and the European Commission under the Horizon 2020 Marie Skłodowska-Curie Research and Innovation Staff Exchange program (MSCA-RISE-2019, project number 873207). Wibhu Kutanan was supported by the Global and Frontier Research University Fund, Naresuan University (grant number: R2566 C051). Open access funding provided by the University of Vienna.

ASEAN-European Academic University Network, ASEA 2019/University of Vienna/4, ASEA 2022–2023/University of Vienna/9 and Bernd Rode Award 2022, ASEA 2019/University of Vienna/4, ASEA 2022–2023/University of Vienna/9 and Bernd Rode Award 2022, Global and Frontier Research University Fund, Naresuan University, R2566 C051, European Commission, 873207.

Data Availability HLA data used for comparative analyses were obtained from publicly available datasets, including previously published studies and the 1000 Genomes Project and from the Allele Frequency Net Database (<https://www.allelefrequencies.net>). However, due to legal and ethical restrictions, the whole-genome sequencing (WGS) data generated for the Maniq population cannot be made publicly

available. Specific details regarding the datasets and sources used in this study are provided within the manuscript and supplementary materials.

Declarations

Ethics approval and informed consent This study was approved by the Ethics Committee of the University of Vienna (reference number 00444) and the Human Research Ethics Committee of Khon Kaen University (reference number HE622223). The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments. The participants all provided their written informed consent to participate in this study.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- An P, Winkler C, Guan L, O'Brien SJ et al (2011) A common HLA-DPA1 variant is a major determinant of hepatitis B virus clearance in Han Chinese. *J Infect* 203(7):943–7. <https://doi.org/10.1093/infdis/jiq154>
- Anuar TS, Al-Mekhlafi HM, Abdul Ghani MK et al (2012) Molecular epidemiology of amoebiasis in Malaysia: highlighting the different risk factors of Entamoeba histolytica and Entamoeba dispar infections among Orang Asli communities. *Int J Parasitol* 42:1165–1175. <https://doi.org/10.1016/j.ijpara.2012.10.003>
- Ardlie KG, DeLuca DS, Segre AV et al (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348:648–660. <https://doi.org/10.1126/science.1262110>
- Arrieta-Bolaños E, Hernández-Zaragoza DI, Barquera R (2023) An HLA map of the world: a comparison of HLA frequencies in 200 worldwide populations reveals diverse patterns for class I and class II. *Front Genet* 14:866407. <https://doi.org/10.3389/fgene.2023.866407>
- Beltrami S, Rizzo S, Strazzabosco G et al (2023) Non-classical HLA class I molecules and their potential role in viral infections. *Hum Immunol* 84:384–392. <https://doi.org/10.1016/j.humimm.2023.03.007>
- Bitarello BD, Brandt DY, Meyer D, Andrés AM (2023) Inferring balancing selection from genome-scale data. *Genome Biol Evol* 15:. <https://doi.org/10.1093/gbe/evad032>
- Blackwell JM, Fakiola M, Castellucci LC (2020) Human genetics of Leishmania infections. *Hum Genet* 139:813–819. <https://doi.org/10.1007/s00439-020-02130-w>
- Byrska-Bishop M, Evani US, Zhao X et al (2022) High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* 185:3426–3440.e19. <https://doi.org/10.1016/j.cell.2022.08.004>

- Caro-Consuegra R, Nieves-Colón MA, Rawls E et al (2022) Uncovering signals of positive selection in Peruvian populations from three ecological regions. *Mol Biol Evol* 39:. <https://doi.org/10.1093/molbev/msac158>
- Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Danecek P, Bonfield JK, Liddle J et al (2021) Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008. <https://doi.org/10.1093/gigascience/giab008>
- Dashti M, Malik MZ, Nizam R et al (2024) Evaluation of HLA typing content of next-generation sequencing datasets from family trios and individuals of Arab ethnicity. *Front Genet* 15:1407285. <https://doi.org/10.3389/fgene.2024.1407285>
- DePristo MA, Banks E, Poplin R et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498. <https://doi.org/10.1038/ng.806>
- Duggal P, Haque R, Roy S et al (2004) Influence of human leukocyte antigen class II alleles on susceptibility to *Entamoeba histolytica* infection in Bangladeshi children. *J Infect Dis* 189:520–526. <https://doi.org/10.1086/381272>
- Garcia OA, Arslanian K, Whorf D et al (2023) The legacy of infectious disease exposure on the genomic diversity of indigenous Southern Mexicans. *Genome Biol Evol* 15:. <https://doi.org/10.1093/gbe/evad015>
- Göllner T, Larena M, Kutanan W et al (2022) Unveiling the genetic history of the Maniq, a primary hunter-gatherer society. *Genome Biol Evol* 14:4. <https://doi.org/10.1093/gbe/evac021>
- Gonzalez-Galarza FF, McCabe A, Santos EJMD et al (2020) Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res* 48:D783–D788. <https://doi.org/10.1093/nar/gkz1029>
- Gourraud P, Khankhanian P, Cereb N et al (2014) HLA diversity in the 1000 Genomes dataset. *PLOS ONE* 9:. <https://doi.org/10.1371/journal.pone.0097282>
- Harris A, DeGiorgio M (2020) A likelihood approach for uncovering selective sweep signatures from haplotype data. *Mol Biol Evol* 37:3023–3046. <https://doi.org/10.1093/molbev/msaa115>
- Hedrick PW (2007) Balancing selection. *Curr Biol* 17:R230–R231
- Herzog T, Larena M, Kutanan W et al (2025) Natural selection and adaptive traits in the Maniq, a nomadic hunter-gatherer society from Mainland Southeast Asia. *Sci Rep* 15(1):4809. <https://doi.org/10.1038/s41598-024-83657-0>
- Hill AVS, Allsopp CEM, Kwiatkowski D et al (1991) Common West African HLA antigens are associated with protection from severe malaria. *Nature* 352:595–600
- Hofmeister RJ, Ribeiro DM, Rubinacci S, Delaneau O (2023) Accurate rare variant phasing of whole-genome and whole-exome sequencing data in the UK Biobank. *Nat Genet* 55:1243–1249. <https://doi.org/10.1038/s41588-023-01415-w>
- Jinam TA, Hong LC, Phipps ME et al (2012) Evolutionary history of continental Southeast Asians: “early train” hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. *Mol Biol Evol* 29(11):3513–3527. <https://doi.org/10.1093/molbev/mss169>
- Jinam TA, Hosomichi K, Nakaoka H et al (2022) Allelic and haplotypic HLA diversity in indigenous Malaysian populations explored using next generation sequencing. *Hum Immunol* 83:17–26. <https://doi.org/10.1016/j.humimm.2021.09.005>
- Jinam TA, Saitou N, Edo J et al (2010) Molecular analysis of HLA class I and class II genes in four indigenous Malaysian populations. *Tissue Antigens* 75:151–158. <https://doi.org/10.1111/j.1399-0039.2009.01417.x>
- Kamatani Y, Wattanapokayakit WS, Ochi H et al (2009) A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. *Nat Genet* 41:591–595. <https://doi.org/10.1038/ng.348>
- Kawaguchi S, Higasa K, Shimizu M et al (2017) HLA-HD: an accurate HLA typing algorithm for next-generation sequencing data. *Hum Mutat* 38:788–797. <https://doi.org/10.1002/humu.23230>
- Kennedy A, Ozbek U, Dorak M (2017) What has GWAS done for HLA and disease associations? *Int J Immunogenet* 44:195–211. <https://doi.org/10.1111/iji.12332>
- Key FM, Teixeira JC, de Filippo C, Andrés AM (2014) Advantageous diversity maintained by balancing selection in humans. *Curr Opin Genet Dev* 29:45–51. <https://doi.org/10.1016/j.gde.2014.08.001>
- Klein J, Sato A, Nagl S, O’Huigin C (1998) Molecular trans-species polymorphism. *Annu Rev Ecol Syst* 29:1–+. <https://doi.org/10.1146/annurev.ecolsys.29.1.1>
- Kricheff DA, Lukas H (2015) Being Maniq: practice and identity in the forests of Southern Thailand. *Hunt Gatherer Res* 1(2):139–155. <https://www.liverpooluniversitypress.co.uk/doi/10.3828/hgr.2015.9>
- Kutanan W, Kampuansai J, Changmai P et al (2018) Contrasting maternal and paternal genetic variation of hunter-gatherer groups in Thailand. *Sci Rep* 8(1):1536. <https://doi.org/10.1038/s41598-018-20020-0>
- Lai S-K, Luo AC, Chiu IH et al (2024) A novel framework for human leukocyte antigen (HLA) genotyping using probe capture-based targeted next-generation sequencing and computational analysis. *Comput Struct Biotechnol J* 23:1562–1571. <https://doi.org/10.1016/j.csbj.2024.03.030>
- Lancaster A, Single R, Mack S et al (2024) PyPop: a mature open-source software pipeline for population genomics. *Front Immunol* 15:1378512. <https://doi.org/10.3389/fimmu.2024.1378512>
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
- Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Mahmud MH, Baharudin UM, Isa ZM (2022) Diseases among Orang Asli community in Malaysia: a systematic review. *BMC Public Health* 22(1):2090. <https://doi.org/10.1186/s12889-022-14449-2>
- Mardian Y, Yano Y, Wasityastuti W et al (2017) Genetic polymorphisms of HLA-DP and isolated anti-HBc are important subsets of occult hepatitis B infection in Indonesian blood donors: a case-control study. *Viro J* 14:201. <https://doi.org/10.1186/s12985-017-0865-7>
- McLaren W, Gil L, Hunt SE et al (2016) The Ensembl Variant Effect Predictor. *Genome Biol* 17:. <https://doi.org/10.1186/s13059-016-0974-4>
- Meyer D, Aguiar VRC, Bitarello BD et al (2018) A genomic perspective on HLA evolution. *Immunogenetics* 70:5–27. <https://doi.org/10.1007/s00251-017-1017-3>
- Miao F, Sun H, Pan N et al (2013) Association of human leukocyte antigen class I polymorphism with spontaneous clearance of hepatitis B surface antigen in Qidong Han population. *Clin Dev Immunol* 2013:145725. <https://doi.org/10.1155/2013/145725>
- Nishida N, Sawai H, Kashiwase K et al (2014) New susceptibility and resistance HLA-DP alleles to HBV-related diseases identified by a trans-ethnic association study in Asia. *PLoS ONE* 9:e86449. <https://doi.org/10.1371/journal.pone.0086449>
- Ou G, Liu X, Xu H et al (2021) Variation and expression of HLA-DPB1 gene in HBV infection. *Immunogenetics* 73:253–261. <https://doi.org/10.1007/s00251-021-01213-w>
- Ou G, Liu X, Yang L et al (2019) Relationship between HLA-DPA1 mRNA expression and susceptibility to hepatitis B. *J Viral Hepat* 26(1):155–161. <https://doi.org/10.1111/jvh.13012>

- Poplin R, Ruano-Rubio V, DePristo MA et al (2018) Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 201178. <https://doi.org/10.1101/201178>
- Prugnolle F, Manica A, Charpentier M et al (2005) Pathogen-driven selection and worldwide HLA class I diversity. *Curr Biol* 15:1022–1027. <https://doi.org/10.1016/j.cub.2005.04.050>
- R Core Team (2024) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Raney B, Barber G, Benet-Pagès A et al (2024) The UCSC Genome Browser database: 2024 update. *Nucleic Acids Res* 52:D1082–D1088. <https://doi.org/10.1093/nar/gkad987>
- Renaud G (2018) glactools: a command-line toolset for the management of genotype likelihoods and allele counts. *Bioinformatics* 34:1398–1400. <https://doi.org/10.1093/bioinformatics/btx749>
- Robinson J, Barker DJ, Georgiou X et al (2020) IPD-IMGT/HLA database. *Nucleic Acids Res* 48:D948–D955. <https://doi.org/10.1093/nar/gkz950>
- Sabeti PC, Varilly P, Fry B et al (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913–U12. <https://doi.org/10.1038/nature06250>
- Sahlan N, Fadzilah MN, Muslim A et al (2019) Hepatitis B virus infection: epidemiology and seroprevalence rate amongst Negrito tribe in Malaysia. *Med J Malays* 74:320–325
- Sanchez-Mazas A, Cerny V, Di D et al (2017) The HLA-B landscape of Africa: signatures of pathogen-driven selection and molecular identification of candidate alleles to malaria protection. *Mol Ecol* 26:6238–6252. <https://doi.org/10.1111/mec.14366>
- Sanchez-Mazas A, Nunes JM, PGAE HLA Consortium of the 18th International HLA and Immunogenetic Workshop (2024) The most frequent HLA alleles around the world: a fundamental synopsis. *Best Pract Res Clin Haematol* 37(2):101559. <https://doi.org/10.1016/j.beha.2024.101559>
- Siewert KM, Voight BF (2017) Detecting long-term balancing selection using allele frequency correlation. *Mol Biol Evol* 34:2996–3005. <https://doi.org/10.1093/molbev/msx209>
- Siewert KM, Voight BF (2020) BetaScan2: standardized statistics to detect balancing selection utilizing substitution data. *Genome Biol Evol* 12:3873–3877. <https://doi.org/10.1093/gbe/evaa013>
- Single R, Meyer D, Nunes K et al (2020) Demographic history and selection at HLA loci in Native Americans. *PLoS ONE* 15(11):e0241282. <https://doi.org/10.1371/journal.pone.0241282>
- Solberg OD, Mack SJ, Lancaster AK et al (2008) Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. *Hum Immunol* 69:443–464. <https://doi.org/10.1016/j.humimm.2008.05.001>
- Song L, Bai G, Liu XS et al (2023) Efficient and accurate KIR and HLA genotyping with massively parallel sequencing data. *Genome Res* 33:923–931. <https://doi.org/10.1101/gr.277585.122>
- Sun B, Andrades Valtueña A, Kocher A et al (2024) Origin and dispersal history of Hepatitis B virus in Eastern Eurasia. *Nat Commun* 15:2951. <https://doi.org/10.1038/s41467-024-47358-6>
- Szpiech ZA, Hernandez RD (2014) selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol* 31:2824–2827. <https://doi.org/10.1093/molbev/msu211>
- Tasnim AR, Allia S, Edinur HA et al (2016) Distribution of HLA-A, -B and -DRB1 alleles in the Kensiu and Semai Orang Asli sub-groups in Peninsular Malaysia. *Hum Immunol* 77:618–619. <https://doi.org/10.1016/j.humimm.2016.06.009>
- Tsuneto L, Probst C, Hutz M et al (2003) HLA class II diversity in seven Amerindian populations. Clues about the origins of the Ache. *Tissue Antigens* 62:512–526. <https://doi.org/10.1046/j.1399-0039.2003.00139.x>
- Van der Auwera GA, Carneiro MO, Hartl C et al (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinforma* 43:11.10.1–11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43>
- Voight BF, Kudaravalli S, Wen XQ, Pritchard JK (2006) A map of recent positive selection in the human genome. *Plos Biol* 4:659–659. <https://doi.org/10.1371/journal.pbio.0040072>
- Wang J, Zhang Q, Liu Y et al (2015) Association between HLA-II gene polymorphism and *Helicobacter pylori* infection in Asian and European population: a meta-analysis. *Microb Pathog* 82:15–26. <https://doi.org/10.1016/j.micpath.2015.03.011>
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. *Evolution* 38:1358–1370
- Wickham H (2016) *Ggplot2: elegant graphics for data analysis*, 2nd edn. Springer, Switzerland
- Zheng X, Levine D, Shen J et al (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28(24):3326–3328. <https://doi.org/10.1093/bioinformatics/bts606>
- Zhu XL, Du T, Li JH et al (2007) Association of HLA-DQB1 gene polymorphisms with outcomes of HBV infection in Chinese Han population. *Swiss Med Wkly* 137:114–20. <https://doi.org/10.4414/smw.2007.11428>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information

Supplementary Table 1. Populations from the 1000 Genomes Project included in the study (https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/).

Number of individuals	Population code	Population description	Super population	Super population code
99	LWK	Luhya in Webuye, Kenya	Africa	AFR
113	GWD	Gambians from The Gambia	Africa	AFR
99	ESN	Esan in Nigeria	Africa	AFR
108	YRI	Yoruba in Ibadan, Nigeria	Africa	AFR
91	GBR	British in England and Scotland	Europe	EUR
99	FIN	Finnish in Finland	Europe	EUR
107	TSI	Toscani in Italia	Europe	EUR
107	IBS	Iberian populations in Spain	Europe	EUR
86	BEB	Bengali from Bangladesh	South Asian	SAS
102	ITU	Indian Telugu from the UK	South Asian	SAS
96	PJL	Punjabi from Lahore, Pakistan	South Asian	SAS
103	GIH	Gujarati Indians in Houston, Texas, USA	South Asian	SAS
103	CHB	Han Chinese in Beijing, China	East Asian	EAS
104	JPT	Japanese in Tokyo, Japan	East Asian	EAS
99	KHV	Kinhin Ho Chi Minh City, Vietnam	East Asian	EAS
93	CDX	Chinese Dai in Xishuangbanna, China	East Asian	EAS

Supplementary Table 2. HLA alleles of class I (*A*, *C*, *B*) and class II (*DRB1*, *DQA1*, *DQB1*, *DPA1*, *DPB1*) genes identified in the Maniq population and their reported associations (reported OR = odds ratio and *p*-values) with common human infectious diseases in human populations.

HLA allele	Association with prevalent infectious diseases	OR	<i>p</i>
<i>HLA-A*02:01</i>	<i>HCV</i> spontaneous clearance [1]	1.839	0.004
<i>HLA-A*11:01</i>	restricted CD8+ immunity against influenza [2]	-	-
<i>HLA-A*24:02</i>	susceptibility to dengue hemorrhagic fever [3]	1.39	<0.05
<i>HLA-A*24:07</i>	–		
<i>HLA-C*03:04</i>	may have a protective role in <i>HIV</i> infection [4]	-	-
<i>HLA-C*07:02</i>	–	-	-
<i>HLA-C*07:199</i>	–	-	-
<i>HLA-B*13:01</i>	increased clearance of <i>HBV</i> surface antigen [5]	2.62	0.0004
<i>HLA-B*18:01</i>	susceptibility to tuberculosis [6]	-	< 0.001
<i>HLA-B*27:06</i>	potentially a protective role <i>HIV</i> , <i>HCV</i> [7, 8]	-	-
<i>HLA-B*38:02</i>	susceptibility to <i>cryptosporidium</i> , <i>astrovirus</i> infection [9]	2.5, 2.3	0.0056, 0.0074
<i>HLA-DRB1*09:01</i>	susceptibility to tuberculosis [10]	1.82	0.0000988
<i>HLA-DRB1*12:02</i>	protective against leprosy [11]	0.69	0.000259
<i>HLA-DRB1*15:01</i>	protective against leishmaniasis [12, 13]; strong association with multiple sclerosis [14]	0.79, 0.63, 3.1	0.0003, 1.15×10^{-13} , 1×10^{-320}
<i>HLA-DRB1*15:02</i>	protective against leishmaniasis [12]	0.79	0.0084
<i>HLA-DQA1*01:01</i>	–	-	-
<i>HLA-DQA1*01:02</i>	protection from chronic <i>HBV</i> infection liver cirrhosis and hepatocellular carcinoma in Chinese [15]	0.39	0.003
<i>HLA-DQA1*03:01</i>	–	-	-
<i>HLA-DQA1*03:02</i>	–	-	-
<i>HLA-DQA1*06:01</i>	protective role in tuberculosis [16]	0.4	0.02
<i>HLA-DQB1*03:01</i>	protective role in tuberculosis [16]; <i>HCV</i> spontaneous clearance [17]	0.4, 3.899	0.01, 0.0017
<i>HLA-DQB1*03:03</i>	protective against <i>Hp</i> infection [18]	0.54	0.006
<i>HLA-DQB1*05:01</i>	protective against <i>HBV</i> infection [19]	0.59	< 0.00001
<i>HLA-DQB1*05:02</i>	increased risk of tuberculosis [16]; associated with chronic <i>HBV</i> infection [19]; susceptibility to <i>HBV</i> infection [20]	2.6, 1.22, 18	0.01, 0.01, <0.05
<i>HLA-DPA1*01:03</i>	protection against <i>HBV</i> infection [21]	0.51	3.15×10^{-10}
<i>HLA-DPA1*02:01</i>	susceptibility to <i>HBV</i> infection [21]	1.88	2.61×10^{-7}
<i>HLA-DPA1*02:02</i>	susceptibility to <i>HBV</i> infection [21]	1.89	2.66×10^{-5}

<i>HLA-DPBI*02:01</i>	protection against <i>HBV</i> infection [21, 22]	0.68, 0.73	5.22×10^{-6} , <0.0001
<i>HLA-DPBI*05:01</i>	susceptibility to <i>HBV</i> infection [21, 22]	1.45, 1.41	1.51×10^{-4} , <0.001
<i>HLA-DPBI*13:01</i>	protective against severe <i>SARS-CoV2</i> [23]	0.49	0.019

Abbreviations: *HCV* (hepatitis C virus), *HBV* (hepatitis B virus), *EBV* (Epstein-Barr Virus), *SARS-CoV-2* (Severe Acute Respiratory Syndrome Coronavirus type 2), *HIV* (Human Immunodeficiency Virus), *Hp* (*Helicobacter pylori*).

References:

- [1] Huang J, Huang K, Xu R, et al. (2016) The Associations of HLA-A*02:01 and DRB1*11:01 with Hepatitis C Virus Spontaneous Clearance Are Independent of IL28B in the Chinese Population. *Sci Rep.* 6:31485.
- [2] Habel JR, Nguyen At, Rowntree LC, et al (2022) HLA-A*11:01-restricted CD8+ T cell immunity against influenza A and influenza B viruses in Indigenous and non-Indigenous people. *PLOS Pathogens* 18(3):e1010337.
- [3] Cook AC, Thibaut D, Pettersen T (2022) Major Histocompatibility Complex Class I and Dengue Hemorrhagic Fever: A Meta-Analysis of Human Leukocyte Antigens A*24 and B*44. *Cureus* 14(11):e31485.
- [4] Ziegler MC, Nelde A, Weber JK, et al (2020) HIV-1 induced changes in HLA-C*03:04-presented peptide repertoires lead to reduced engagement of inhibitory natural killer cell receptors. *AIDS* 34(12):1713-1723.
- [5] Miao F, Sun H, Pan N, et al (2013) Association of human leukocyte antigen class I polymorphism with spontaneous clearance of hepatitis B surface antigen in Qidong Han population. *Clin Dev Immunol* 2013:145725.
- [6] Kone A, Diarra B, Cohen K, et al (2019) Differential HLA allele frequency in *Mycobacterium africanum* vs *Mycobacterium tuberculosis* in Mali. *HLA*, 93:24-31.
- [7] Goulder PJR., Watkins DI (2008) Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nature Reviews Immunology* 8:619-630.
- [8] McKiernan SM, Hagan R, Curry M, et al (2004) Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. *Hepatology*, 40(1):108-14.
- [9] McCowin S, Moreau G, Haque R, et al (2021) HLA class I and II associations with common enteric pathogens in the first year of life. *EBioMedicine*, 67:103346.
- [10] Toyo-Oka L, Mahasirimongkol S, Yanai H, et al (2017) Strain-based HLA association analysis identified HLA-DRB1*09:01 associated with modern strain tuberculosis, *HLA*, 90:149-156.
- [11] Dallmann-Sauer M, Fava V, Gzara C, et al (2020) The complex pattern of genetic associations of leprosy with HLA class I and class II alleles can be reduced to four amino acid positions. *PLOS Pathogens*, 16(8):e1008818.
- [12] Singh T, Fakiola M, Oommen J, et al (2018) Epitope-Binding Characteristics for Risk versus Protective DRB1 Alleles for Visceral Leishmaniasis. *J Immunol.* 200(8):2727-2737.
- [13] Blackwell JM, Fakiola M, Castellucci LC (2020) Human genetics of leishmania infections. *Hum Genet* 139:813–819.

- [14] The International Multiple Sclerosis Genetics Consortium & The Wellcome Trust Case Control Consortium 2, Sawcer S, Hellenthal G, et al (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476(7359):214-9.
- [15] Liu C, Cheng B (2007) Association of polymorphisms of human leucocyte antigen-DQA1 and DQB1 alleles with chronic hepatitis B virus infection, liver cirrhosis and hepatocellular carcinoma in Chinese. *Int J Immunogenet.* 34(5):373-8.
- [16] Vejbaesya S, Chierakul N, Luangtrakool K, et al (2002) Associations of HLA class II alleles with pulmonary tuberculosis in Thais. *Eur J Immunogenet* 29:431-434.
- [17] Huang J, Xu R, Wang M, et al (2019) Association of HLA-DQB1*03:01 and DRB1*11:01 with spontaneous clearance of hepatitis C virus in Chinese Li ethnicity, an ethnic group genetically distinct from Chinese Han ethnicity and infected with unique HCV subtype. *J Med Virol.* 91(10):1830-1836.
- [18] Wang J, Zhang Q, Liu Y, et al (2015) Association between HLA-II gene polymorphism and *Helicobacter pylori* infection in Asian and European population: A meta-analysis. *Microb Pathog* 82:15–26.
- [19] Ou G, Xu H, Yu H, et al (2018) The roles of HLA-DQB1 gene polymorphisms in hepatitis B virus infection. *J Transl Med.* 16(1):362.
- [20] Zhu XL, Du T, Li JH, et al (2007) Association of HLA-DQB1 gene polymorphisms with outcomes of HBV infection in Chinese Han population. *Swiss Med Wkly*, 137:114-120.
- [21] Nishida N, Sawai H, Kashiwase K, et al (2014) New Susceptibility and Resistance HLA-DP Alleles to HBV-Related Diseases Identified by a Trans-Ethnic Association Study in Asia. *PLOS ONE* 9:e86449.
- [22] Ou G, Liu X, Xu H, et al (2021) Variation and expression of HLA-DPB1 gene in HBV infection. *Immunogenetics* 73:253–261.
- [23] Farias TDJ, Brugiapaglia S, Croci S, et al (2024) HLA-DPB1*13:01 associates with enhanced, and KIR2DS4*001 with diminished protection from developing severe COVID-19. *HLA* 103(1):e15251

Supplementary Table 3. Tests for deviation from Hardy–Weinberg Equilibrium (HWE) across classical and non-classical HLA loci in the Maniq population. HWE was assessed using Guo and Thompson’s exact Monte Carlo test (100,000 iterations). The number of observed and expected homozygotes are shown for each locus. No significant deviations from HWE were detected ($p > 0.05$ for all loci). Slatkin’s exact test p -values are also provided for comparison.

Locus	N	k (alleles)	Homozygotes (Obs/Exp)	Guo & Thompson p -value	Slatkin p -value
HLA-A	12	4	7/7.75	1	0.9170
HLA-B	12	4	7/7.75	1	0.9170
HLA-C	12	3	8/8.54	1	0.8102
HLA-DRA	12	2	10/10.17	1	0.7193
HLA-DRB1	12	4	7/7.75	1	0.9170
HLA-DQA1	12	5	7/7.67	1	0.9859
HLA-DQB1	12	4	7/7.75	1	0.9170
HLA-DPA1	12	3	9/9.29	1	0.9069
HLA-DPB1	12	3	9/9.2	1	0.9069
HLA-E	12	2	7/8.04	1	0.3867
HLA-F	12	1	12/12	-	-
HLA-G	12	3	9/7.88	0.4024	0.7277
HLA-DMA	12	3	10/10.12	1	1.0000
HLA-DMB	12	2	11/11.04	1	1.0000
HLA-DOA	12	2	11/9.38	0.1292	0.5730
HLA-DOB	12	1	12/12	-	-

Supplementary Table 4. Slatkin's implementation of Ewens-Watterson (EW) homozygosity test of neutrality. Given are Observed F (Obs F), Expected F (Exp F), normalized deviate of the homozygosity F (Fnd) and the corresponding *p*-value. The column 'Relation to F Neutral' specifies whether the observed homozygosity is higher, lower, or about equal to the expected homozygosity. Entries marked with “–” indicate loci that are monomorphic. Significant deviations from neutrality (directional selection or drift) are indicated by *p*-values > 0.975 (highlighted in bold).

HLA genes	Obs F	Exp F	Fnd	<i>p</i> -value of F	Relation to F Neutral
Classical HLA genes					
<i>HLA-A</i>	0.6458	0.4602	1.4630	0.9170	Higher
<i>HLA-C</i>	0.7188	0.5765	0.9127	0.8102	Higher
<i>HLA-B</i>	0.6458	0.4602	1.4630	0.9170	Higher
<i>HLA-DRA</i>	0.8472	0.7441	0.6765	0.7193	Higher
<i>HLA-DRB1</i>	0.6458	0.4602	1.4630	0.9170	Higher
<i>HLA-DQA1</i>	0.6389	0.3759	2.4866	0.9859	Higher
<i>HLA-DQB1</i>	0.6458	0.4602	1.4630	0.9170	Higher
<i>HLA-DPA1</i>	0.7743	0.5765	1.3345	0.9096	Higher
<i>HLA-DPBI</i>	0.7743	0.5765	1.3345	0.9096	Higher
Non-classical HLA genes					
<i>HLA-E</i>	0.6701	0.7441	-0.4847	0.3867	Lower
<i>HLA-F</i>	–	–	–	–	–
<i>HLA-G</i>	0.6562	0.5765	0.5378	0.7277	Higher
<i>HLA-DMA</i>	0.8438	0.5765	1.8031	1.0	Higher
<i>HLA-DMB</i>	0.9201	0.7441	1.1547	1.0	Higher
<i>HLA-DOA</i>	0.7812	0.7441	0.2439	0.5730	Higher
<i>HLA-DOB</i>	–	–	–	–	–

Supplementary Table 5. HLA SNPs with the highest $|iHS|$ values and global locus-specific F_{ST} values (standard deviation, sd) calculated across all analyzed populations. SNP functions were obtained from Ensembl Variant Effect Predictor (Assembly: GRCh38.p14) database and from GTEx v8.

Locus	SNP	Position	Function	$ iHS $	$F_{ST}(sd)$
<i>B</i>	rs36057188	31360983	upstream, eQTL	3.8	0.878 (0.067)
<i>DRB1</i>	rs9270187	32588188	intron, eQTL	5.8	0.683 (0.034)
<i>DQA1</i>	rs9273067	32644436	downstream, eQTL	4.1	0.541 (0.143)
<i>DQB1</i>	rs1130432	32661352	missense, eQTL	5.4	0.507 (0.040)
<i>DPA1</i>	rs9380336	33072959	intron, eQTL	4.1	0.221 (0.182)
<i>DPB1</i>	rs9277516	33086503	intron, eQTL	5.3	0.236 (0.146)

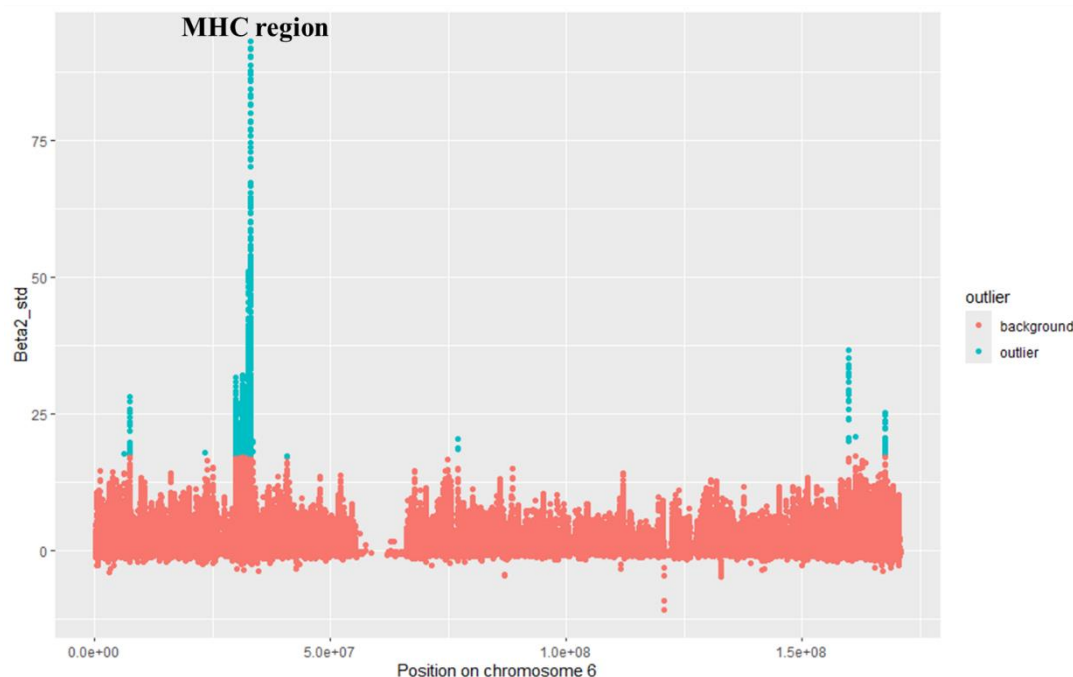
Supplementary Table 6. HLA SNPs with the highest xp-EHH values and F_{ST} values (standard deviation, sd) calculated for the pairwise comparison Maniq versus East Asians. SNP functions were obtained from Ensembl Variant Effect Predictor (Assembly: GRCh38.p14) database and from GTEx v8.

Locus	SNP	Position	Function	xp-EHH	$F_{ST}(sd)$
<i>DRB1</i>	rs9269465	32574470	downstream, eQTL	2.7	0.240 (0.109)
<i>DOB</i>	rs12207915	32809975	downstream, eQTL	2.5	0.801 (0.147)

Supplementary Table 7. HLA SNPs with top-scoring $Beta2_std$ ($\beta^{(2)}_{std}$) scores and global locus-specific F_{ST} values (sd = standard deviation) calculated across all analyzed populations. SNP functions were obtained from Ensembl Variant Effect Predictor (Assembly: GRCh37/hg19) database and from GTEx v8 – see for further information in Material and Methods.

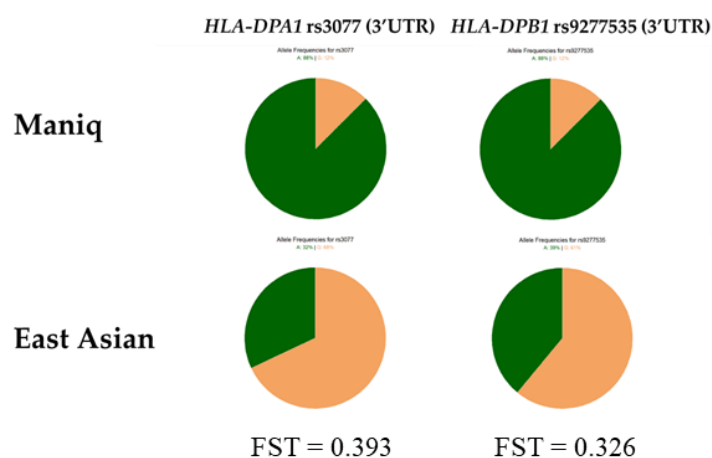
Locus	SNP	Position	Function	$\beta^{(2)}_{std}$	$F_{ST}(sd)$
<i>A</i>	rs9260080	29909071	5'UTR, eQTL	34.04	0.074 (0.068)
<i>C</i>	rs9264532	31234381	downstream, eQTL	30.07	0.062 (0.022)
<i>DRB1</i>	rs28366307	32560995	upstream, eQTL	29.53	0.032 (0.023)
<i>DQA1</i>	rs9272036	32598841	intron, eQTL	28.36	0.309 (0.077)
<i>DOA</i>	rs9276986	32980222	upstream, eQTL	20.35	0.219 (0.073)
<i>DPA1</i>	rs77527547	33034118	intron, eQTL	64.58	0.197 (0.181)
<i>DPB1</i>	rs9277511	33054215	intron, eQTL	93.08	0.201 (0.145)

Supplementary Figure 1. Plot of Beta2_std scores across chromosome 6; the MHC region appears as distinct outlier region.



Supplementary Figure 2. Allele frequencies of the two 3'UTR SNPs rs3077-G/A and rs9277535-A/G in the Maniq and East Asian populations (1000 Genomes Project).

These two variants function as eQTLs in several tissues such as liver, whole blood, and spleen, and are in complete linkage disequilibrium (LD) with positively selected variants at the *HLA-DPA1* and *HLA-DPB1* loci in the Maniq population. The A alleles (shown in green) have been associated with a decreased risk of chronic *HBV* infection in Asian populations (e.g., Kamatani et al. 2009).



Concluding Discussion

This thesis presents novel insights into the evolutionary history, population genetics, genetic adaptations, and immunogenetics of the Maniq.

The Maniq as an Isolated Population

The genomic evidence, especially the results of the ROH analysis, points towards a population of small size and long stretches of endogamy (Chapter I). This result is congruent with the known history of the Maniq. Early written publications on the Maniq already speak of their small group size (Brandt 1961; Evans 1968). Even though the Maniq were never fully cut off from neighbouring societies, their contacts were always limited. This of course changed in recent years due to deforestation and a reduction in their natural habitat. Contacts with outsiders rarely resulted in gene flow towards the Maniq. There have always been accounts of Maniq leaving their respective group, yet we know of no ‘Hamiq’ joining the Maniq and even producing offspring. Nevertheless, ties with ‘Hamiq’ could have been extensive. In one interview a Maniq even referred to a non-Maniq person as their grandfather, a designation intended in a non-biological sense. More widely known is the story of Kanang (or Khanung) who was adopted by King Rama V (Chulalongkorn) of Siam (Hamilton 2006). The king even wrote a play about the (imagined) life of the Semang boy titled ‘Ngo Pa’. Our main informant (who self-identifies as a Maniq, and whom Helmut Lukas has known for over 25 years) and other sources tell of at least 11 different groups of Maniq wandering through the south of Thailand. Some groups are more open to contact with ‘Hamiq’ and their way of life, while others do not want to have any contact with ‘Hamiq’ and actively avoid contact. One can assume that this diversity in opinion regarding contact with outsiders has been an ongoing debate within the Maniq as long as there have been ‘Hamiq’ in the area. The genomic evidence supports the data obtained from linguistic and cultural sources that the Maniq are most closely related to the Semang (Chapter I). For instance, the language of the Maniq is classified within the Aslian branch of the Austroasiatic language family. They can be modelled as a mixture of Andamanese- and East Asian ancestry (Chapter I). When compared with other Semang populations, the Maniq consistently display the highest values for Andamanese-related ancestry,

which again points to limited contact with outsiders (Chapter I). Additionally, the Maniq possess one of the highest levels of population F_{ST} in the worldwide set of populations used in our study (Chapter I). Together with our other results this simply highlights the extremely high amount of genetic drift in the Maniq. In the context of the peopling of Southeast Asia and the further expansion towards Australia, the analysed data of the Maniq are congruent with current scientific discussion but do not provide additional insights (Chapter I). Due to their extreme genetic drift paired with the low sample size, the statistical power for some algorithms was limited. This is also the reason why we refrained from publishing any results on time to most recent common ancestor (TMRCA) or estimations on effective population sizes. These topics can be revisited if the Maniq decide to participate in further studies. As discussed above, the Maniq, like other Semang groups, carry both Andamanese and East Asian-related ancestries, but exhibit the highest proportion of the Andamanese-related ancestry. This pattern is consistent with the hypothesis that when traveling southward the Maniq remained in their area and became more isolated from other populations with East Asian-related ancestries. Those groups could either have travelled southwards along the Mekong River, or northward from the Malay Peninsula. The latter migration could explain some Austronesian words present in the language of the Maniq. Yet, one must remain careful in interpreting that evidence, since language of course could change more quickly (e.g. adaptation to the use of loanwords), and does not immediately imply genetic admixture. Revisiting the peopling of Mainland Southeast Asia, additional data continue to complicate inferred migration and gene flow patterns.

Lastly, when looking at ancient ancestry components of Neanderthal and Denisovan, the Maniq exhibit introgression of Neanderthal DNA and no signs of introgression of Denisovan DNA (Chapter I). This is also in line with evidence from other Semang groups. The level of Neanderthal introgression seems to be elevated when compared to some Semang groups, but the standard errors are quite large and do not allow for a sound statistical inference (Chapter I). Overall, the compiled evidence presented in Chapter I supports the view that the Maniq likely are the most pristine Semang group living today. This interpretation is also consistent with geography, since they are the northernmost Semang group.

Adapting to a Hunter-Gatherer Lifestyle and the Tropical Rainforest

Since there are no signs of the Maniq ever having changed their lifestyle, one can assume that their genomics adapted to their hunter-gatherer lifestyle over the course of time. Similarly, they also remained in their natural habitat, the tropical rainforest in southern Thailand. It must be mentioned that this rainforest is rapidly dwindling, and thus the environment for hunting and gathering is shrinking. We scanned the genomes of the Maniq for positive and balancing selection and found different genome-wide association studies (GWAS) traits, as well as enriched biological pathways. Several genes that we found to be under positive selection, have also been detected by other studies on various Asian populations (Chapter II). We could identify four genes (*PTPRN2*, *NRXN3*, *AMZ1*, *FRAS1*) to be under positive selection, which have also been found in other Orang Asli groups (Chapter II). They could be linked to the so called ‘negrito-like’ phenotype. While shared genotypes among those populations are evident, a specific genotype for the proposed phenotype remains unclear (Chapter II). Height is a classic example of a phenotypic trait that is highly polygenic (e.g. Yengo et al. 2018).

Similar to the findings in the Batak of the Philippines, the Maniq exhibit signals of positive selection at *DNM3* (Chapter II). While other populations tend to increase in height over time, the Maniq remained rather short in stature for a long time. Early works put their average height at around 150 cm (Brandt 1961). In our own fieldwork we did not produce any data on body height of the Maniq. Our guesstimates would put the Maniq nowadays at around 160 cm, with some young males being around 175 cm in height. Of course, this is not formal evidence but should just underline that the increased access to nutrient-rich foods and rice over the last two decades has left an imprint on the Maniq.

There are two main competing hypotheses as to why the phenotype of ‘short stature’ has arisen in rainforest hunter-gatherers. The first one hypothesizes that it is mainly an adaptation to the rainforest environment and enables more agile tree climbing, better movement and thermoregulation, as well as reduced food requirements (e.g. Lopez et al. 2019; Little 2020). The second hypothesis posits that there is actually a trade-off between accelerated growth and early sexual maturation in a high-mortality environment, which favours earlier sexual maturation and reproduction (Migliano et al. 2007; Becker et al. 2010). In our study, we did not find any signals of selection regarding sexual maturation or reproduction, yet we did find the aforementioned genes under selection (Chapter II). Our

study in Chapter II therefore does support the first hypothesis and finds no evidence for the alternative hypothesis.

Further evidence for adaptation to a rainforest environment includes enriched pathways regarding extracellular matrix organization and collagen synthesis (Chapter II). The enriched collagen genes include *COL11A2*, *COL5A1*, *COL4A2*, and *COL18A1* (Chapter II). Where *COL11A2* was also found to be under positive selection in other populations ranging from East Asian, over indigenous Mesoamericans, to the Luhya in Kenya (Chapter II). Those genes are linked to structural support and mobility, which probably play a role in humid rainforest environments (Chapter II). Additionally, we found evidence for positive selection acting upon the adrenergic signaling in cardiomyocytes pathway (Chapter II). This is potentially a cardiovascular adaptation to increase stress resilience and physical performance. Interestingly we also found genes of the oxytocin signaling pathway to be under positive selection (Chapter II). Since oxytocin is involved in many different processes from labour (uterine contraction) to social bonding, selection could be acting on either or all of those processes. The Maniq are an egalitarian, face-to-face society. Any forces increasing cooperative social behaviour or parenting may have a positive impact on survival in those societies.

We found several genes under positive selection which are associated with the trait ‘Type II Diabetes Mellitus’ in the GWAS-catalog (Chapter II). Additionally, we also found the gene set for Type I Diabetes Mellitus to be enriched for positive and balancing selection (Chapter II). This – at first glance – puzzling picture should inspire future research. At the same time, together with the other positively selected genes associated with body height or BMI-adjusted waist circumference this could point towards a polygenic selection on metabolism in general (Chapter II). All in all, the evidence may contribute to a ‘thrifty genotype’, where, in times of food abundance, energy is stored efficiently. Nowadays the Maniq are in a phase of transitioning, and a sedentary lifestyle could lead to metabolic disorders (e.g. diabetes or obesity) for some of them.

The last topic we investigated, immunity, could also be affected by a change in environment. As presented in Chapter II, both forces of selection acted upon immune-related genes in the Maniq. Two notable examples of genes under selection are *PTPRC*, for positive selection, and *MUC19* for balancing selection (Chapter II). *PTPRC* encodes CD45, which is central for antigen recognition by T and B

lymphocytes in pathogen-host interactions (Saunders and Johnson 2010). It was also found to be under positive selection in various populations in the lower Amazon rainforest (Borda et al. 2020), which could point towards an environmental influence for this adaptation. *MUC19* encodes one member of the gel-forming mucins that are part of mucus, and play a role in bacterial clearance (Culp et al. 2015). Mucus itself is part of the innate immune system of humans and may protect against different respiratory pathogens that are prevalent among the Orang Asli, like *Streptococcus*, *Neisseria*, and *Haemophilus* (Cleary et al. 2021). Interestingly, *MUC19* has been found to be under positive selection amongst North American indigenous populations (Reynolds et al. 2019). The major histocompatibility complex (MHC) contains the human leukocyte antigen (HLA) genes, which are an elementary component of the human immune system. This highly polymorphic region of the genome shows both signs of positive and balancing selection in the Maniq (Chapter II). This points towards a pathogen-driven adaptation which maintains diversity.

Immunogenetic Patterns of the Maniq

The demographic history, including genetic drift and different selection pressures, of the Maniq produced skewed allele-frequency patterns and a reduced HLA diversity (Chapter III). A few alleles at each HLA locus occur at very high frequency in the Maniq, similar to the Aché of Eastern Paraguay, who were formerly also nomadic hunter-gatherers (Tsuneto et al. 2003; Single et al. 2020). These parallels suggest that small, isolated populations experiencing strong genetic drift and local adaptation, exhibit reduced HLA diversity, together with a dominance of some specific alleles. Even though we observe this reduced diversity, the Maniq also share some HLA alleles with Southeast Asian populations (Chapter III). Interestingly, certain rare HLA alleles (*B*27:06:01*, *B*38:02:01*, and *C*07:199:01*) of the Maniq are uncommon in the Asian region (Chapter III). This could either be due to unique evolutionary pressures acting on the genomes of the Maniq, or the retention of ancestral alleles which may have been lost in other populations.

A simple principal component analysis (PCA) on HLA frequencies shows a clustering of Maniq, other Semang groups, and indigenous populations of Taiwan and Papua New-Guinea (Chapter III). This

cluster indicates shared immune pressures from similar environments or some ancient genetic links between Mainland and Island Southeast Asia (Jinam et al. 2012).

We are not aware of any publicly available health data on the Maniq, but there are studies on Orang Asli groups (Sahlan et al. 2019; Mahmud et al. 2022). The studied groups in Malaysia showed high infectious disease burdens due to soil-transmitted helminths, protozoan parasites and viruses such as hepatitis B (HBV). Furthermore, HBV infection rates in some Semang populations were nearly three times higher than the Malaysian average. We presume that the Maniq are exposed to similar pathogen pressures and high infectious disease burdens. For each of the mentioned infectious diseases, we find at least one HLA allele that is potentially protective against it (Chapter III). Interestingly, for HBV we also find susceptibility markers (Chapter III). To dive a bit deeper, we also applied selection tests, where we found balancing selection to act upon HLA-A, -C, DPA1, DPB1 and DOA, and positive selection acting upon HLA-B, DPA1, DPB1 and DOB (Chapter III). Crucially, the loci of DPA1 and DPB1 which positive selection acts on, are linked to HBV infection outcomes. Similar findings have been reported for some indigenous American populations (Caro-Consuegra et al. 2023; Garcia et al. 2023). It seems that HBV has been a major selective force, shaping the HLA allele diversity in the Maniq (Chapter III).

The Future of the Maniq

The Maniq and their lifestyle as hunter-gatherers are endangered. Hunter-gatherers all over the world struggle to maintain their way of living and are faced with the phenomenon called ‘transitioning’, where they abandon their ancestral way of living. The Maniq are a very heterogenous community, where some groups are open and interested in contact with Hamiq, while others want to be left alone. I am no expert in navigating this difficult landscape of political, legal and social decisions that must be made. I think protecting their environment by stopping excessive land use and regulating tourist flows will help them to maintain their hunter-gatherer lifestyle. Currently, the future of the Maniq remains uncertain. The Maniq have survived for centuries, probably even longer, and have a chance to exist even longer, if we, the Hamiq, acknowledge their value to our world.

References

- Aghakhanian F et al. 2015. Unravelling the Genetic History of Negritos and Indigenous Populations of Southeast Asia. *Genome Biol Evol.* 7(5):1206–1215. <https://doi.org/10.1093/gbe/evv065>
- Becker NSA, Verdu P, Hewlett B, Pavard S. 2010. Can Life History Trade-Offs Explain the Evolution of Short Stature in Human Pygmies? A Response to Migliano et al. (2007). *Hum Biol.* 82(1):17–27. <https://doi.org/10.3378/027.082.0101>
- Bergey CM et al. 2018. Polygenic adaptation and convergent evolution on growth and cardiac genetic pathways in African and Asian rainforest hunter-gatherers. *Proc Natl Acad Sci.* 115(48):E11256–E11263. <https://doi.org/10.1073/pnas.1812135115>
- Bergström A et al. 2021. Origins of modern human ancestry. *Nature.* 590(7845):229–237. <https://doi.org/10.1038/s41586-021-03244-5>
- Borda V et al. 2020. The genetic structure and adaptation of Andean highlanders and Amazonians are influenced by the interplay between geography and culture. *Proc Natl Acad Sci U S A.* 117(51):32557–32565. <https://doi.org/10.1073/pnas.2013773117>
- Brandt JH. 1961. The negrito of peninsular Thailand. *J Siam Soc.* 49:123–158
- Caro-Consuegra R, Lucas-Sánchez M, Comas D, Bosch E. 2023. Identifying signatures of positive selection in human populations from North Africa. *Sci Rep.* 13(1) ://WOS:000993082000004. <https://doi.org/10.1038/s41598-023-35312-3>
- Cleary DW et al. 2021. The upper respiratory tract microbiome of indigenous Orang Asli in north-eastern Peninsular Malaysia. *Npj Biofilms Microbiomes.* 7(1) ://WOS:000733236200001. <https://doi.org/10.1038/s41522-020-00173-5>
- Culp DJ et al. 2015. Salivary Mucin 19 Glycoproteins INNATE IMMUNE FUNCTIONS IN STREPTOCOCCUS MUTANS-INDUCED CARIES IN MICE AND EVIDENCE FOR EXPRESSION IN HUMAN SALIVA. *J Biol Chem.* 290(5):2993–3008. <https://doi.org/10.1074/jbc.M114.597906>
- Diamond J, Bellwood P. 2003. Farmers and Their Languages: The First Expansions. *Science.* 300(5619):597–603. <https://doi.org/10.1126/science.1078208>
- Endicott P. 2013. Introduction: Revisiting the “Negrito” Hypothesis: A Transdisciplinary Approach to Human Prehistory in Southeast Asia. *Hum Biol.* 85(1–3):7–20. <https://doi.org/10.3378/027.085.0301>
- Evans IHN. 1968. The negritos of Malaya. Psychology Press.
- Garcia OA et al. 2023. The Legacy of Infectious Disease Exposure on the Genomic Diversity of Indigenous Southern Mexicans. *Genome Biol Evol.* 15(3) ://WOS:000952777100002. <https://doi.org/10.1093/gbe/evad015>
- Hamilton A. 2006. Reflections on the ‘Disappearing Sakai’: A Tribal Minority in Southern Thailand. *J Southeast Asian Stud.* 37(2):293–314
- Hill C et al. 2006. Phylogeography and Ethnogenesis of Aboriginal Southeast Asians. *Mol Biol Evol.* 23(12):2480–2491. <https://doi.org/10.1093/molbev/msl124>
- Hublin J-J et al. 2017. New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. *Nature.* 546(7657):289–292. <https://doi.org/10.1038/nature22336>

- Jinam TA et al. 2012. Evolutionary History of Continental Southeast Asians: “Early Train” Hypothesis Based on Genetic Analysis of Mitochondrial and Autosomal DNA Data. *Mol Biol Evol.* 29(11):3513–3527. <https://doi.org/10.1093/molbev/mss169>
- 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. <https://doi.org/10.1093/gbe/evx118>
- Kelly RL. 2013. The lifeways of hunter-gatherers: the foraging spectrum. Cambridge University Press.
- Kennedy A, Ozbek U, Dorak M. 2017. What has GWAS done for HLA and disease associations? *Int J Immunogenet.* 44 <https://doi.org/10.1111/iji.12332>. <https://doi.org/10.1111/iji.12332>
- Kricheff DA, Lukas H. 2015. Being Maniq: Practice and identity in the forests of Southern Thailand. *Hunt Gatherer Res.* 1(2):139–155. <https://doi.org/10.3828/hgr.2015.9>
- Little MA. 2020. Evolutionary Strategies for Body Size. *Front Endocrinol.* 11 ://WOS:000525264900001. <https://doi.org/10.3389/fendo.2020.00107>
- Lopez M et al. 2019. Genomic evidence for local adaptation of hunter-gatherers to the African rainforest. *Curr Biol.* 29(17):2926–2935. e4
- Lukas H. 2004. Can " They" save" Us", the Foragers? Indonesian and Thai Hunter-Gatherer Cultures under Threat from Outside.
- Mahmud MH, Baharudin UM, Isa ZM. 2022. Diseases among Orang Asli community in Malaysia: a systematic review. *Bmc Public Health.* 22(1) ://WOS:000884740900004. <https://doi.org/10.1186/s12889-022-14449-2>
- Migliano AB, Vinicius L, Lahr MM. 2007. Life history trade-offs explain the evolution of human pygmies. *Proc Natl Acad Sci U S A.* 104(51):20216–20219. <https://doi.org/10.1073/pnas.0708024105>
- 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. <https://doi.org/10.1038/ng.3621>
- Rambo AT. 1988. Why are the Semang? Ecology and ethnogenesis of aboriginal groups in Peninsular Malaysia. *Ethn Divers Control Nat Resour Southeast Asia.* 19–35
- Rees JS, Castellano S, Andrés AM. 2020. The Genomics of Human Local Adaptation. *Trends Genet.* 36(6):415–428. <https://doi.org/10.1016/j.tig.2020.03.006>
- Reynolds AW et al. 2019. Comparing signals of natural selection between three Indigenous North American populations. *Proc Natl Acad Sci U S A.* 116(19):9312–9317. <https://doi.org/10.1073/pnas.1819467116>
- Richter D et al. 2017. The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. *Nature.* 546(7657):293–296. <https://doi.org/10.1038/nature22335>
- Sahlan N et al. 2019. Hepatitis B virus infection: Epidemiology and seroprevalence rate amongst Negrito tribe in Malaysia. *Med J Malays.* 74(4):320–325
- Saunders AE, Johnson P. 2010. Modulation of immune cell signalling by the leukocyte common tyrosine phosphatase, CD45. *Cell Signal.* 22(3):339–348. <https://doi.org/10.1016/j.cellsig.2009.10.003>
- Schebesta P. 1927. Bei den Urwaldzwerge von Malaya. FA Brockhaus.
- Single R, Meyer D, Nunes K. 2020. Demographic history and selection at HLA loci in Native Americans. *PLoS ONE.* 15 <https://doi.org/10.1371/journal.pone.0241282>. <https://doi.org/10.1371/journal.pone.0241282>

- Skoglund P, Mallick S, Patterson N, Reich D. 2018. No evidence for unknown archaic ancestry in South Asia. *Nat Genet.* 50(5):632–633. <https://doi.org/10.1038/s41588-018-0097-9>
- Stock JT. 2013. The skeletal phenotype of “Negritos” from the Andaman Islands and Philippines relative to global variation among hunter-gatherers. *Hum Biol.* 85(1/3):67–94
- Stringer C. 2016. The origin and evolution of *Homo sapiens*. *Philos Trans R Soc B Biol Sci.* 371(1698):20150237. <https://doi.org/10.1098/rstb.2015.0237>
- Tasnim AR, Allia S, Edinur HA et al (2016) Distribution of HLA-A, -B and -DRB1 alleles in the Ken-siu and Semai Orang Asli sub-groups in Peninsular Malaysia. *Hum Immunol* 77:618–619
- Tsuneto L, Probst C, Hutz M. 2003. HLA class II diversity in seven Amerindian populations. Clues about the origins of the Ache. *Tissue Antigens.* 62 <https://doi.org/10.1046/j.1399-0039.2003.00139.x>. <https://doi.org/10.1046/j.1399-0039.2003.00139.x>
- Woodburn J. 1982. Egalitarian Societies. *Man.* 17(3):431. <https://doi.org/10.2307/2801707>
- Yengo L et al. 2018. Meta-analysis of genome-wide association studies for height and body mass index in ~700 000 individuals of European ancestry. *Hum Mol Genet.* 27(20):3641–3649. <https://doi.org/10.1093/hmg/ddy271>

AI Statement

All texts were initially written by the author. ChatGPT (OpenAI) was used to support the structuring of the Preamble and Concluding Discussion. Additionally, it was used to check spelling and enhance readability in the Acknowledgements, Abstract, Preamble, Concluding Discussion and AI Statement. DeepL was used to generate an initial translation of the Abstract, which was used for the Zusammenfassung. All texts were subsequently reviewed and approved by the author.