THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Universite Toulouse III - Paul Sabatier

Présentée et soutenue par :
KUSUMA Pradiptajati

Le 14 Septembre 2017

Titre :

In search of Asian Malagasy ancestors in Indonesia

ED Biologie, Santé, Biotechnologies: Anthropobiologie

Unité de recherche :
Anthropologie Moléculaire et Imagerie de Synthèse (AMIS) UMR-5288

Directeur(s) de Thèse :
RICAUT François-Xavier, Chargé de Recherche
LETELLIER Thierry, Chargé de Recherche

Rapporteurs :
FORESTIER Hubert, Professeur, Paris, France
KAYSER Manfred, Professeur, Rotterdam, Pays-Bas

Autre(s) membre(s) du jury :
BRAGA José, Professeur, Toulouse, France
SUDOYO Herawati, Professeure, Jakarta, Indonésie
ADELAAR Alexander, Professeur, Melbourne, Australie
RAKOTOARISOA Jean-Aimé, Professeur, Antananarivo, Madagascar

Résumé

AUTEUR : KUSUMA Pradiptajati

TITRE : *In search of Asian Malagasy ancestors in Indonesia*

DIRECTEUR DE THESE : RICAUT François-Xavier & LETELLIER Thierry

LIEU ET DATE DE SOUTENANCE : Toulouse, le 14 Septembre 2017

RESUME en français

Introduction

L'Indonésie présente une grande diversité linguistique, ethnique et génétique, avec plus de 600 groupes autochtones et 700 différentes langues réparties sur plus de 17 000 îles. Au carrefour de deux océans et deux masses continentales, l'archipel indonésien a été l'objet de migrations humaines successives de la fin du Pléistocène à aujourd'hui. Le dernier évènement d'importance a été la dispersion Austronésienne qui a débuté il y a environ 5 000 ans, probablement depuis Taiwan, pour ensuite se propager à travers les Philippines et l'Indonésie, puis toucher une aire géographique allant de l'Océanie lointaine à l'est, à Madagascar et la côte est Africaine à l'ouest. Malgré de nombreuses recherches en génétique sur la dispersion Austronésienne vers l'est, il y a très peu de données sur la dispersion vers l'ouest, laissant sans réponse de nombreuses questions, liées notamment au peuplement de Madagascar. Reposant sur l'analyse des données culturelles et biologiques, les populations d'Indonésie semblent avoir joué un rôle majeur dans la colonisation de Madagascar, au milieu du premier millénaire de notre ère. Cependant, le peu de populations Indonésiennes étudiées à ce jour n'a pas permis jusqu'à présent d'identifier la/les population(s) indonésienne(s) source(s) ou de proposer un modèle de peuplement. Dans ce présent travail mené en collaboration avec plusieurs équipes internationales, j'ai réalisé des études en génétique des populations de 12 nouvelles populations Indonésiennes, qui *à priori* devraient éclairer l'histoire des migrations austronésiennes dans l'Océan Indien. Parmi elles sont inclus l'ethnie des Ma'anyan du sud-est de Bornéo qui sont les plus proches linguistiquement des Malgaches. En utilisant différents marqueurs génétiques (chromosome Y, ADN mitochondrial et marqueurs autosomaux), ma recherche a amélioré nos connaissances de la diversité génétique Indonésienne, et du lien génétique entre l'Indonésie et Madagascar.

Résultats

L'analyse des marqueurs uniparentaux (chromosome Y et ADNmt) à l'aide d'approches phylo-géographiques et d'analyses des distances génétiques, suggère que les Malgaches proviennent de plusieurs régions d'Indonésie, avec un lien privilégié avec le sud-est de Bornéo, le sud de Sulawesi et les îles de la Sonde. Etonnamment, les Ma'anyan partagent un nombre limité de lignées paternelles et maternelles avec les Malgaches, malgré leur proximité linguistique. Par ailleurs, en combinant l'analyse de fréquences des SNPs et des analyses haplotypique à partir des données autosomales, il a été confirmé que la diversité génétique des Ma'anyan ne correspond pas à l'ancestralité asiatique des Malgaches. Cependant, en centrant l'analyse sur les populations du sud-est de Bornéo, l'origine de l'ancestralité asiatique des Malgaches est ancrée dans la population Banjar, un mélange de population Ma'anyan et Malaise, résultat des activités commerciales de l'empire Malais dans le sud-est de Bornéo, qui se sont poursuivies à travers l'océan Indien. Par ailleurs nos résultats ont aussi permis d'accroitre notre compréhension de la diversité génétique de l'Indonésie en identifiant (1) une nouvelle composante génétique austronésienne présente chez les Ma'anyan, et retrouvée à faible fréquence à travers l'Asie du Sud-Est, suggérant une plus grande complexité du modèle d'expansion austronésien dans la région et (2) le rôle joué par les nomades de la mer dans la structuration de la diversité génétique et les échanges entre populations dans l'Indonésie Centrale, soulignant l'histoire génétique complexe de populations suivant un mode de vie nomade.

Conclusions
L'histoire génétique des îles d'Asie du Sud-Est est plus complexe qu'envisagé précédemment, et de nouveaux événements de l'histoire des populations sont révélés par le présent travail, comme une nouvelle composante Austronésienne et une plus grande compréhension du rôle joué par les derniers nomades de la mer Indonésiens. Par ailleurs, malgré une affinité linguistique, les Ma'anyans ne sont pas la source première des populations Malgaches. Ce désaccord entre données linguistiques et génétiques reflète la complexité de l'histoire génétique des populations du sud de Bornéo. La présente étude propose un nouveau scénario de l'origine asiatique du peuplement de Madagascar qui réconcilie les données génétiques, historiques et linguistiques, scénario dans lequel les Banjar -un métissage génétique entre Malais et Ma'anyan- ont participé aux voyages maritimes menés par les Malais. Ce groupe (Banjar) est, parmi les populations des Iles d'Asie du Sud-Est analysées, la source asiatique la plus probable de la composante asiatique du génome Malgache. Cependant, les routes maritimes exactes de migration de Bornéo à Madagascar restent encore à élucider.

Mots-clés: Indonésie, Madagascar, Migration, ADNmt, Chromosome Y, SNP de l'ADN autosomal

---

RESUME en anglais

Background
Indonesia hosts a wide range of linguistic, ethnic and genetic diversity, comprising of more than 600 ethnic groups and 700 living languages across a total of ~17,000 islands. Acting as a crossroad between two oceans and two landmasses, the Indonesian archipelago has facilitated multi-layered human migrations from the Late Pleistocene to the modern day. The last substantial wave of human migration was the Austronesian dispersal ~5,000 years ago, which is thought to have originated in Taiwan. Its influence spread through the Philippines and Indonesia, ultimately impacting a wide geographical area ranging from Remote Oceania in the east, to Madagascar and the eastern coast of Africa in the west. Despite considerable genetic research on the eastward Austronesian expansion, there is little equivalent research on the western edge, leaving major issues unresolved regarding the settlement of Madagascar. Based on cultural and biological studies, it has been suggested that Indonesian peoples played a major role in the colonization of Madagascar from around the middle of the first millennium CE (Current Era). However, poor geographical coverage of Indonesian populations has prevented the Indonesian source populations from being identified and no clear settlement or migration patterns have been proposed. Here, in collaboration with an international team of colleagues, I performed human population genetic studies on 12 new Indonesian populations, which were *a priori* expected to shed light on the westward migration of Austronesians across the Indian Ocean. This includes the Ma'anyan ethnic group from Southeast Borneo, who are the closest linguistic siblings to modern Malagasy. Using different genetic markers (Y-chromosome SNPs, mitochondrial DNA and genome-wide SNPs), my research has improved the description of Indonesian genetic diversity, and investigated the genetic links between Indonesia and Madagascar.

Results
Based on uniparental markers (Y-chromosome and mtDNA), a combination of phylogeographic and genetic distance analyses suggest that Malagasy derive from multiple regional sources in Indonesia, with a focus on southeastern Borneo, southern Sulawesi and the Lesser Sunda Islands. Interestingly, the Ma'anyan share limited paternal and maternal lineages with the Malagasy, despite their linguistic connection. Furthermore, combining SNP frequency and haplotype-based analyses from autosomal genome-wide data, it was confirmed that the genetic diversity of the Ma'anyan does not match the Asian ancestry of the Malagasy. However, by focusing on Southeast Borneo populations, strong support was found for an origin of the Asian ancestry of Malagasy among the people of Banjar, an admixed population of Ma'anyan and Malay, likely resulting from trading activities by the Malay Empire in Southeast Borneo, and later continuing across the Indian Ocean arena. These results increase our understanding of the genetic diversity across Indonesia by: 1) identifying the unique and undiscovered Austronesian genetic component carried by the Ma'anyan, which occurs at low levels across Island Southeast Asia and suggests a more complex model for the Austronesian expansion in this region, and 2) describing the role played by sea-nomads in structuring genetic diversity and population exchanges in central Indonesia, thus revealing the complex genetic history of populations living this rare nomadic lifestyle.

Conclusions
The genetic history of Island Southeast Asia is more complex than previously thought, and new layers of population history are revealed by the work presented here, including a new Austronesian component and greater understanding of the role played by the last Indonesian sea-nomads. Moreover, despite close linguistic affinity, the Ma'anyan are shown not to be the primary source population of the Malagasy. This discrepancy between linguistic and genetic evidence reflects the complex history of South Borneo populations. These studies provide strong support for a new scenario of the Austronesian settlement of Madagascar by reconciling genetic, historical and linguistic data, in which the Banjar – who carry Malay-Ma'anyan genomic ancestry – were participants in Malay-led voyages. This group is the most probable Asian source for the founding Asian population of the Malagasy gene pool among all the analysed Island Southeast Asian populations. However, the exact maritime route(s) of migration from Borneo to Madagascar still remain an open question.

Keywords: Indonesia, Madagascar, Migration, mtDNA, Y-chromosome, genome-wide SNP

DISCIPLINE ADMINISTRATIVE : Anthropobiologie

Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse (AMIS) UMR 5288.
37 Allées Jules Guesde, Université Paul Sabatier, CP 31000, Toulouse, Midi-Pyrénées,
France.

This page is intentionally left blank

Acknowledgements

I have had a wonderful three and a half years with the Evolutionary Medicine research group (AMIS Laboratory), not forgetting days full of discussions, debates (typically French, never-ending) and productive scientific work at the Evolutionary Medicine research group, AMIS Laboratory. This work, entitled "*In search of Asian Malagasy ancestor in Indonesia*", together with the 5 scientific publications in international peer-reviewed journals, would not have been accomplished without significant contributions from the parties mentioned below.

I thank Dr. Francois-Xavier Ricaut, my supervisor, for his tutoring, dedication, and encouragement. I have really enjoyed working with him professionally. I also thank him for the lunch invitations every time we published a paper. Those were the only days where I could taste proper European dishes and improve my nutritional intake quality, as a foreign student living far away from his homeland.

I thank Dr. Thierry Letellier, my supervisor, for his critical thinking and in depth scientific discussions. Through him I have learnt valuable lessons on the different viewpoints of how to undertake science and research. He is a biochemist, but also excels in population genetics problems, and he has reminded me to focus and understand the core issues of my research that I will do in the future.

I thank Prof. Herawati Sudoyo, director of the Eijkman Institute for Molecular Biology, Jakarta, a woman in science who has pioneered population genetics research in Indonesia. I thank her for her encouragement, trust, and the opportunity she has given me to pursue my studies in France.

I thank Dr. Nicolas Brucato, a cool post-doctoral researcher in our group. I consider him as my third supervisor and he has improved my knowledge to work on technical details. I also thank him for his updates on published population genetics tools and the discussions we have shared together.

I thank Prof. Murray Cox, a computational biology scientist at Massey University, New Zealand, for his support and help on the technical discussions in every article that we have published together. The red ink, comments and suggestions he made on all my manuscript drafts were priceless.

I thank Prof. Alexander Adelaar and Dr. Philippe Grange, linguists from the University of Sydney, Australia, and from Universite de La Rochelle, France, for their support and critical discussions on the linguistic and anthropological aspects of the genetics study I performed. One day, I would really like to go to the field together, to see how these great linguists work and collect data, so I could learn from them.

I thank Dr. Hannes Schroeder, a researcher at the Centre of GeoGenetics, Copenhagen, Denmark, for the opportunity to practice and learn the Capture Method. It was a great pleasure to meet him and the other members of Prof. Thomas Gilbert's team.

I thank Prof. Antoine Gessain and Dr. Olivier Cassar, microbiology/genetics researchers at the Pasteur Institute, Paris, for the time and opportunity they gave me to learn methods on the detection and genetic analysis of HTLV-1 and HHV-8 viruses. I hope that this collaboration between Paris and Jakarta will continue, so that this study on viruses can be replicated in Indonesia.

I thank Laure Tonasso and Stephanie Schiavinato, the laboratory technicians at the AMIS Laboratory, for their laboratory tutorials and all their help on the laboratory work performed in the AMIS Laboratory, Toulouse.

I thank Dr. Safarina G. Malik, Dr. Helena Suryadi, Gludhug Purnomo, Chelzie Crenna, and Isabella Apriyana for their assistance during the sampling campaigns in Indonesia. Also to other colleagues in the Genome Diversity Laboratory, Eijkman Institute for Molecular Biology, Jakarta, for their love and moral support.

I thank Dr. Rebecca Coles for the great assistance during the redactional phase of my thesis writing.

I thank all the members of the AMIS laboratory who have shared these past three years with me and for their support: Dr. Denis Pierron, Dr. Margit Heiske, Dr. Morgane Gibert, Dr. Patricia Balaresque, Dr. Harilanto Razafindrazaka, Dr. Xavier Mata, and my 'companion in arms', Dr. Cesar Fortes-Lima, and Barbara Da Costa. I thank also Prof. Eric Crubezy, director of the AMIS Laboratory, for allowing me to work in the laboratory.

Last but not least, I thank all my family: my wise and strong parents, my beautiful wife Uci, and my three little angels Thariq, Maryam and Asiya, who have given me the power to survive in the wilderness of doctoral study. Thank you for being proud of me and for all your support, which gave me a shoulder whenever I was tired, and raised me back up again. This work is all for them. Kisses~

My doctoral study is financially supported by the Embassy of France in Indonesia through Campus France, also by "Beasiswa Unggulan" of Ministry of National Education of Indonesia.

This page is intentionally left blank

This page is intentionally left blank

"Urip iku urup"

(Javanese philosophy)

(transl.) Life is to illuminate

Live a life that can give benefit to the others around us, despite how small it is,
but never be a nuisance to the others.

This page is intentionally left blank

Table of contents

I. Introduction

1.1. Background

Throughout human history, the Indian Ocean region has been an important route for human migration, dispersal and contact. In the late Pleistocene, southern India was situated in the southern dispersal corridor that facilitated the migration of modern humans across the world (Field et al., 2007; Mellars et al., 2013). Millennia later, during the mid-Holocene, the development of long-distance seafaring technologies placed this region, and particularly its island components, at the centre of new maritime corridors (Beaujard, 2012a, 2012b; Mahdi, 2016). The settlement history and population interactions within this vast region are still not well-understood, and to date no significant human genetic study has been undertaken that focuses on the role played by Island Southeast Asia (ISEA) populations on the westward Austronesian migration from Indonesia to Madagascar.

Indonesia has been a busy intersection for human migrations since the early arrival of modern humans from 'Out-of-Africa' into Papua and Australia about 60,000 – 50,000 years ago (Allen and O'Connell, 2014; Aubert et al., 2014; Barker, 2013; Grenet et al., 2016). The country also was impacted by the dispersal of Austronesian language speakers from south China and Taiwan into the Pacific and Madagascar that started around 4,000 – 3,000 years ago. This latter migration is strongly associated to specific cultural items (Bellwood, 2007; Bulbeck, 2008; Gray and Jordan, 2000; Spriggs, 2007), and is supported by many genetic studies (Jinam et al., 2012; Karafet et al., 2010; Soares et al., 2016; Wollstein et al., 2010). It created the important cultural and biological population diversity that can be seen today in Indonesia; 700 languages are spoken by ~500 ethnic groups (Lewis et al., 2017). Many researchers have investigated the eastward Austronesian migration into the Pacific (Bellwood, 2007; Duggan and Stoneking, 2014; Kayser et al., 2008; Skoglund et al., 2016; Soares et al., 2016; Spriggs, 1995), while the westward migration to Madagascar has remained understudied. Archaeological and paleoenvironmental evidence, such as changes in the biota (tropical rainforest turning into savanna, extinction of native fauna), suggests that people settled on the island of Madagascar, in the western part of the Indian Ocean, only recently (Burney et al., 2004; Gommery et al., 2011). There may have been sporadic early arrivals from Africa around 2,400 years before present (BP) (these populations have unknown connections to the present-day Malagasy, current human populations in Madagascar), but a burst of continuous settlement events is clearly evident around the middle of the first millennium CE (Dewar and Wright, 1993; MacPhee and Burney, 1991). The Malagasy are believed to be the result of admixture between East African Bantu and Southeast Asian Austronesian, according to biological, cultural, and linguistic data (Adelaar, 2009; Ferrand, 1908; Hurles et al., 2005; Serva et al., 2012).

The Malagasy speak the language Malagasy, which is a member of the Western-Malayo-Polynesian group of the Austronesian language family. About 90% of its lexicon comes from the Maa'nyan Dayak language, a language spoken by an ethnic group living in the Southeast Barito region in East Kalimantan and South Kalimantan, on Borneo Island (Dahl, 1951). In Malagasy there are loanwords from Javanese, Bugis, Malay, Sulawesi (all from Indonesia), Bantu (African), Arabic, and Sanskrit (India), which all together compose ~10% of Malagasy word diversity (Adelaar, 2005, 2006). Linguistic data (number of words from Javanese, Malay and Sanskrit transmitted through Malay/Javanese) suggest a close relationship between the Malagasy and Austronesians, due to human population movements, possibly as subordinates (ship crew or slaves) in Malay-led voyages (Adelaar, 2006; Blench, 2007) from Indonesia to Madagascar during the reign of the Srivijaya empire (western Indonesia, 6th - 13th century CE), as depicted in a relief in an Borobudur temple in central Java (Figure 1.1). This hypothesis is consistent with material culture found in Madagascar originating from Southeast Asia (iron-work techniques, outrigger boats, musical instruments such as the stick-zither and xylophone, and the cultivation of 'tropical food kits', such as rice, bananas, yams and taro (Blench, 2014; Crowther et al., 2016).

Several genetic studies have attempted to investigate the relationship between the Malagasy and the people from ISEA. Many researchers have concluded that Indonesia was the Malagasy's Asian ancestor, in correspondence with linguistic and cultural evidence (Hurles et al., 2005, 2005; Razafindrazaka et al., 2010; Tofanelli et al., 2009). A recent study has supported the hypothesis that Asian Malagasy ancestral populations originated from central Indonesia (Java-Kalimantan-Borneo region) (Pierron et al., 2014). Another study, using mitochondrial DNA data, modeled the estimated number of Indonesian female settlers that arrived in Madagascar to be around 30 women (Cox et al., 2012). Despite these studies, there is still missing information that is needed to propose a detailed historical and demographic scenario of the connection between Indonesian and Malagasy populations. None of these previous studies have determined the level of involvement of the Malagasy's linguistic sister group, the Ma'anyan Dayak from Borneo, or have clarified whether linguistic and historical evidence are correlated with genetic data. Here, in collaboration with an international research network, I performed a human population genetic study on new Indonesian population samples, including the Ma'anyan ethnic group from Southeast Borneo, the closest linguistic sibling to the modern Malagasy. This overall goal of this study was to generate new information on the westward migration of Austronesians across the Indian Ocean.

Figure 1.1. A relief in Borobudur temple illustrating the oceanic voyage from Indonesia to eastern Africa during the Srivijaya era in 6th century using triggered-boat.

1.2. Aims and questions

This thesis uses genetic analyses to investigate the demographic history of Austronesian-speaking people and their dispersal in Indonesia, across the Indian Ocean and into Madagascar. The thesis has two main aims:

1. To determine the origin of human genetic diversity in Indonesia.
2. To assess how this diversity impacts the gene pool of populations in the Indian Ocean region.

Genetic diversity in ISEA, especially in Indonesia, is very high (Lipson et al., 2014; Tumongor et al., 2014; Karafet et al., 2010; HUGO Pan Asian SNP, 2009). There are still regions, however, that have not been explored, and thus, potentially more populations that play key roles in the genetic diversity of Indonesia. In this thesis I focus on Bornean populations, because this island has not been investigated thoroughly by researchers studying areas of human history in ISEA, such as archaeology, anthropology, ethnography and population genetics. Moreover, this island exhibits unique anthropological and linguistic connections with other populations in Madagascar, and the exact ancestral population(s) of the Malagasy in Indonesia has not been found. To address the aims of this thesis I have published five articles in peer-reviewed scientific international journals, which are appended in section III (Published articles) of this thesis. The articles address specific questions and use two different genetic tools: uniparental markers (mtDNA and Y-chromosome) and autosomal genotype data, within a large interpretative framework (linguistic, historical, anthropological, and biological). This interdisciplinary study was only feasible through discussions with collaborators to define the research questions, select sample populations, design sampling campaigns, and for data generation and interpretation.

The articles focus on the following specific questions:
Question 1: Do Bornean populations give additional genetic diversity to the diversity of Indonesia?
Question 2: How has the historical influence of western Eurasia shaped Indonesian genetic diversity?
Question 3: Is Malagasy's linguistic sister, the Ma'anyan from Borneo, their direct ancestor? If not, which other population(s) can potentially be the ancestor(s) of the Malagasy?
Question 4: Did Indonesian sea-nomads contribute to Austronesian migration to Madagascar?

1.3. Literature review

1.3.1. History of the people of Island Southeast Asia

ISEA, which encompasses Indonesia, East Malaysia (Borneo) and the Philippines, was first inhabited by modern humans about 70,000 to 60,000 years ago (Allen and O'Connell, 2014; Détroit et al., 2004) (Figure 1.2). Evidence of upper Pleistocene human occupation was discovered in the Philippines in the Tabon cave (human remains dated to 47,000 years ago (Détroit et al., 2004), in the Callao cave (67,000 years ago) (Mijares et al., 2010), and in the Wallacea region (Sulawesi island where Talepu rock art was uncovered and dated at 40,000 years ago) (Aubert et al., 2014). More than 100 archaeological sites in ISEA, however, are from the Holocene era (Neolithic, Bronze age, and later periods) (Spriggs, 2007). Borneo

Island remains a *terra incognita,* although there have been rare archaeological excavations in the area, near the iconic site of the Niah Cave (Sarawak, Malaysian Borneo), which had a modern human presence around 46,000 years ago up to late Holocene (Barker et al., 2007; Harrison, 1957). Recent archaeological studies have found a number of well-dated sites, dated to under 10,000 years ago (Grenet et al., 2016; Plutniak et al., 2014).

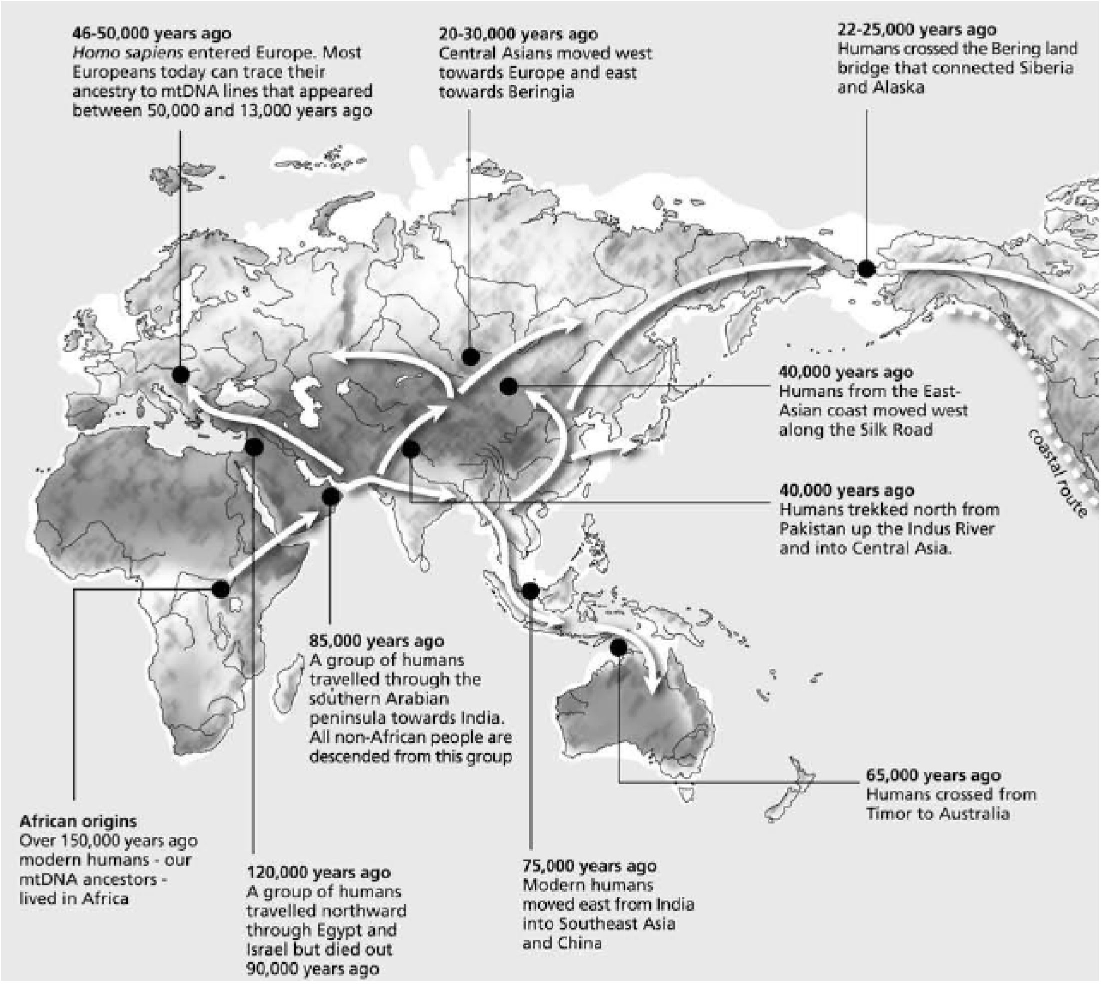


Figure 1.2. An outline of the modern human migration route out of Africa until East Asia and Indonesia, taken from Oppenheimer (2008)

The first modern humans likely settled and crossed ISEA no later than 70,000 years ago. At that time the Southeast Asian archipelago was divided into three main regions: Sundaland (i.e. Sumatra, Java, Borneo, the Malay Peninsula), Wallacea islands and Sahuland (i.e. Australia, New Guinea, Tasmania) (Voris, 2000) (Figure 1.3). Sea level rose 18,000 to 8,000 years ago, shifting coastlines and forming the ISEA topography observed today (Wurster et al., 2017).

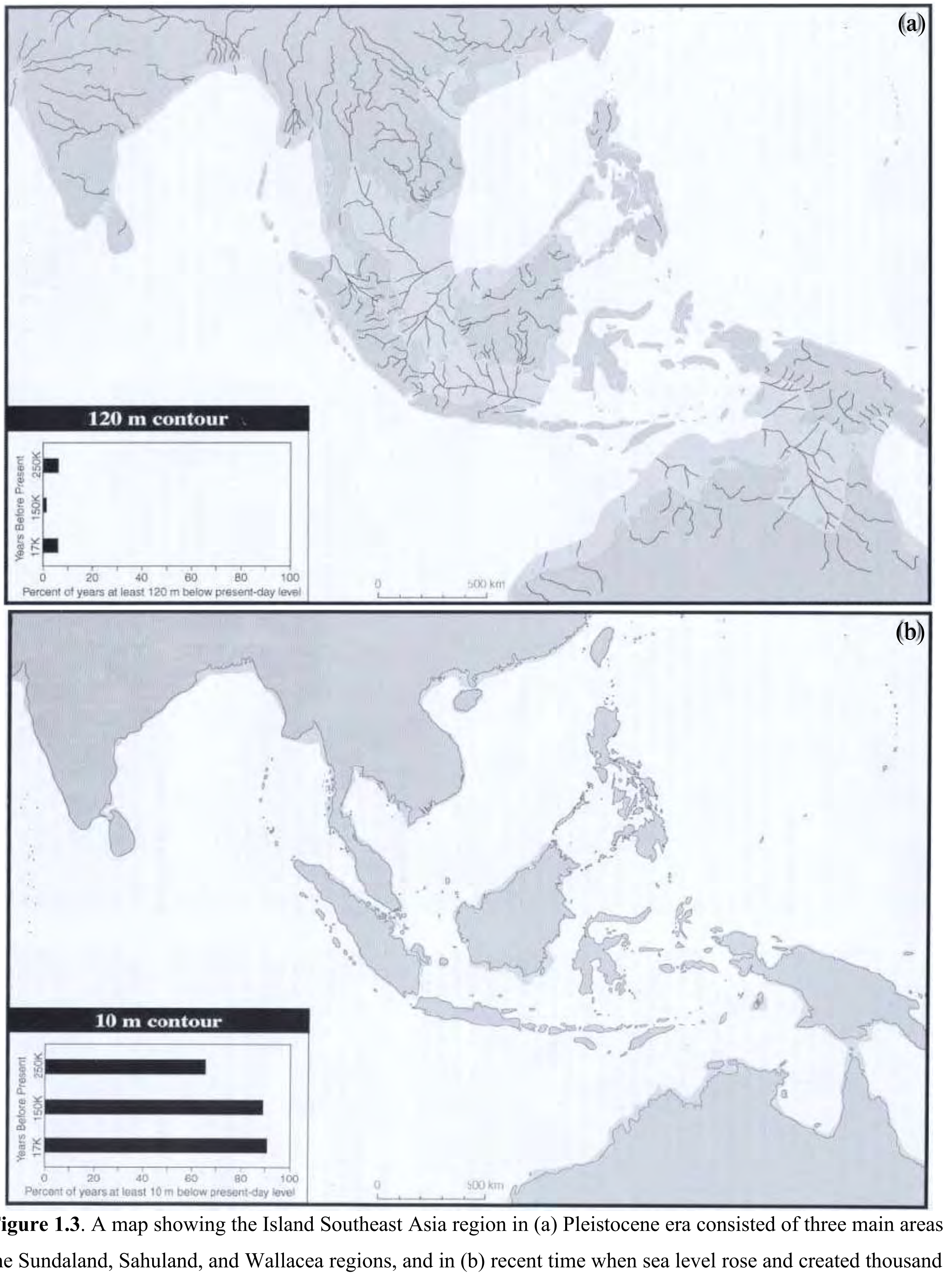


Figure 1.3. A map showing the Island Southeast Asia region in (a) Pleistocene era consisted of three main areas: the Sundaland, Sahuland, and Wallacea regions, and in (b) recent time when sea level rose and created thousands of islands in the Indonesian archipelago and other regions across ISEA , taken from Voris (2000).

The first wave of modern humans arriving in ISEA from Africa in the Upper Pleistocene were called "Australo-Melanesians". Traces of these early settlers, their specific morphological features and genetic profiles, can be seen today in some rare and remote populations, including the Mamanwa and Batak from the Philippines, the New Guineans/Papuans and Aboriginal Australians (Bulbeck, 2013; Malaspinas et al., 2016; Migliano et al., 2013). A recent hypothesis suggests that these early settlers survived through admixture with new settlers during the Neolithic, or by withdrawing into isolated refugia and evolving into so-called 'negrito' populations (Higham, 2013). Nevertheless, their genetic traces and distribution in ISEA is today very scarce. They were hunter-gatherers with a subsistence strategy based on foraging, and were well-adapted to the wide range of environmental conditions they encountered. They occupied three types of sites: rock shelters in the highland regions, river floodplains and coastal sites (Higham, 2013). This subsistence strategy and lifestyle lasted in ISEA for more than 40,000 years, but underwent a major shift starting around 5,000 years ago, driven by the invention and development of crop cultivation and farming. The post-glacial climatic changes led to environmental changes, which made food production in some specific regions of East Asia viable (i.e. 8,000 – 7,000 years ago, Yangtze River valley, south China) (Crawford and Shen, 1998; Z. Deng et al., 2015; Fuller, 2011). This may have led to a sedentary way of life, and an increase in population density leading to population expansion (Bellwood, 2013).

Later, Neolithic human migration occurred during the late Holocene from south China and Taiwan to ISEA. There were two waves of human migration from south China. The first wave was through Thailand and Vietnam, and then through Malaysia to Sumatra and Borneo, suggested by the distribution of archaeological artifacts (basket or cord-marked ceramics) (Anderson, 2005). This has been associated with the early expansion of the Austroasiatic language to the Indonesian archipelago (Higham, 2014; Higham et al., 2015). The second wave was from south China through Taiwan and the Philippines to Sulawesi and Borneo, later reaching Malaysia, as suggested by the distribution of red-slipped pottery (Bellwood et al., 1995; Paz, 2002). This was later associated with the Austronesian expansion "Out-of-Taiwan". These latter events, together with the population substrate from the first settlement wave, are supported by genetic studies showing that present-day ISEA populations have five main genetic layers (Jinam et al., 2012; Karafet et al., 2010; Lipson et al., 2014) (Figure 1.4). These populations were the main ancestors of Indonesian and ISEA populations in general, together with the former group, the Australo-Melanesian group.

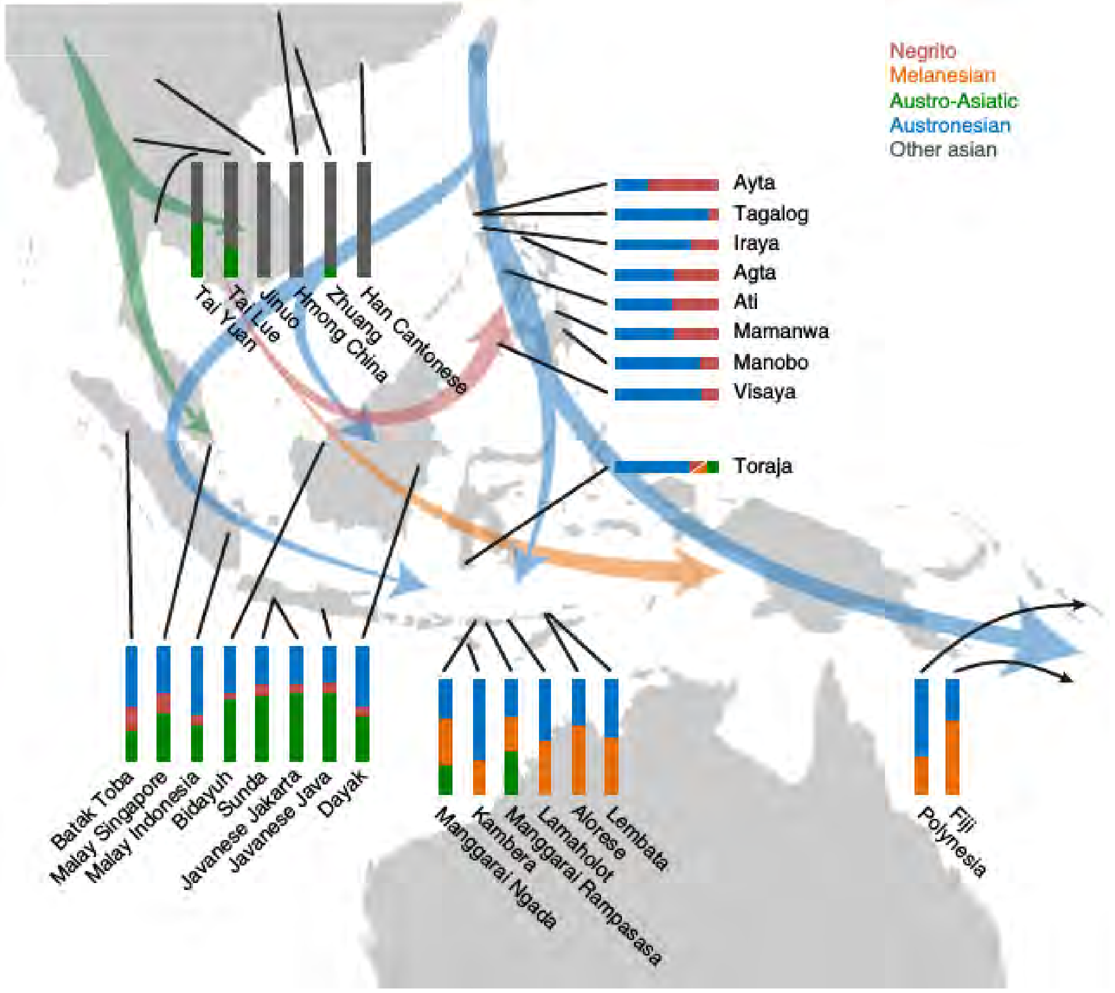


Figure 1.4. Genetic reconstruction of human settlement and migration waves in Island Southeast Asia, taken from Lipson et al. (2014).

The Neolithic migration southward from Taiwan to the Indonesian archipelago, was encouraged by the development of farming in Taiwan and south China, during the late Pleistocene to early Holocene. The Yangtze River basin has been suggested as the region where people first domesticated rice around 8,000 - 5,000 years ago (Crawford and Shen, 1998; Z. Deng et al., 2015; Fuller, 2011). This preceded the significant movement of agricultural populations originating in south China, south to Southeast Asia, and finally into the Pacific Islands, over a time span of about 5,000 years (Bellwood et al., 1995; Bellwood, 2007, 2013). This southward expansion may have only been possible due to maritime foraging and trade, as shown by archaeological data (Bulbeck, 2008). It is widely accepted that this expansion led to almost the entire population residing in the western part of ISEA speaking Malayo-Polynesian languages, a branch of the Austronesian language, which is also spoken by the Malagasy population in Madagascar (Adelaar, 1995a, 2005). In contrast, populations from the eastern part of ISEA speak other branches of Austronesian (i.e. Central-Eastern Malayo-Polynesian) and Papuan languages (Bellwood et al., 1995; Greenhill, 2015; Lewis et al., 2016). There are nine other branches of the Austronesian language family that are only spoken by the indigenous people of Taiwan, the Formosan (Bellwood et al., 1995). Therefore, it was concluded that the Austronesian language developed in Taiwan, then spread to other regions during the Holocene (Diamond and Bellwood, 2003) (Figure 1.5).

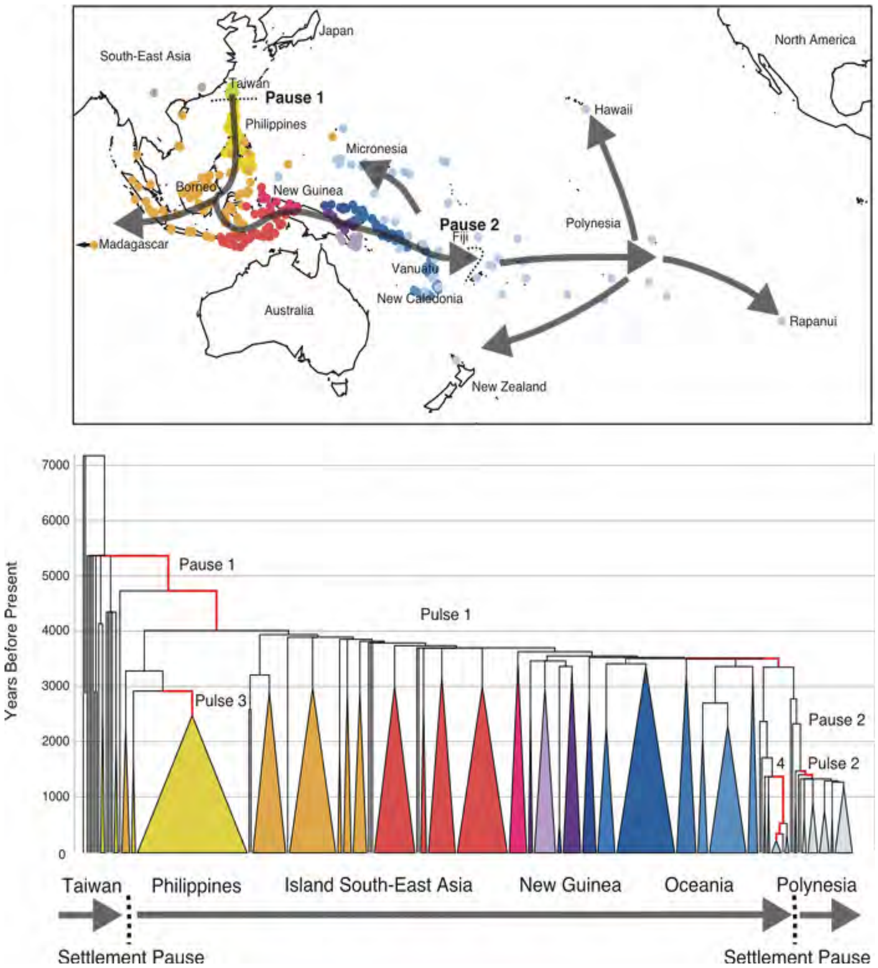


Figure 1.5. Dispersal of Austronesian-speaking people, taken from Gray et al. (2009).

There were subsequent migrations from south and southwestern Eurasia, mainly from India and the Middle East through trading activities. Indianisation and Islamisation in Indonesia led to the development of Hindu, and later Islamic, kingdoms between the 5th and 16th centuries CE. These kingdoms reinforced interactions and influence from western Eurasia to all the regions in western and central Indonesia spanned by these kingdoms (Beaujard, 2012a; Beaujard and Norel, 2014; Gonda, 1975; Kanchan, 1990; Karafet et al., 2010). These interactions brought not only new ideas and trade goods into the Indonesian archipelago, but they also encouraged, to some degree, gene flow, in a number of Indonesian populations. This is one aspect that I will investigate in this study.

1.3.2. Austronesian language

Austronesian is the largest language family in the world and the most geographically widespread of all languages. The dispersal of Austronesian languages over their current geographic range (from the coast of East Africa to the eastern Pacific) is a relatively recent event that started about 4,000 years ago, triggered by Neolithisation in south China and Taiwan (Bellwood, 2007). Currently, Austronesian languages are spoken by native populations inhabiting the territories of Indonesia, Malaysia, the Philippines, Madagascar, Taiwan, Vietnam, southern Cambodia and south China (Adelaar, 2005; Bellwood et al., 1995). In the eastern regions, the language is also spoken by people in the coastal areas of West Papua, Papua New Guinea and the Melanesian Islands (New Ireland, New Britain, Solomon islands, Fiji, and Vanuatu), until Polynesia (Pawley and Ross, 1995). More than 1,200 languages are affiliated to Austronesian language family, which is currently spoken by approximately 300 million individuals.

The Austronesian linguistic family is divided into ten primary groups, of which nine belong to the Formosan language subgroup, found only in Taiwan (Adelaar, 2005) (Figure 1.6). All Austronesian languages outside Taiwan belong to the Malayo-Polynesian subgroup of the Austronesian language. The great diversity of the Formosan language has been interpreted as a product of longer in situ divergence (Blust, 2013). Therefore, the Austronesian language family tree strongly favours Taiwan as the origin of the language.

In Indonesia there are three branches of the Proto-Austronesian language: 1) West-Malayo-Polynesian (WMP) spoken by populations in the western region of Indonesia and also in Madagascar; 2) Central-Malayo-Polynesian (CMP) spoken by populations in the central region of Indonesia; and 3) South Halmahera-West New Guinea (SHWNG) spoken by the majority of inhabitants in Maluku, and Papuans in coastal/lowland areas (Adelaar, 2005; Blust, 1978). Within each of these branches, each region may have different languages. In Indonesia there are in total ~700 different languages. Population dispersal over 17,000 islands (emphasised when sea-level rose in the early Holocene) led to some degree of isolation and linguistic differentiation, thus producing the vast diversity of Austronesian languages seen today (Tryon, 1995).

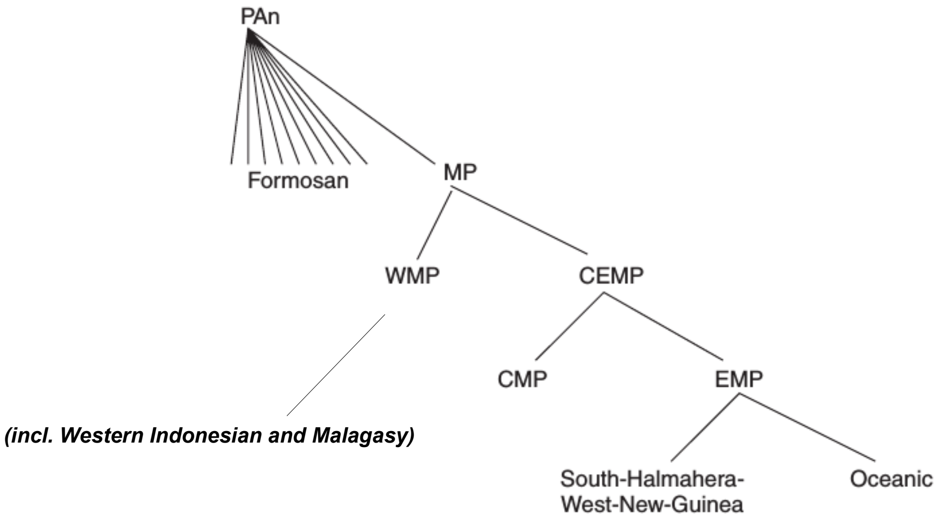


Figure 1.6. Tentative tree diagram of Austronesian language, taken and modified from Adelaar (2013).

1.3.3. Borneo: an overview

Among the ~17,000 islands in the Indonesian archipelago, Borneo, the third largest island in the world, has contributed greatly to the cultural diversity of Indonesia. Borneo hosts some of the first evidence for Late Pleistocene modern human settlement in Sundaland, now ISEA (Barker, 2013; Détroit et al., 2004; Krigbaum and Manser, 2005). The deep skull of Niah cave in Sarawak (Malaysian Borneo), is one of the key specimens that has shed light on human migration into ISEA (Barker, 2013; Harrison, 1957). Rather than resembling Australo-melanesian skull morphology, the deep skull has unique features that are similar to autochthonous people living in Borneo today, the "Dayak" people, who probably originated from an early migration from mainland East Asia to Borneo around 30,000 years ago (Curnoe et al., 2016). This finding may provide evidence for a degree of population continuity in this region of Borneo over the last 40,000 years. When sea-level rose and separated Sundaland into component islands, Borneo was separated from peninsular Malaysia about 12,000 years ago, and from Java and Sumatra around 10,000 years ago (Bird et al., 2005). Changes in geographical conditions with this newly formed coastline stimulated populations to adapt to their new environment (Bulbeck, 2008). Moreover, progressive ponding of the rivers and saline contamination of coastal ground waters and rivers, was likely to significantly affect the ecology of the area, hence encouraging population movement away from water edges (Barton et al., 2013). This factor might have further influenced language and genetic diversity in Borneo.

People in Kalimantan, and the rest of Borneo Island, as well as other major western Indonesian populations, speak the Austronesian language family subgroup called the Western-Malayo-Polynesian group (Adelaar, 1995a; Lewis et al., 2016). There are 74 languages (~11% of total Indonesian language diversity) spoken in Kalimantan (Figure 1.7), clustered into 10 different subgroups: 7 endo-Bornean groups (Land Dayak, Rejang-Baram, Kenyah-Kayan, Apo Duat, West Barito, Barito-Mahakam and East Barito) and 3 exo-Bornean groups (Malayic Dayak, Tamanic and Sabahan) (Adelaar, 1995a, 2005; Robert Blust, 2010). Furthermore, there are also Sama-Bajau languages spoken by the Bajo sea-nomad people, who live in eastern coastal areas of Kalimantan (Lewis et al., 2016; Nuraini, 2016). Although all languages spoken in Borneo are classified as Austronesian, there is a significant influence from the Austroasiatic language substrate (Blench, 2011). Austronesian and Austroasiatic languages are generally recognised as coherent and internally consistent language phyla. However, it has been acknowledged, since Schmidt (1906), that they share some common vocabulary, which has led to proposals merging them in a single phylum. Consequently, similarities between lexical items would be due to inheritance from a common ancestor. However, Borneo and Southeast Asia were part of the same land mass until 12,000 years ago, suggesting that there was late Holocene contact between mainland Asia, hence Austroasiatic speakers, and Borneo. These two hypotheses of an Austronesian-Austroasitic link, a common ancestor and recent interactions can be concomitant. There are three lines of evidence for language contact between Austronesians and Austroasiatics in Borneo: 1) resemblance between mainland Southeast Asia branches of Austroasiatic and Borneo Austronesian; 2) similarities between Asian languages and Borneo Austronesian; and 3) between Austronesian and Austroasiatic due to trade-driven expansion of the Malay (Blench, 2011). This evident connection was observed mainly in populations living in western/northwestern Borneo, such as the Land Dayak speaking group (Adelaar, 1995a), where some cultural and agricultural evidence was observed (taro cultivation, mouth organ musical instrument). Apart Austroasiatic, another influence from mainland Southeast Asia is also observed with the absorbency of Malay words; this is mainly found in the Malayic Dayak speaking group, and also in the non-indigenous Banjar group that speak the Malay language in South Kalimantan.

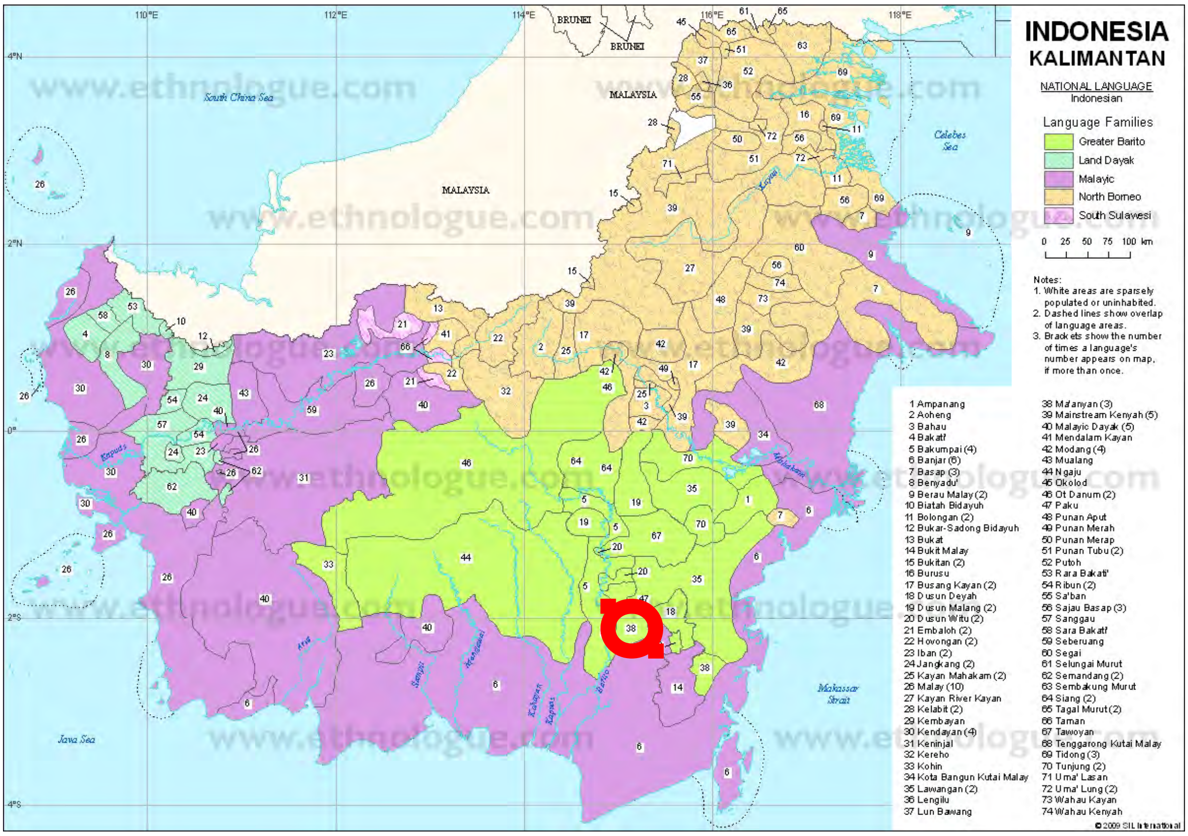


Figure 1.7. Maps of language diversity in Kalimantan, taken from http://ethnologue.com. Red circle indicates where the Ma'anyan (Southeast Barito) speakers live.

Malay language and culture in South Kalimantan was most probably brought by the invasion of the Malay Srivijaya Empire in the region between the 6th to 8th centuries. Historical information on south Borneo is very scarce, but there is one valuable piece of information written in the 'Hikajat Banjar' ('Tale of Banjar'), which describes the history of Banjarmasin city (Ras, 1968). The story explains that there was an old Malay settlement in south Borneo in a region called "Tanjung". Tanjung is now located further up north from the coastal area. However, a record states that during that time south Borneo was divided by a 200 km gulf (Van Bemmelen, 1949). This enabled the Malay Empire to reach the Tanjung area and establish an outpost, which later moved southward to the current city of Banjarmasin (Ras, 1968). The Malay invasion triggered two human dispersal events from south Borneo: the dispersal of proto-Bajo speakers (Blust, 2007) and the oceanic migration into Madagascar (Adelaar, 1989, 2017). The strong influence of the Malay empire in Indonesia brought about the Indianisation of almost the entire archipelago (Beaujard, 2012b; Kanchan, 1990). In parallel to cultural and trading influences, it is also highly probable that the movement of people from South and Southwest Asia to Indonesia occurred.

One unique Bornean language group spoken by people living around the Barito River is the Southeast Barito language. Intriguingly, it is also spoken by the Malagasy people of Madagascar. Around the Barito River, this language is now spoken mainly by the Ma'anyan people, and other smaller groups such as the Samihim, Paku, Dusun Witu and Dusun Deyah (Adelaar, 1995b, 2005; Hudson, 1972). However, there are cultural dissimilarities between the Ma'anyan people and the Malagasy. Some of the Malagasy are wet rice cultivators, while Dayaks are, as a rule, dry rice cultivators. Some Malagasy use outrigger canoes, whereas Southeast Barito Dayaks never do. The Malagasy migration to East Africa presupposes navigational skills, common to some Indonesian populations, but which can hardly be attributed to the Dayaks, who are forest dwellers today. These factors can be accounted for if we adopt the hypothesis that the Southeast Barito migrants did not cross the Indian Ocean themselves in order to colonise Madagascar, but that they were brought there as subordinates (slaves, ship crew, laborers) by the Malays (Adelaar, 1995a, 2005). It is important to resolve this issue; this issue it is investigated and discussed in this thesis with genetic evidence.

The huge linguistic diversity in Borneo is also enriched by the different lifestyles of its populations. Today, the majority of autochthonous populations make a living through agriculture, but a few Dayak sub-ethnic groups, like the Punan Dayak, have retained their original lifestyles as hunter-gatherers (Sellato, 1994). The later migration waves of Malay and Sulawesi people to Kalimantan added to the diversity of languages and cultures of the Kalimantan people (Adelaar, 1995a). Apart from geographical distance, different languages and cultures would have been barriers between populations. This resulted in a unique genetic background for each population, and consequently the vast human genetic diversity of Kalimantan today. For example, the Bidayuh, living in northwestern Kalimantan, have a significant genetic influence from mainland Southeast Asia, which corresponds to their linguistic characteristics (Jinam et al., 2012). On the other hand, Murut and Dusun groups, living in northeastern Kalimantan are genetically related to the Formosan (Austronesian speaking) in Taiwan and the Igorot people in the Philippines, as they live in geographic proximity to the Philippines (Mörseburg et al., 2016). The Tamanic people, living in central-northern Kalimantan, speak a language close to Buginese and other Sulawesi languages (Adelaar, 1995a), and may also have Sulawesi genetic influences in their understudied genome. Despite these recent studies, there remain many understudied populations in Kalimantan and its human genomic diversity is largely unknown. To date, there are extensive genetic studies at the regional level (Karafet et al., 2010; Lipson et al., 2014; Tumonggor et al., 2013), but Kalimantan populations were not included in these studies. In order to fill these gaps in knowledge of the region, this thesis aims to investigate human genetic diversity in Kalimantan and its influence on other populations.

1.3.4. Arrival of Austronesians in Madagascar

The populations on the island of Madagascar have strong cultural, linguistic and genetic connections to both Africa and ISEA (Adelaar, 1995a, 2005, 2009; Beaujard, 2012b; Hurles et al., 2005; Soodyall et al., 1995). It is well-documented that the main phase of settlement took place around 2,000 – 1,000 years ago (Burney et al., 2004; Hurles et al., 2005; Pierron et al., 2014), but there is also evidence for earlier (4,000 – 3,000 years ago) sporadic human presence on the island, with unknown relationships to the present-day Malagasy and hunter-gatherers, most likely, from Africa (Burney et al., 2004; Dewar and Wright, 1993). Cut marks were found on hippopotamus bones in northwestern Madagascar, supporting a human presence around 2,400 years ago (MacPhee and Burney, 1991), as well as environmental changes (deforestation) (Burney et al., 2004), dated from around 2,000 years ago. Furthermore, there are reports stating the presence of anthropic artifacts around 4,000 years ago (Gommery et al., 2011). Decades of archaeological surveys and excavations have not found an ISEA presence in Madagascar. However, the latest archaeobotanic study showed an early Asian presence in the eastern Africa region, particularly in the Comoros Islands and northern Madagascar, at around the early 7th and 11th century CE respectively (Crowther et al., 2016) (Figure 1.8). Researchers found fossil Asian crops (rice and cotton) in those regions. This finding rises many further questions and drove research on how, when, and which Asian people brought these ancient crops to the East of the Indian Ocean, and ultimately participated in the settlement of Madagascar.

The presence of Asian people in Madagascar is strongly supported by linguistic research. The Malagasy are subdivided into 18 different ethnic groups, and this roughly corresponds to the division of dialects (Rasoloson and Rubino, 2005). Central dialects and southwestern dialects form two clearly separate groups, while the other dialects are more difficult to cluster (Adelaar, 2005). The most important groups, in order of their numbers, are the Merina (northern central mountains), the Betsimisaraka (along the east coast), the Betsileo (southern central mountains), the Tsimihety (northern mountains), the Sakalava (western savannas) and the Antandroy (southern dry lands).

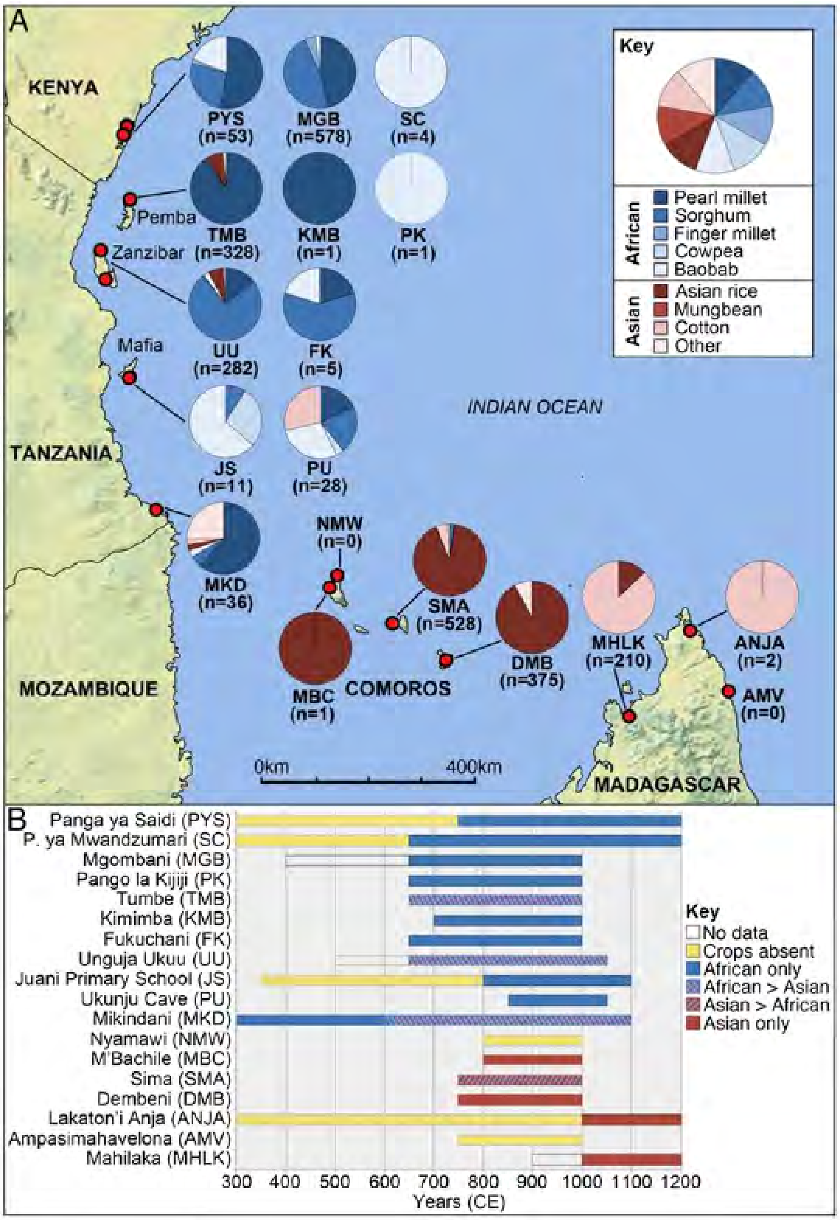


Figure 1.8. A map of eastern Africa with archaeobotanical findings showing Asian traces in eastern Africa as early as mid 7th century, taken from Crowther et al (2016).

All ethnic groups speak Malagasy, but with different dialects. The differences between the Malagasy dialects are small enough to support the arrival of linguistically similar migrant groups (Adelaar, 2005; Rasoloson and Rubino, 2005). The Malagasy language is part of the Western-Malayo-Polynesian (WMP) language of Austronesian. Most of the Malagasy language lexicon comes from Maa'nyan Dayak languages, languages spoken by tribes inhabiting the region of the Barito River in East Kalimantan and South Kalimantan (Dahl, 1951). The Ma'anyan language has striking phonetic, grammatical and lexical similarities with Malagasy, and is the most closely related language to Malagasy. Words are mostly found in terms associated with the domain of navigation and parts of the body. Loanwords from Javanese, Bugis, Bantu (African), Arabic and Sanskrit are found in lower frequency in the Malagasy language (Adelaar, 2006), however Malagasy has many borrow-words from Malay, which appear to be from two different dialects: Sumatra (Srivijaya?) Malay and Banjar Malay. For example, Sumatran Malay has the following words: *səmbah* 'gesture of worship or honour'; *ləmah* 'weak'; *cecak* 'lizard'; and *kəmbar* 'twins'. Banjar Malay, which does not have the vowel *ə* and uses *a* instead, has the corresponding forms sambah, lamah, cacak and kambar (all with same meanings). These words were borrowed into Malagasy as *samba* 'expression of gratitude to God, *lama* 'weak', *tsatsaka* 'lizard' and (sakalava) *hamba* 'twins' (Adelaar, 2006). Sanskrit words in Malagasy were probably also brought via Malay/Asian people. Some linguists suggest that the close relationship between Malagasy and Austronesian, especially in Indonesia, is strong evidence for an Austronesian-speaking people migration from Indonesia to Madagascar, as ship workers or slaves (Blench, 2007), possibly during the reign of the Srivijaya Malay empire in the 6th - 8th centuries (Beaujard, 2012a, 2012b). The Malay influence on the Malagasy language would have lasted until after the introduction of Islam in Southeast Asia, as evidenced by the term 's*ombily*' which has a Malay origin to designate the ritual oxen slaughter (Adelaar, 1995a).

Apart from linguistic evidence, there are also cultural, material, and food plant transmission documentation from Indonesia to Madagascar, and eastern Africa in general (Blench, 1996). Food plants, such as taro and banana species that are specifically of Southeast Asian origin, are widely present in Madagascar and in off-coast small islands, indicating anthropogenic translocations (Boivin et al., 2013). Furthermore, the existence of outrigger-canoes in eastern Africa that are widely practiced by sea-faring Vezo people in Madagascar, signifies the distinct introduction of Indonesian maritime technology (Adelaar, 1995a; Blench, 1996; Lambek, 1998). Indonesian ethnomusicological evidence in Madagascar is also evident. The stick-zither is a characteristic instrument of Indonesia, especially in South Sulawesi and Sumba (Blench, 1996) (Figure 1.9). In Madagascar, the stick-zither's one or two strings are made from several twisted raffia fibers, although more recently, steel strings are also used.

Zithers are played only by men and not everyone is allowed to play them. Knowledge is passed down from father to son (McLeod, 1977).

Figure 1.9. The stick-zither musical instrument found in Sumba, Indonesia (left) and in Madagascar (right), photo courtesy of Roger Blench (http://www.rogerblench.info/Images/Artefacts/SEAsi/Indonesia/)

In contrast to the extensive studies focusing on the linguistic and cultural relationships between Malagasy and Austronesian populations, there have only been a few studies investigating their genetic relationship, although there is strong support for a genetic influence of Austronesian populations from ISEA in Madagascar. In addition to a major African (Bantu) genetic input, there is at least a 30% Asian (Austronesian) input into the Malagasy gene pool (Pierron et al., 2014; Tofanelli et al., 2009), which has an affinity to populations located in the present-day Borneo-Java-Sumatra region in Indonesia (Hurles et al., 2005; Pierron et al., 2014). The Malagasy even exhibit a unique Asian genetic signature, the B4a1a1b mitochondrial haplogroup called the Malagasy motif (Razafindrazaka et al., 2010). This haplogroup is a subgroup of the Polynesian motif, a haplogroup marker of Polynesian, Oceanian and eastern Indonesian populations that evolved along the way of the south- and eastward Austronesian migration (Duggan et al., 2014; Duggan and Stoneking, 2014; Soares et al., 2011). Interestingly, however, the Malagasy motif haplogroup is only found in Madagascar (Cox et al., 2012; Razafindrazaka et al., 2010), and it is within this context that this thesis aims to further investigate the ‘missing’ genetic link between Indonesia and Madagascar.

Indonesia and Madagascar are separated by ~6,000 km of sea; historically, humans could have only made this long-distance oceanic voyage with the help of nature. Oceanic south-equatorial currents connect Indonesia and northern Madagascar and Mozambique, and the north-equatorial current enables sea travel between Indonesia and the Horn of Africa. Seafaring simulations using a computer simulated approach suggest that the 6,000 km (3,200 nautical miles) between Indonesia and Madagascar were not a significant boundary (Fitzpatrick and Callaghan, 2008) (Figure 1.10a).

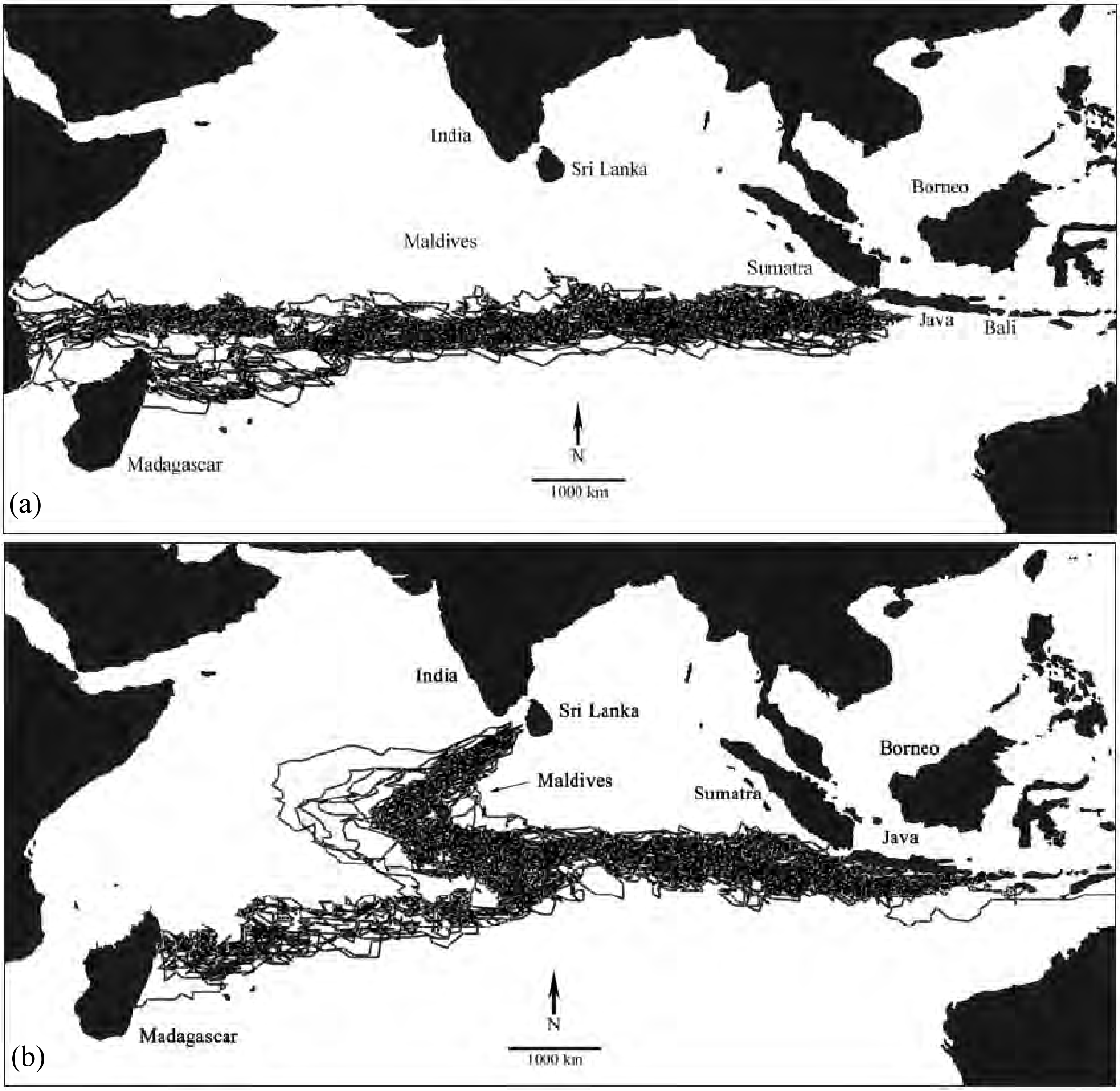


Figure 1.10. Possible direct (a) and downwind (b) sailing from Indonesia to Madagascar and eastern Africa in March and January respectively from computational simulation, taken from Fitzpatrick & Callaghan (2008).

Instead, simulation models support several direct migration routes from Sunda/Sumatra to Madagascar, or an indirect migration route through south India/Sri Lanka, and suggest the Maldives as a possible en route stopping location between ISEA and Madagascar (Fitzpatrick and Callaghan, 2008) (Figure 1.10b). Thus, long-distance oceanic voyage was feasible, as supported by oceanic currents and advanced sea-faring technologies.

1.3.4. Genetic markers as tools to study human history

1.3.4.1. Mitochondrial DNA

Mitochondrion (plural: mitochondria) is an organelle that has important roles in the life of a cell, including cellular respiration, signaling, cell cycle and apoptosis control. Mitochondria have their own genome, distinct from the cell's nucleus. Every eukaryotic cell has hundreds to thousands of mitochondrial genome copies. Mitochondrial DNA (mtDNA) is inherited maternally. Males cannot transmit their mitochondria because during conception sperms' mitochondria, which are located on the neck, do not penetrate the ovum with the sperms' head (Gilbert and Singer, 2006). This prevents recombination event in the mitochondrial genetic material, although there are some rare cases of these events (Kraytsberg et al., 2004). These special characteristics of mitochondria are used in anthropological genetic studies to trace the history of maternal lineages. Mutations that occur in the mitochondrial genome can cause diseases (McFarland et al., 2010; Sudoyo et al., 2003; Wallace, 1994), however, other specific mutations can be used as female-specific population markers for investigating human migration history (Kivisild, 2015; Pierron et al., 2008; Stoneking, 2008).

MtDNA is a circular double-stranded molecule. It has 16,569 base pairs (bp), which is only 0.0006% of nuclear DNA in length. MtDNA codes for 37 genes: 2 ribosomal RNAs (rRNAs), 13 subunits of the oxidative phosphorylation (OXPHOS) system, and 22 transfer RNAs (tRNAs) (Anderson et al., 1981). Within the genome there are ~1,100 bp of nucleotides that are not translated and have regulating functions. This region is called the control region. This region has been exploited widely for applications in population genetic studies. The control region contains three hypervariable segments (HVS), called HVS-I, HVS-II, and HVS-III, and also includes a short structure called the displacement loop (D-Loop) (Clayton, 2003). The control region has a very high nucleotide mutation rate, reaching 0.165×10^{-6} substitutions/site/year, while in the coding region the rate is only 0.017×10^{-6} substitutions/site/year (Pakendorf and Stoneking, 2005). This high sequence variability creates specific polymorphism patterns in individuals within a population that are often different from individuals in other populations. This feature makes mtDNA a prominent marker in population genetic studies and allows robust phylogenetic and phylogeographic studies.

Variation in mtDNA sequences is correlated with population geographical distributions and their histories (Cann et al., 1987). Polymorphisms in mtDNA nucleotides produce haplotypes that can characterise a population or group of populations, forming a haplotype group called a haplogroup (which includes mtDNA sequences sharing a common mtDNA ancestor). Currently 5,400 haplogroups have been identified worldwide (http://phylotree.org). The study of the geographic distribution of these haplogroups is called phylogeography. The history of haplogroup divergence and its evolution through time (based on the mutation rate we can obtain 'time to the most recent ancestor' (TMRCA)) is called phylogeny. MtDNA gives information about maternal lineages, providing a complementary picture to the scenarios that are built from the analysis of Y- or autosomal chromosome data (Kayser et al., 2008; Tobler et al., 2017).

The nomenclature of mtDNA haplogroups is firstly based on letters, and secondly on numbers to represent different and/or deeper groups. A comparison of mtDNA variation in global populations showed that modern humans originated in Africa; this region has the highest mtDNA diversity, confirming the 'Out-of-Africa' scenario (Macaulay et al., 2005; Pagani et al., 2015; Rasmussen et al., 2011). The African signature haplogroup is classified as the L haplogroup, and has the highest degree of variation observed in Africa lineages. This haplogroup is divided into the sub-haplogroups L0, L1, L2, and L3. Outside Africa, specific Eurasian macrohaplogroups are classified as the M and N haplogroups, which evolved from the L3 haplogroup. European populations are classified as I, J, K, H, T, U, V, W, and X haplogroups. In ISEA we found lineages (E, B, F, M and N) originating in Sundaland (previous landmass located west of the Wallace Line) and other haplogroups (S, O, P, M, and Q) from the Sahul (Australia, Papua New Guinea, Melanesia islands) which both evolved in isolation from each other over the last 50,000 years. Currently, the mtDNA nomenclature has a robust branch structure, which has been constructed through the rigorous and detailed analyses of whole mtDNA genomes from more than 30,000 individuals (van Oven and Kayser, 2009).

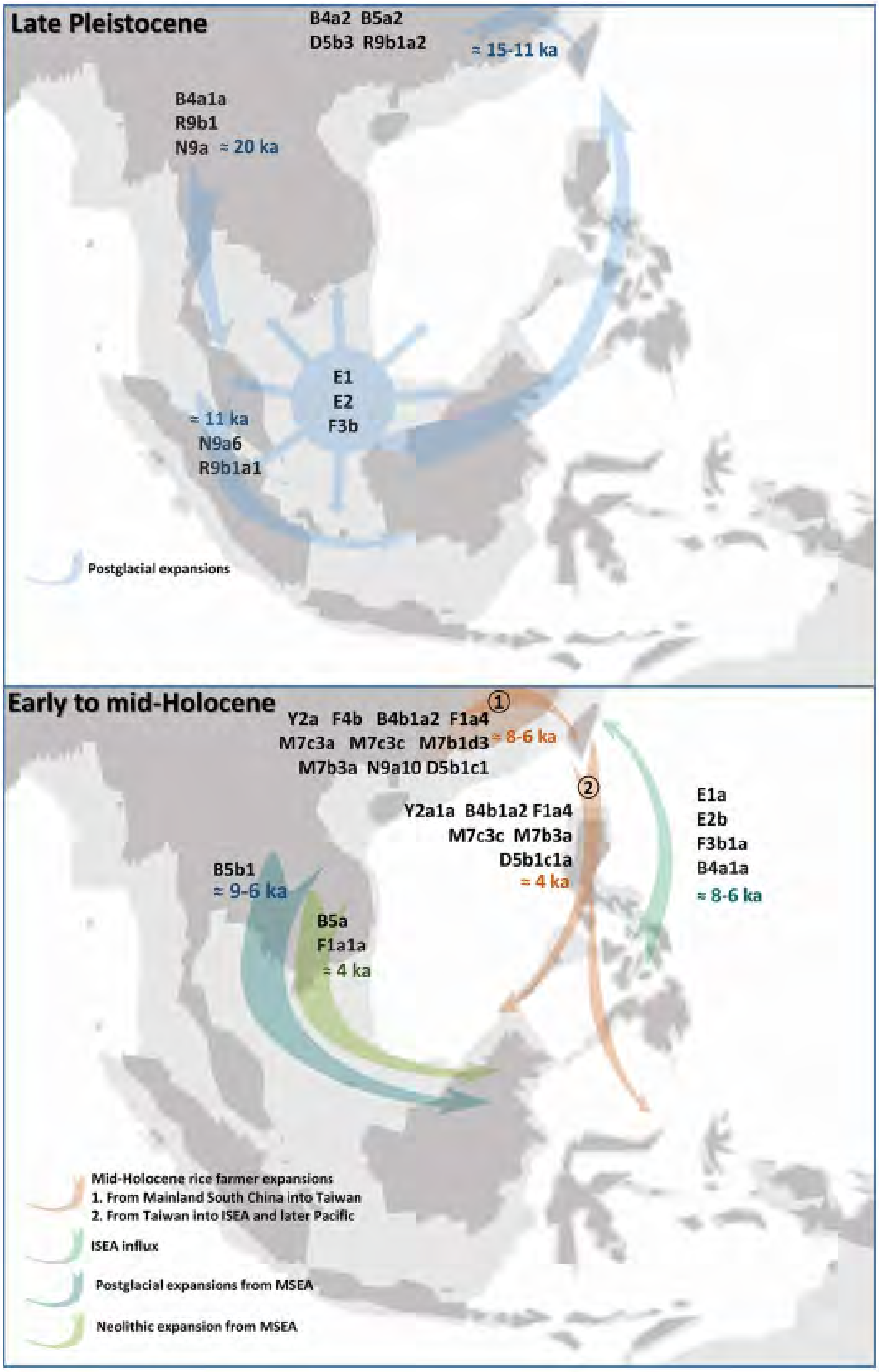


Figure 1.11. Maps of East Asian mtDNA lineages distributions and their migration scenarios through the Indonesian archipelago, taken from Brandao et al. (2016).

A number of haplogroups are found in the ethnic populations scattered throughout the Indonesian archipelago, the majority of which are thought to be related to small-scale late-Holocene migrations, notably the Neolithic Austronesian dispersal out-of-Taiwan (i.e., haplogroup B4*, M7b*, M7c*, and F1a4*) and earlier Neolithic dispersals from mainland

Southeast Asia (B5a*, B5b*, F1a1*) (Bellwood, 2007; Brandão et al., 2016; Soares et al., 2016). These two events were preceded by ISEA population dispersals triggered by rises in sea-level after the Last Glacial Maximum 14-8 kya (E and F3* haplogroup) (Brandão et al., 2016; Hill et al., 2007; Oppenheimer and Richards, 2001) (Figure 1.11). In eastern Indonesia, populations carry haplogroups affiliated to Papuan populations (i.e. haplogroup P and Q). These haplogroups are also found in Austronesian and Papuan speaking populations, suggesting complex interactions driven by social behavior, language shifts, and admixture between Austronesian and Papuan speaking populations (Brandão et al., 2016; Hill et al., 2007; Kayser et al., 2003, 2008; Tumonggor et al., 2013; Wilder et al., 2011). Meanwhile in Madagascar, Asian mtDNA haplogroups (such as B4a1a1, B4a1a1b, F1a1a, and M7c1) are found more frequently than African mtDNA haplogroups (L*), at a roughly 2:1 ratio (Capredon, 2011; Razafindrazaka, 2010; Tofanelli et al., 2009). Interestingly, a unique ISEA haplogroup has been identified in Malagasy populations but not in other populations – a haplogroup classified as B4a1a1b (Malagasy motif), a sub-branch of the Polynesian motif (Cox et al., 2012; Razafindrazaka et al., 2010).

In conclusion, in the era of whole nuclear genome sequencing, mitochondrial genomes continue to be highly informative and are a widely used tool in human evolutionary and population genetic studies. MtDNA has offered many opportunities to explore genealogical relationships among individuals, and to study the frequency differences of matrilineal lineages among human populations at continental and regional scales, thus, assessing the female-specific aspects of human demographic history (Kivisild, 2015).

1.3.4.2. Y chromosomes

The Y-chromosome is a male-unique chromosome; it has specific roles as a sex-determining chromosome and in male fertility. Compared to other chromosomes, it is the smallest in size and has the least number of genes (Jobling and Tyler-Smith, 1995). The Y-chromosome is believed to have been a degenerated chromosome during its evolution (Charlesworth, 1996); there are nine genes located on the short arm, and the remaining genes are located on the long arm (Skaletsky et al., 2003). The human Y-chromosome is normally unable to recombine with the X chromosome, except for small pieces of pseudoautosomal regions at the telomeres. The bulk of the Y-chromosome is called the non-recombining region of the Y-chromosome (NRY), and it has little functional genetic content. The recombination suppression mechanism in the Y-chromosome promotes its gradual degeneration, which is a common feature of non-recombining sex chromosomes (Charlesworth and Charlesworth, 2000; Steinemann and Steinemann, 1998). The Y-chromosome is an important tool in population genetics and forensic research.

In human population genetics research, three types of polymorphisms are observed in the Y-chromosome: insertions-deletions, SNPs and microsatellites (Hammer et al., 1998; Jobling and Tyler-Smith, 1995; Karmin et al., 2015; Underhill et al., 2001). These polymorphisms have been widely used to investigate male phylogeography and global migration. Insertions and deletions are normally found at specific locations along chromosomes, known as Alu insertions (Hammer et al., 1998). SNPs are also present along the NRY. Microsatellites, also called short-tandem repeats (STR), are short sequences of nucleotides that are repeated several times in tandem. Due to the Y-chromosome's non-recombining nature, these three polymorphisms are often found physically linked to each other, and researchers use this feature to define a Y-chromosome's haplotype. Determining Y-chromosome haplotype diversity in different populations is essential in reconstructing the history of global human settlement from a male perspective (Jobling and Tyler-Smith, 2003; Kivisild, 2017; Underhill et al., 2001; Underhill and Kivisild, 2007).

A comprehensive description of Y-chromosome genetic diversity, using high-coverage next-generation sequencing (NGS), has been proposed for many Y-chromosome phylogenies worldwide (Hallast et al., 2014; Karmin et al., 2015; Poznik et al., 2013; Trombetta et al., 2015), and more specifically in the African continent (Cruciani et al., 2011; Scozzari et al., 2014). Like mtDNA haplogroups, Y-chromosome haplogroups are also represented by alphanumerical codes. Around 800 haplogroups have been identified in the minimum Y-haplogroup tree (http://phylotree.org/Y/), but there are thousands of Y-haplogroups in total (https://isogg.org/tree/ISOGG_YDNATreeTrunk.html). The major haplogroups A, B, and E, are found across all the continents, being associated with ancient African populations and later with specific Bantu dispersals. In European populations, the sub-haplogroup R1 is highly represented, while in Native American populations the sub-haplogroup Q is highly represented (Battaglia et al., 2013). Sub-haplogroups K and O are found in East Asia (Karafet et al., 2010).

A specific scenario of male migration to Indonesia has been proposed (Karafet et al., 2010): four migration phases within different time-frames, marked by the presence of specific Y-chromosome haplogroups or lineages. The haplogroups C-M38, S-M230 and M-P256 represented the first male arrivals into Indonesia (50,000 – 45,000 years ago); these haplogroups are now commonly found in Papua and eastern Indonesian. The influx of various O haplogroups into Indonesia from mainland Asia (O-P203, O-M95, O-M122, O-P201, O-M119) occurred in the late Pleistocene (35,000 – 8,000 years ago), and later from Taiwan with Austronesian haplogroups (O-P203, O-M110, P-P201) (6,000 – 4,000 years ago). Subsequent migrations from the Indian subcontinent and the Middle East into Indonesia were marked with the presence of H-M69, R* and L-M20 haplogroups during the historical era (Figure 1.12).

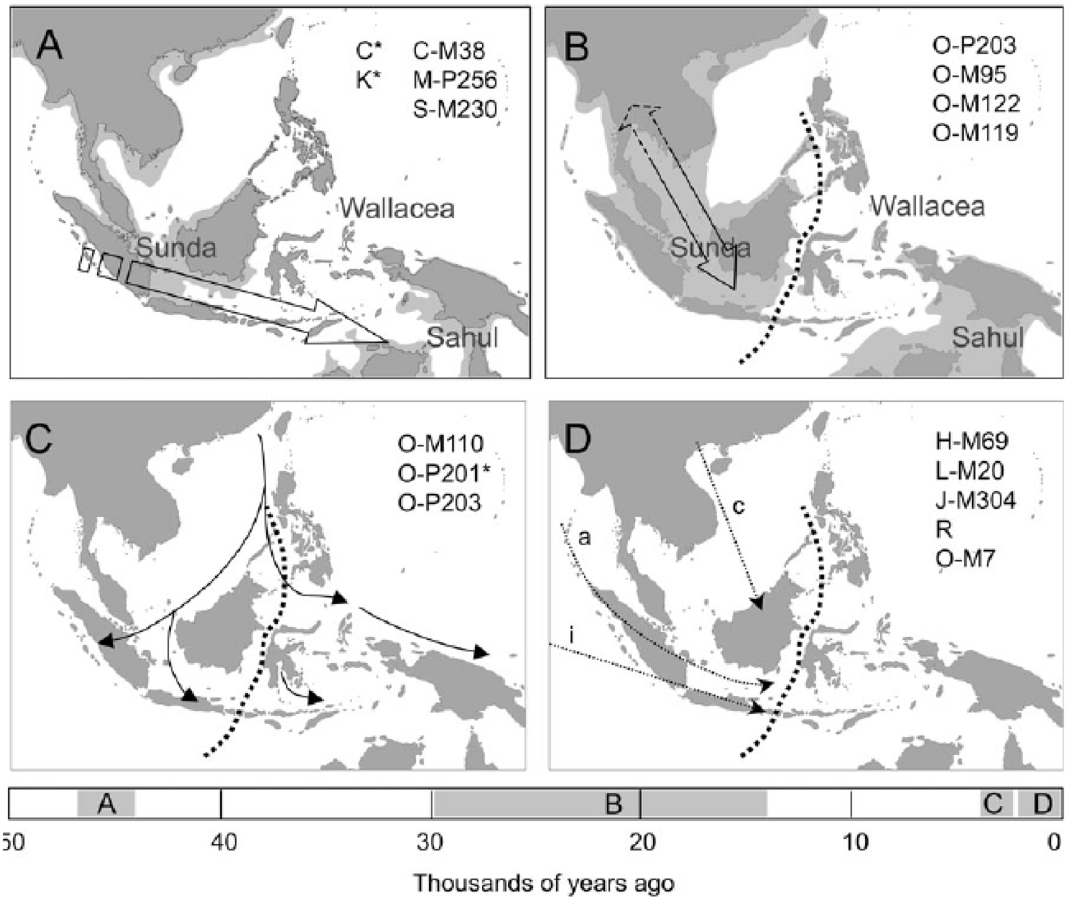


Figure 1.12. Maps illustrating the phases of human settlement in the Indonesian archipelago over 50,000 years taken from Karafet et al. (2010).

In Madagascar, in contrast to the distribution of mtDNA haplogroups, African Y-chromosome haplogroups (B2*, E2*, and E3*) are more frequent than Asian haplogroups (O1a* and O2a*) at a 2:1 ratio, with a minor input from the Middle East with the presence of haplogroup T* and J* (Capredon, 2011; Razafindrazaka, 2010; Tofanelli et al., 2009). This indicates sex-bias origins of Asian and African lineages in Madagascar. By the further study and comparative analysis of mtDNA and Y-chromosome uniparental markers, complementary scenarios of the parental ancestries of the studied populations will be generated in this thesis.

1.3.4.3. Autosomal SNPs

Technological and scientific developments over the last decades have made large-scale, genome-wide investigations feasible and cost-effective. Today, Next Generation Sequencing (NGS) and genome-wide SNP screening is readily accessible to population genetics and clinical genomic research worldwide (Goodwin et al., 2016). Diverse platforms (such as Illumina and Affymetrix) are available to survey a wide spectrum of genetic markers, to investigate human origins and population settlement histories.

Genome-wide data from SNP arrays/sequencing produce large public databases, which are analysed with specific bioinformatic tools to detect demographic/migratory events in population histories (at continental, regional and local scales) and selection signals. For instance, HapMap (International HapMap Consortium, 2010) and HGDP (Li et al., 2008) studies have mapped population diversity in four different continents (Africa, South and East Asia, Europe, and South America). Data from the 1000 Genomes Project was another milestone; the project performed whole-genome sequencing on 2,500 individuals from populations across the world (The 1000 Genomes Project Consortium, 2015). By screening a wide diversity of human population genomes, the results of these large projects help the interpretation of more focused studies. Currently large endeavors are being made to characterise population substructure and history, notably within Africa (Busby et al., 2016; Henn et al., 2011), Europe (Novembre et al., 2008), the Near East (Broushaki et al., 2016); India/South Asia (Moorjani et al., 2013; Reich et al., 2009), East Asia including Indonesia (Aghakhanian et al., 2015; L. Deng et al., 2015; HUGO Pan-Asian SNP Consortium et al., 2009; Lipson et al., 2014; Mörseburg et al., 2016), South America (Skoglund et al., 2015), and Oceania (Duggan and Stoneking, 2014; Skoglund et al., 2016; Wollstein et al., 2010). Population genetic studies focusing on Indonesia and Madagascar, using a SNP array technology approach, are still relatively limited (Cox et al., 2016; Pierron et al., 2014; Xu et al., 2012).

Such genetic studies depend crucially on the accurate and unbiased inference of ancestry, at the genome-wide level and at each locus in the genome, for a bi-fold analysis approach (Pasaniuc et al., 2013). Determining the ancestry of a population has applications ranging from the inference of population history to the estimation of population structure (Novembre et al., 2008; Rosenberg et al., 2003), population stratification and admixture history (Patterson et al., 2012), such that most current global populations can trace their ancestry to more than one distinct ancestral group. Populations that result from the admixture of different populations are referred to as admixed populations. For example, African-Americans have 80% African ancestry and 20% European ancestry (International HapMap Consortium, 2010). In Indonesia, using sets of Ancestry Informative Markers (AIM) SNPs, Indonesian populations can be divided into Papuan and Austronesian/Asian components, forming east-to-west gradient (Cox et al., 2010). This study was later confirmed by one that used a larger number of SNPs (Xu et al., 2012). In Madagascar, the Malagasy uniformly have around 70% African ancestry and 30% Asian ancestry (Pierron et al., 2014) (Figure 1.13).

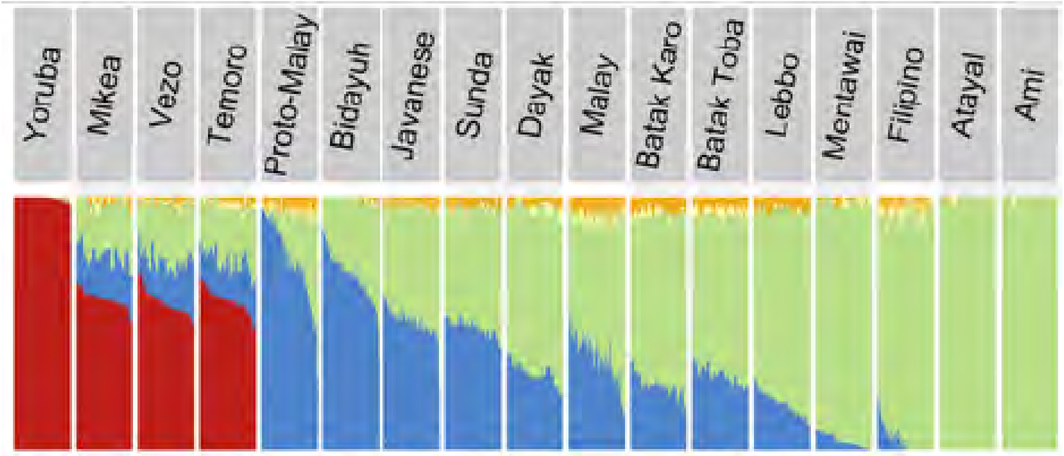


Figure 1.13. A diagram of admixture in Madagascar populations (Mike, Vezo, Temoro) showing a mixture between African (red) and Asian ancestry (blue and green), taken from Pierron et al. (2014).

Many methods have been developed to analyse admixed populations, to estimate the proportion of ancestry derived from each ancestral group, and to infer the time of the admixture event(s). Methods are mostly divided into two types of analyses: descriptive estimation and inference estimation methods (global and local ancestry). Descriptive analysis involves estimating the proportion of ancestry contributed by different populations averaged across the entire genome. Clustering methods such as Principal Component Analysis (PCA)

(Patterson et al., 2006), STRUCTURE (Pritchard et al., 2000) and ADMIXTURE (Alexander et al., 2009) have become indispensable tools for investigating population structure in genome-wide data. STRUCTURE and ADMIXTURE have been applied to determine regional levels of clustering and ancestries. The power of PCA analysis can recreate a geographical map of a continent by studying genetic variation within that region (Novembre et al., 2008). These methods are useful to determine population substructure, yet they do not provide any formal tests, such that the Isolation-by-Distance effect can generate similar results to long-distance historical migrations.

There are two types of inference methods: global and local ancestry inference analyses. Global ancestry usually involves demographic models and phylogenetic. This method infers change in population size, time of admixture and split, and other parameters that can be included in the analysis. This method is run by the Approximate Bayesian Clustering (ABC) computation (Beaumont et al., 2002; Csilléry et al., 2012). Local ancestry, such as LAMP (Sankararaman et al., 2008), HAPMIX (Price et al., 2009), PCAdmix (Brisbin et al., 2012), fineSTRUCTURE (Lawson et al., 2012) and GLOBETROTTER (Hellenthal et al., 2014), determine ancestry at each locus in the genome, and provide individual-level information about ancestry, admixture mapping and time of admixture. This method relies on accurate population data given by the exact mixing populations and its sources. The absence of accurate mixing populations can disrupt the precision of the results. This method, however, can provide very useful information on recent admixture events (~5-150 generations ago), but it has a reduced capacity to detect older events (Hellenthal et al., 2014). Hence, the combination of both descriptive and inference methods, an approach used in this study, gives a broader insight into the admixture history of a population.

1.3.2.4. Disease related markers

During their vast migrations, human groups have encountered a wide spectrum of micro-organisms and viruses that have affected their biology. In parallel to the epidemiological interest of these interactions, the study of hijacked genomes and their imprint on human biology is a powerful tool to further explore human history (Cassar et al., 2012, 2013; Moodley et al., 2009; Thedja et al., 2011) (Figure 1.14). The geographic distribution of pathogen diversity and host resistance genes are powerful population markers, as humans and pathogens would have co-evolved along the migratory route.

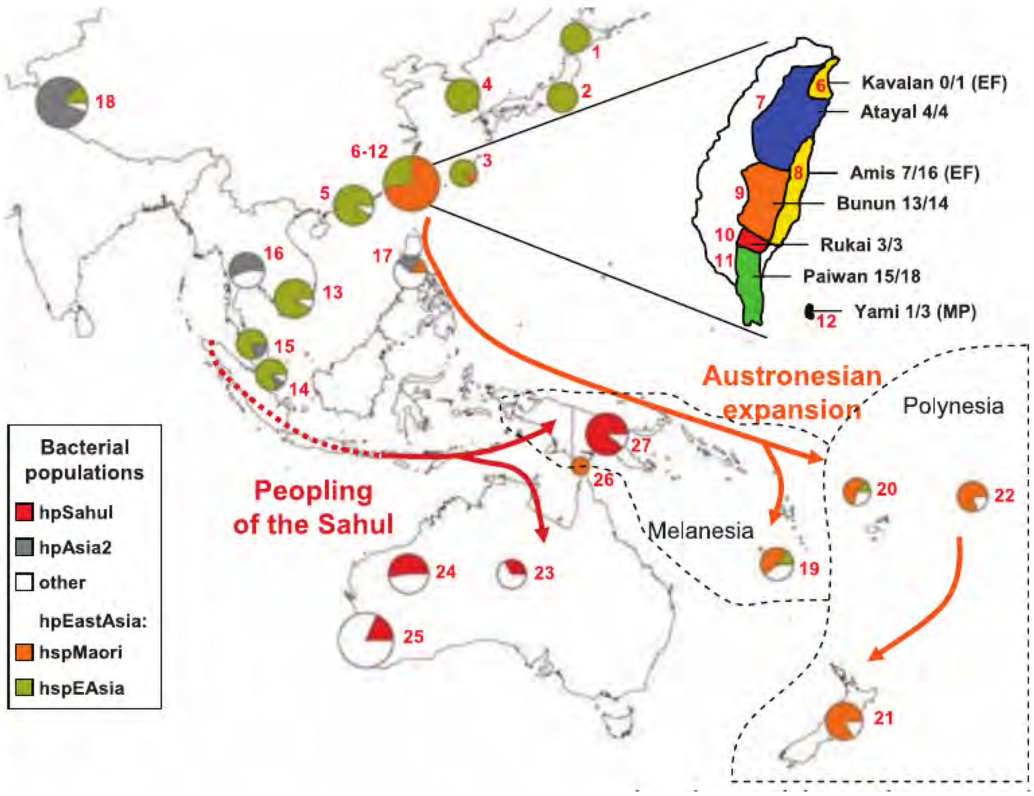


Figure 1.14. The distribution of *Helicobacter pylorii* genotype mimics the Austronesian-speaking dispersal route from South China and Taiwan into the Pacific, taken from Moodley et al. (2009).

One example is the geographic distribution of the Hepatitis B virus overlapping the dispersal route of human mtDNA in Indonesia. Hepatitis B virus (HBV) infection is a major cause of hepatitis. HBV genetic variants have important roles in the development of chronic hepatitis B, and so different prognoses and treatments are made according to the HBV genetic profile present (Kao et al., 2004; Zhang et al., 2016). Until now, researchers have identified eight HBV genotypes: genotypes A to H. Genotype B and C are common in the region of Southeast Asia (Lusida et al., 2008). A total of eight sub-genotypes HBV/B are present in different regions: B2 in China, B3 in Indonesia, B4 in Vietnam, B5 in the Philippines, and B7 and B8 in Nusa Tenggara. Similarly, the sub-genotype HBV/C is present in different regions: C1, C5 and C6 in Southeast Asia, C2 in East Asia, C3 in the Pacific islands, and C4 in Aboriginal Australian populations. Recent study has shown that the distribution pattern of the HBV genotype in Indonesia has a positive correlation the virus sub-genotype and the ethnic background of the population. For example, HBV sub-genotype B3 is common in ethnic populations in the western part of Indonesia, while HBV sub-genotype B7 is mainly found in ethnic populations in eastern Indonesia (Thedja et al., 2011).

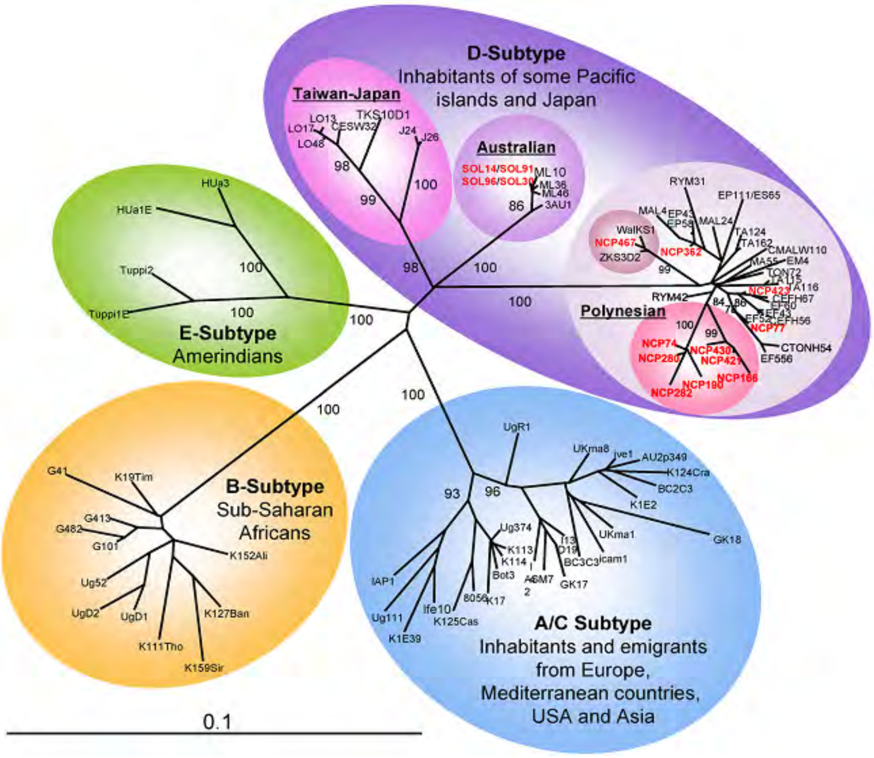


Figure 1.15. Phylogenetic tree of HHV8 subtypes clustered according to their regional geographical distribution, taken from Cassar et al. (2012).

Another example comes from recent studies on the distribution patterns of the Human T-Lymphotropic Virus type 1 (HTLV1) and the Human Herpes Virus 8 (HHV8); both have been used successfully as markers of human migration into Australo-Melanesian regions (Cassar et al., 2012, 2013, 2017), but they have not yet been used for investigating migrations within ISEA. Both viruses have the ability to integrate their DNA into the human genome, and they have unique characteristics which make them relevant population markers for infected groups: low mutation rate, relative molecular stability and mother-to child mode of transmission. The HTLV1 genotypes are determined from the non-coding Long Terminal Repeat (LTR) region, and the phylogeographical tree is highly discriminative (Cassar et al., 2013). There are seven main subtypes of HTLV based on the LTR genomic proviral region: subtype A has a global distribution, subtypes B, D, E, F and G are found in Africa, and the Australo/Melanesian subtype C is found only in Oceania. On the other hand, molecular epidemiology studies on HHV-8 have focused mainly on the variable K1 region (ORF-K1). This has led to the identification of five main viral subtypes (A, B, C, D, E) that exhibit geographical clustering. Subgroup A1–4 and subtype C are predominant among populations of European descendant in some regions of Asia. Subgroup B1–4 and clade A5 are predominant in Sub-Saharan Africa (Betsem et al., 2014). Subgroups E are found in Native American population, while subgroups D are found in Southeast Asia, Melanesia - Oceania region and Taiwan. Melanesia and Oceania regions have the highest diversity of HHV-8 (Cassar et al., 2007) (Figure 1.15).

In addition to viral diseases, Indonesia, as a tropical country, has many endemic parasitic diseases, including malaria which is caused by the Plasmodium parasite. Parasites and humans have a long history of interaction, and this has been implicit in the evolution of genetically determined disease resistance in humans. In malaria-endemic environments, different biological mechanisms have been selected for in human groups, for example, genetic diseases such as sickle-cell disease in Africa and the Southeast Asian Ovalocytosis (SAO) in Asia. SAO is characterised by the deformation of a patient's erythrocyte membranes into an oval shape (O'Donnell et al., 1998). SAO has autosomal dominant inheritance, and all SAO carriers are heterozygous, because the dominant form of the SAO allele is lethal. In some cases, SAO is not characterised as a disease as it is asymptomatic, causing no health problems to the carrier (Allen et al., 1999; Liu et al., 1994). The cause of SAO is a 27-bp deletion in the Solute Carrier Family 4 (AE1) member 1 (SLC4A1) gene (Mohandas et al., 1984). This causes a loss of nine amino acids in the protein erythrocyte band 3, which prevents Plasmodium parasites binding with erythrocytes and causing malaria.

SAO has a worldwide distribution in Malaysia, Melanesia, Papua New Guinea, the Philippines, Indonesia and Madagascar (Paquette et al., 2015), a distribution which is hypothesised to be related to the Austronesian dispersal towards Indonesia, the Pacific and Madagascar (Rabe et al., 2002; Wilder et al., 2009) (Rabe et al., 2002; Wilder et al., 2009). A study conducted on Papua New Guinean populations concluded that the distribution pattern of SAO is positively associated to populations with distinctive Austronesian markers (Tsukahara et al., 2006). The age of SAO has been estimated, by sequencing the region in the SLC4A1 gene, to around 10,000 years, predating the spread of agricultural practices and the dispersal of the Austronesian language (Paquette et al., 2015). This evidence indicates that SAO was brought by Austronesian people to Papua New Guinea about 3,500 years ago, in agreement with the origin of the Lapita cultural complex in the Bismarck Archipelago (Papua New Guinea). Using diseases as migratory markers can enhance our understanding of global human migration through history, especially in the Indo-Pacific region (Madagascar and the Indonesian archipelago), where studies on viruses and host resistance are scarce.

II. Samples and technical works

2.1. Collaborative framework

My thesis is a collaborative study between the Laboratory of Molecular Anthropology and Synthesis Imaging (AMIS) UMR-5288, University of Toulouse (France), and the Eijkman Institute for Molecular Biology (Jakarta, Indonesia). My host laboratory, AMIS, undertakes fundamental research focusing on humans and human diversity, from the first hominids to modern human groups inhabiting different environments across the world. The Eijkman Institute for Molecular Biology is a research institute that undertakes a wide range of studies related to the molecular analysis of the Indonesian populations, from forensic analyses to population genetic histories and disease-related molecular studies. Understanding human genetic structure has fundamental implications for understanding the evolution and impact of human diseases. The study on the genetic structure of Indonesian populations by researchers from the Eijkman Institute has focused on its connection to disease-related molecular epidemiology. This collaboration enriches the exploration of Indonesian population history and global migrations through multidisciplinary studies.

2.2. Study populations

2.2.1. Data collection and data privacy

The historical and medical advantages of studying Indonesian population structure is the approach that we used to explain the significance of our study to the people in the villages where we collected samples. Sample collection followed protocols for the protection of human subjects established by both the Eijkman Institute and by the French scientific institution. The study was approved by the Eijkman Research Ethic Commission under Research Ethic clearance number 90 for the study of Indonesia Human Genome Diversity and Diseases, and by the *Cellule Bioéthique* of the *Secretariat d'état à l'enseignement supérieur et à la recherche* (n°IE-2015-837(1)). With the assistance of the Indonesian Public Health clinic and Eijkman Institute staff, we also provided medical care and medicines to the participants in each sampling location. The study overview was presented to all the participants, they then filled out a questionnaire that asked for information on their ethnicity, languages spoken and geographic location, for at least the last two to three generations. Informed consent was signed by each participant, who could withdraw from the study at any time without giving reason. A statement guaranteed the confidentiality of the research participants, indicating that the participants had understood all the information in the consent form and were willing to volunteer/participate in the research. We collected biological samples in the form of blood and saliva samples for genetic analyses (using Oragene (DNA Genotek, Inc.) and/or Norgen (Norgen Biotek Corp.) sampling kit). I collected around ~500 samples during four different fieldwork campaigns, which I organised and conducted. The documentation of sampling activities are displayed in Annex A. I performed the DNA extractions from both blood and saliva samples in the Eijkman Institute laboratory, with the assistance of the Eijkman Institute staff. The detailed methods of the DNA extraction is described in the appended manuscripts.

Table 2.1. List of the populations sampled

No.	Ethnic group	Language	Region	Latitude	Longitude	n
1	Ma’anyan	Ma’anyan	East Barito, Central Kalimantan	-2.108057	115.163496	167
2	SK Dayak	Various southern Dayak languages	Banjarmasin, South Kalimantan	-3.319621	114.589684	38
3	Ngaju	Ngaju	Banjarmasin, South Kalimantan	-3.319621	114.589684	28
4	Banjar	Banjar/Malay	Banjarmasin, South Kalimantan	-3.319621	114.589684	100
5	Samihim	Samihim (a subgroup of Ma’anyan)	Kotabaru, South Kalimantan	-2.485020	115.944103	71
6	Bajo Kotabaru	Bajo	Kotabaru, South Kalimantan	-3.224478	116.267831	60
7	Bajo Derawan	Bajo	Berau, East Kalimantan	2.284365	118.243355	42

2.2.2. Fieldwork

Sampling campaigns were performed in the Indonesian part of Borneo Island (Kalimantan) during different phases of my doctoral study. The first sampling campaign was undertaken in 2013, a few months prior to the starting the doctoral program. The subsequent sampling campaigns were performed in late 2014 and twice in 2015. Each campaign required 2 weeks of paperwork (clearance and permits from local authorities), and on average 2 weeks for sampling individuals in each ethnic group with a team of 5 to 6 people (scientists, technicians and medics). Sampling activities were conducted in different areas of Borneo Island to include 6 distinct ethnicities and 1 Dayak group of various ethnicities. All samples collected in the fieldwork campaigns are summarised in Table 2.1.

Table 2.2. List of archived samples of populations involved in this study.

No.	Ethnic group	Language	Region	Latitude	Longitude	n
8	Lebbo’	Lebbo’	Malinau, North Kalimantan	1.785167	114.829353	19
9	Bajo Kendari	Bajo	Kendari, Southeast Sulawesi	-4.001891	122.509459	27
10	Bugis	Bugis	Luwu, South Sulawesi	-3.311117	120.347988	50
11	Mandar	Mandar	Majene, West Sulawesi	-3.264819	118.880867	50
12	North Maluku	Various local Austronesian and Papuan languages	Ternate, North Maluku	0.786298	127.374594	38

Figure 2.1. The geographical distribution of the new and archived samples included in this study. Numbers on the map correspond to the population ID number in Table 2.1 and Table 2.2. The Indonesian map is taken from http://freevectormaps.com

2.2.3. Archived samples

I incorporated samples in my data analyses from previous independent sampling campaigns in Indonesia conducted by Dr. François-Xavier Ricaut (AMIS laboratory, University of Toulouse) through the MAFBO (*Mission Archéologique Francaise à Borneo*) research project, and also through a project led by the Human Genome Diversity and Diseases

Laboratory of Eijkman Institute for Molecular Biology. All archived samples are stored in each institution, and those included in my study are listed in Table 2.2. The geographical locations of all new and archived samples are depicted in Figure 2.1.

2.2.4. Comparative dataset

I included SNP data from public databases, such as the 1000 Genomes (The 1000 Genomes Project Consortium, 2015) (http://www.internationalgenome.org), HGDP (Li et al., 2008) (http://www.cephb.fr/en/hgdp_panel.php), HUGO Pan Asian SNP (HUGO Pan-Asian SNP Consortium et al., 2009) (http://www4a.biotec.or.th/PASNP/front-page), and other published data from several publications. For mtDNA and Y-chromosome haplogroup comparative datasets, I incorporated published data on studies of Indonesian populations (Karafet et al., 2010; Gunnarsdottir et al., 2011; Tumonggor et al., 2013) and Malagasy populations (Capredon, 2011; Razafindrazaka, 2010; Tofanelli et al., 2009). Each article published in this doctoral study used a different comparative population dataset, and all the data can be found in each article's appended annexes and supporting materials (section VI).

2.3. Laboratory and technical works

After the sampling campaigns, I performed DNA extractions from saliva samples using the manufacturer's standard protocols. All DNA extracts were quantified using Nanodrop 2000 (Thermo Scientific). For Y-chromosome haplogroup analysis, I sent the DNA extract to a designated service company (Integragen) in Paris to conduct SNP survies on the Y-chromosome. Haplogroups were then estimated from all male samples from surveyed SNPs following international nomenclature (van Oven et al., 2014) (http://www.phylotree.org/Y/). For mtDNA haplogroup analysis, I worked from the hyper-variable region (HVR) in laboratories in Jakarta and Toulouse. I performed the standard PCR protocol, the mtDNA hypervariable region I was sequenced using primers F15989 (5'CCCAAAGCTAAGATTCT AAT-3') and R389 (5'-CTGGTTAGGCTGGTGTTAGG-3'), the PCR product was purified, Cycle Sequencing PCR using Big Dye Terminator Kit (ThermoFisher Scientific) was undertaken, along with purification using the Ethanol precipitation and EXOSAP procedure.

The purified cycle-PCR products were then sent to sequencing services in both institutions (Genopole Toulouse Midi-Pyrénées, France, and Sequencing Laboratory of the Eijkman Institute for Molecular Biology, Jakarta), to conduct the Sanger sequencing. Mitochondrial haplogroups from all samples were estimated using polymorphisms of HVR sequences of all samples according to Phylotree v.16 (van Oven and Kayser, 2009).

To complement the above work, I performed whole mitochondrial genome sequencing using Next-Generation Sequencing, to confirm haplogroup assignment from HVR sequencing, focusing on the Indonesian maternal lineages connected with Madagascar (the B4a* haplogroups). Due to the late time-frame of data generation, analyses of these mitogenomes are not included in this thesis, but they will be investigated further in the near future. I used the capture-based and direct long-range PCR method. For the capture method, I prepared libraries from whole DNA samples using Kit NEBNext E7370S (New England Biolabs) following the method of Meyer and Kircher (2010). I then conducted the capture protocol from the Centre of GeoGenetics in Copenhagen in collaboration with Dr. Hannes Schroeder, following the method of Maricic et al. (2010). For the direct long-range PCR method, I prepared libraries from the long-range mtDNA PCR products using a similar library preparation method as above. The libraries were sent to a sequencing service institution in Toulouse (Genopole Toulouse Midi-Pyrénées) to generate the sequencing reads. I built a bioinformatic pipeline to generate sequences from the raw sequencing reads (fastq format). The haplogroups were estimated using Haplogrep (http://haplogrep.uibk.ac.at) according to Phylotree v.16

To generate data using SNP arrays, I used different types of SNP chips. Firstly, I used Illumina Omni Express (700K SNPs surveyed) conducted by Gene by Gene (Houston, Texas, USA), then Illumina Omni 2.5 (2M SNPs surveyed) in collaboration with Dr. Denis Pierron, conducted in the *Centre National de Génotypage* (CNG - Paris), and also Illumina Omni 5 (4.5M SNPs surveyed) conducted in CNG. The data were retrieved in PLINK ped/bed format, then followed by quality control checks for SNP calling quality in each locus and individual. Detailed parameters were specified in each published article in section III. Samples to which each SNP array were applied are mentioned in the respective annex of each published article (Section III and VI). All different analyses performed on each population (mtDNA, Y-chromosome, Illumina Omni Express SNP array, Illumina Omni 2.5 SNP array, Illumina Omni 5 SNP array) are listed in Annex B.

III. Indonesian genetic diversity and their genetic impact in Madagascar

3.1. Uniparental marker analyses

In this sub-chapter, I present my research on the two main aims of this thesis – the origin of Indonesian genetic diversity and its relationship to the genetic background of the Malagasy population - by using uniparental markers (mtDNA and Y-chromosome). I published two articles on these topics, using data from newly sampled Indonesian populations and previously published data. The first article was published in the journal **BMC Genomics**, entitled "*Mitochondrial DNA and the Y-chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations*", and the second was published in **Quaternary International** journal, entitled "*West Eurasian genetic influences in the Indonesian archipelago*". The first article focuses on question 1, 3 and 4 (see p13), and explains the position of the Indonesian populations within the current genetic diversity of Indonesia and ISEA, making a clear distinction between western and eastern Indonesia. Furthermore, I also observed some links between Indonesian and Malagasy populations from uniparental markers, suggesting a complex scenario in the role that Indonesian populations have played in the settlement of Madagascar. Based on the same dataset, the second article focused on question 2 (see p13). The study revealed western Eurasian (South West and/or South Asia) genetic traces in various Indonesian populations related to the last 2,000 years of historical interactions, and a different geographic origin in western Eurasia and geographic distribution in Indonesia of the western Eurasian maternal and paternal lineages.

3.1.1. Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations

(Published in BMC Genomics, 16:191, March 17th 2015. doi: 10.1186/s12864-015-1394-7)

Pradiptajati Kusuma[1,4], Murray P. Cox[2], Denis Pierron[1], Harilanto Razafindrazaka[1], Nicolas Brucato[3], Laure Tonasso[1], Helena Loa Suryadi[4], Thierry Letellier[1], Herawati Sudoyo[4,5], François-Xavier Ricaut[1§]

[1] Laboratoire d'Anthropologie Moléculaire et Imagérie de Synthèse UMR-5288, Université de Toulouse, Toulouse, France
[2] Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
[3] Center for Linguistics, University of Leiden, Leiden, Netherlands
[4] Eijkman Institute for Molecular Biology, Jakarta, Indonesia
[5] Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia

3.1.1.1. Abstract

Background

Linguistic, cultural and genetic characteristics of the Malagasy suggest that both Africans and Island Southeast Asians were involved in the colonization of Madagascar. Populations from the Indonesian archipelago played an especially important role because linguistic evidence suggests that the Malagasy language branches from the Southeast Barito language family of southern Borneo, Indonesia, with the closest language spoken today by the Ma'anyan. To test for a genetic link between Malagasy and these linguistically related Indonesian populations, we studied the Ma'anyan and other Indonesian ethnic groups (including the sea nomad Bajo) that, from their historical and linguistic contexts, may be modern descendants of the populations that helped enact the settlement of Madagascar.

Result

A combination of phylogeographic analysis of genetic distances, haplotype comparisons and inference of parental populations by linear optimization, using both maternal and paternal DNA lineages, suggests that Malagasy derive from multiple regional sources in Indonesia, with a focus on eastern Borneo, southern Sulawesi and the Lesser Sunda islands.

Conclusion

Settlement may have been mediated by ancient sea nomad movements because the linguistically closest population, Ma'anyan, has only subtle genetic connections to Malagasy, whereas genetic links with other sea nomads are more strongly supported. Our data hint at a more complex scenario for the Indonesian settlement of Madagascar than has previously been recognized.

Keywords: Ma'anyan, Indonesia, Madagascar, sea nomad, mitochondrial DNA, Y chromosome

3.1.1.2. Introduction

Prior to the European colonial period, Austronesian-speaking populations were the most widespread of any language family (Bellwood et al., 1995; Blust, 2011). While most groups speaking Austronesian languages moved eastward, settling the Pacific Ocean, others moved westward through the Indian Ocean, reaching eastern Africa and Madagascar. Dispersing halfway around the world within the past two millennia, the Austronesian expansion is often considered the last substantial wave of migration in human prehistory (Blench, 2006; Burney et al., 2004; Kumar, 2011).

Despite considerable research on the eastward Austronesian expansion, there is little equivalent research on the western edge, leaving major issues unresolved regarding the settlement of Madagascar. Although the exact nature and route of this movement is largely unknown, linguistic and anthropological evidence indicates strong Indonesian influences, as recorded in the vocabulary and socio-cultural life of Malagasy, the modern people of Madagascar (Adelaar, 2005; Blench, 2010; Dewar and Wright, 1993; Serva et al., 2012). Linguistic research suggests that the Malagasy language is derived from Southeast Barito (SEB), a subgroup of Austronesian languages, and is most closely related to the language spoken by the land-locked forest-dweller Ma'anyan in central and southeastern Kalimantan (Borneo) (Adelaar, 2005, 1995a, Dahl, 1991, 1977, 1951), one indigenous language among 73 others spoken in Borneo (Lewis et al., 2016). However, there is evidence of word borrowings from a small number of Austronesian languages spoken on other Indonesian islands as well (Adelaar, 2009, 2006). This probably reflects multiple Austronesian arrivals to Madagascar from about 700 AD onward (although earlier dates cannot be completely excluded). One hypothesis is that earlier movements were perhaps linked to Southeast Barito speakers, with later arrivals during the 12th-15th centuries connected instead to the Srivijaya and Majapahit kingdoms of Southeast Asia (Beaujard, 2012a, 2003).

These linguistic findings are broadly supported by genetic studies, which emphasize the shared Indonesian and African genetic heritage of Malagasy. A recent study of genome-wide SNP data suggests that the western and central regions of Indonesia (Java/Borneo/Sulawesi) have the closest genetic connections with Malagasy (Pierron et al., 2014). This is in agreement with previous studies of uniparental markers (mtDNA and the Y chromosome), which found genetic affinity between Malagasy and western Indonesian populations (Hurles et al., 2005; Tofanelli et al., 2009). A key lineage linking Indonesia and Madagascar is the

Polynesian motif (a mitochondrial DNA haplogroup, B4a1a1, characterized by the polymorphisms A14022G, T16217C, A16247G and C16261T) (Soodyall et al., 1995). More recently, it has been recognized that Malagasy carry specific point mutation variants (mtDNA nucleotides 1473 and 3423), which together have been termed the Malagasy motif (Razafindrazaka et al., 2010). This Malagasy version of the Polynesian motif is distributed throughout Madagascar with frequencies in specific ethnic groups ranging from 11-50% (Capredon et al., 2013; Razafindrazaka et al., 2010; Tofanelli et al., 2009). While still debated, this relatively homogenous distribution has been interpreted as supporting the first arrival of the Polynesian motif during an early phase of Madagascar's settlement (Cox et al., 2012). To date, the Malagasy motif has not been found in Indonesia, or anywhere else outside Madagascar. However, this may simply reflect the paucity of Indonesian populations available for study.

The westward Austronesian expansion was likely associated with trading activities of the Srivijaya empire, as suggested by the many Malay loanwords present in Malagasy (Adelaar, 2009, 2005, 1989), and this trade has been hypothesized to involve some sea nomad groups (i.e., the Orang Laut, Bajo and Bugis) (Beaujard, 2011, 2012a; Pelras, 1997). This trading network was dominated by men, thus hinting at a potential male bias in the Indonesian contribution to Malagasy, in concordance with the standard matrilineal/matrilocal bias of traditional Austronesian society (Jordan et al., 2009; Lansing et al., 2011; Marck, 2008; Peng et al., 2010). Contact between the Srivijaya empire and southeast Borneo may have stimulated the dispersal of Southeast Barito speakers to Madagascar – possibly at the same time as the dispersal of Sama-Bajaw speakers from the same area, but a different subgroup of Barito languages (Blust, 2007). In this context, the Bajo are one sea nomad population of particular interest. Today, the Bajo live in several coastal communities around East Borneo, Sulawesi, the Lesser Sunda islands and the Maluku islands (Nuraini, 2008; Stacey, 2007). Because the Austronesian migration to Madagascar and the Sama-Bajaw dispersal may be interrelated, we compare genetic data from recent seafaring populations, such as the Bajo, with the more settled Malagasy.

Similarly, whether Barito populations such as the Ma'anyan, the closest linguistic siblings to modern Malagasy, share close genetic lineages with the Malagasy also remains unanswered. For the first time, we report genetic data for the Ma'anyan and the Lebbo' (a population from Borneo with no presumed role in the settlement of Madagascar) to determine whether the

Ma'anyan have an especially close genetic connection with Malagasy. We also include Bajo sea nomads from Sulawesi to determine whether there is a common genetic link based on their shared involvement in long distance maritime trading networks. A large data set of published and unpublished Indonesian populations is included for comparative analysis (Karafet et al., 2010; Tumonggor et al., 2013) [38, 39]. To investigate sex-specific genetic connections between Indonesia and Madagascar, we analyze both maternal (mtDNA) and paternal (Y chromosome) variation. We propose that the genetic connections of Malagasy to Indonesia are not restricted to Borneo, but instead include maternal and paternal lineages from a wide range of source populations from southern Sulawesi and the Lesser Sunda islands. We therefore propose that the settlement of Madagascar may have been mediated, at least in part, by sea nomad groups.

3.1.1.3. Methods

Population Samples

All samples analyzed in this study were collected with informed consent from unrelated individuals. Subjects were surveyed for language affiliation, current residence, familial birthplaces, and a short genealogy of four generations to establish regional ancestry. A total of 205 DNA samples were analyzed from three ethnic groups: 159 Ma'anyan individuals were collected in Tamiang Layang (East Barito), Central Kalimantan, and Banjarmasin (South Kalimantan), representing the largest ethnically-defined population sample from Borneo to date; 19 Lebbo' in East Kalimantan; and 27 sea nomad Bajo in Kendari (Sulawesi). Collection and use of these samples was approved by the Research Ethics Commissions at both the Eijkman Institute for Molecular Biology, Indonesia, and the University of Toulouse, France. We also included data for additional Indonesian populations from published (Gunnarsdóttir et al., 2011; Karafet et al., 2010; Tumonggor et al., 2013) and unpublished sources (Gayo, North Maluku, and a mixed assemblage of other Dayak ethnic groups from the southern part of South Kalimantan province ("SK Dayak") from the Eijkman Institute's archived samples) (Annex C: Table S1). We also included published Malagasy data: seven Malagasy populations located in the southwest, southeast, and central highlands of Madagascar. The Malagasy were pooled for most analyses as these groups are genetically highly similar (between group $F_{ST} < 0.05$ and 95% of F_{ST} values non-significant ($P > 0.05$) for both mtDNA and the Y chromosome) (Capredon et al., 2013; Razafindrazaka, 2010; Tofanelli et al., 2009). In total, the mtDNA dataset comprises 529 Malagasy and 2,841 Indonesians, and the Y chromosome dataset comprises 371 Malagasy and 2,095 Indonesians.

DNA Extraction and Genotyping

We collected blood samples for the Ma'anyan, except the Lebbo' and Bajo, for which saliva samples were collected using the Oragene DNA Collection kit (http://dnagenotek.com). DNA was extracted from blood using a standard salting-out procedure, and from saliva using the manufacturer's standard protocol. For paternal lineage analysis, 96 binary markers on the non-recombining region of the Y chromosome were analyzed. We used a nanofluidic dynamic array (Fluidigm, USA) high-throughput genotyping system. This system is developed for SNP genotyping assays and able to perform 9,216 real-time polymerase chain reactions (PCRs) (96 primers × 96 samples) on a single chip. The results were analyzed using the BioMark™ HD system (Fluidigm, USA) which integrated the Real-Time PCR Analysis software. Each haplogroup was assigned based on the updated ISOGG's Y-DNA haplogroup tree [64] and the Y-Phylotree [65]. The full list of markers is shown in Annex C: Table S2. The mtDNA hypervariable region I was sequenced using primers F15989 (5'-CCCAAAGCTAAGATTCT AAT-3') and R389 (5'-CTGGTTAGGCTGGTGTTAGG-3'). Sequences (GenBank accession numbers: KM590988-KM591192) were then analyzed and aligned against the revised Cambridge Reference Sequence (rCRS) (Andrews et al., 1999) using the MAFFT aligner v.7 (Katoh and Standley, 2013). Mitochondrial haplogroups were determined with the Haplogrep program (http://haplogrep.uibk.ac.at) based on Phylotree v.16 (van Oven and Kayser, 2009). The Malagasy motif, defined by mitochondrial coding region polymorphisms at nucleotides 1,473 and 3,423, were typed on all individuals carrying haplogroup B4a* using the method previously described (Razafindrazaka et al., 2010).

Statistical Analysis

Pairwise F_{ST} distances between Indonesian and Malagasy populations were computed from haplogroup frequency data using Arlequin v.3.5 (Excoffier and Lischer, 2010) with 5,040 permutations. Multidimensional scaling (MDS) from F_{ST} values based on Y chromosome and mitochondrial DNA haplogroup frequencies (Annex C: Table S3 and Table S4) was performed to visualize inter-population relations. The nonparametric Mann-Whitney U-test was applied to analyze the statistical significance of genetic affinity between Malagasy and Western/Eastern Indonesian groups. This phylogeographical division was defined by Wallace's line, in agreement with previous human genetic population studies (Cox et al., 2010; Tumonggor et al., 2013; Xu et al., 2012). F_{ST} values obtained for the pairwise comparison of maternal and paternal lineages between Malagasy and Indonesian populations were plotted geographically with Surfer v.12.0 using the Kriging method. To determine which linear combination of Indonesian populations produces the closest genetic profile to that observed in Malagasy, we employed a statistical analysis of least squares with equalities and inequalities (lsei) algorithm in the R package, limSolve (Van den Meersche et al., 2008). To capture sampling variance and drift dynamics, the genetic data were resampled 5,000 times and the linear optimization results visualized with box plots using the R package, ggplot. This analysis used mitochondrial DNA and Y chromosome haplogroup frequency distributions for both Malagasy and Indonesian populations (Annex C: Table S3 and Table S4). An African reference group was used to represent the non-Asian contribution to Malagasy. This reference comprised African samples from populations in North-, East-, Central- and South Africa (Cruciani et al., 2004; Harich et al., 2010; Rosa et al., 2006; Salas et al., 2004, 2002; Tishkoff et al., 2007; Wood et al., 2005). Sharing of mitochondrial haplotypes was ascertained using Arlequin v.3.5. For this analysis, sequences from Tofanelli et al. (2009) were excluded due to their short length (360 bp compared to 520 bp for the present study).

Table 1. Y chromosome haplogroup frequencies in the Ma'anyan, Lebbo' and Bajo.

Haplogroups		Lebbo'	Ma'anyan	Bajo	Geographic origin
C*	C-RPS4Y*	0.1333	0.5222	0.0370	SEA
K*	K-M9*	0.0667	--	--	SEA
KxLT	K-M526*	0.1333	0.0222	0.2222	SEA
O1a	O-M119*	--	0.0222	0.0370	SEA
O1a1	O-P203	--	0.1000	--	SEA
O2a1	O-M95*	0.3333	0.0778	--	SEA
O3	O-M122*	--	0.0111	0.0370	SEA
C1c	C-M38*	--	--	0.2222	SEA
O1a2	O-M110	0.2667	0.0222	--	SEA
O2a1a	O-M88	--	0.0111	--	SEA
O3a2	O-P201*	0.0667	0.1778	0.0741	SEA
O3a2b	O-M7	--	--	0.0370	SEA
M1a	M-186	--	--	0.0741	SEA
P*	P-M45*	--	0.0222	--	SEA
R*	M-207*	--	--	0.1481	WE
R1a	R-M17	--	0.0111	--	WE
T1a	T-M70	--	--	0.0741	WE
L1a	L-M76	--	--	0.0370	WE

Note.

SEA: Southeast Asian origin; WE: Western Eurasian origin

3.1.1.4. Results

Y chromosome and mitochondrial DNA classification

Based on analysis of 96 Y chromosome binary markers (Annex C: Table S2), the majority of men in the Ma'anyan, Lebbo' and Bajo carry haplogroups previously found in Southeast Asia, particularly C*, K*, and O* (Table 1). Only a few individuals carry Y chromosomes belonging to Western Eurasian haplogroups: R* (M207) (Hammer and Zegura, 2002; Karafet et al., 2008) was found in four Bajo individuals, R1a (M17) (Pamjav et al., 2012; Underhill et al., 2010) was found in one Ma'anyan individual, while the western Eurasian haplogroups L1a (M76) and T1a (M70) (Mendez et al., 2011) were found in one and two Bajo individuals, respectively. Indian haplogroup R*, which includes R1a, has previously been identified in Bali, Java, Borneo, and Mandar (Annex C: Table S3), and thus could conceivably have transited through Indonesia (as opposed to a direct connection), but T1a and L1a have not been identified to date in any Indonesian population.

On the mitochondrial DNA (Table 2), the frequency distributions of haplogroups found in the Ma'anyan, Lebbo' and Bajo are broadly similar, and consistent with, patterns of maternal lineages in Indonesia. Indeed, four main geographical/historical affiliations can be observed: mainland Asia, the Austronesian expansion, western Eurasia/India, and New Guinea. In brief, mainland Asian mtDNA haplogroups (such as B4c2, M73, M74, M12) are carried by a majority of individuals (64%), followed by haplogroups that have been putatively linked with an Austronesian expansion out of Taiwan (such as B4a1a1, M7c1a4, F1*, E*; 32%). The remaining lineages likely derive from India and west Eurasia, and were only observed among the Ma'anyan (M2, M5a4, and M35a, ranging in frequency from 0.6-1.9%). The presence of Indian and other western Eurasian genetic traces has been observed previously in Borneo, as well as Sumatra, Java and Bali (Gunnarsdóttir et al., 2011; Karafet et al., 2005) (Annex C: Table S4). Indian haplogroups are restricted to western Indonesia, particularly in regions historically involved in the ancient trading networks of the Hindu kingdoms (such as Srivijaya and Majapahit). Among the Bajo, we also observed the M1a Y haplogroup and Q1 mitochondrial haplogroup, which likely traces its ancestry to New Guinea or eastern Indonesia (Friedlaender et al., 2007; Kayser et al., 2003; Mona et al., 2009; Tumonggor et al., 2013). These haplogroups represent a trace of Papuan genetic input. This is perhaps due to the extensive trading network of the Bajo eastward to New Guinea (Stacey, 2007) and/or earlier westward expansions of Papuan speakers from New Guinea to eastern Indonesia (Donohue and Denham, 2010).

Table 2. Mitochondrial haplogroup frequencies in the Ma’anyan, Lebbo’ and Bajo.

Haplogroups	Lebbo'	Ma'anyan	Bajo	Geographic origin
B4a	0.2105	0.0943	0.0741	MA
B4c1b	--	0.0252	0.0741	MA
B4c2	--	0.1887	--	MA
B5a	0.1579	0.0377	0.0370	MA
B4a2a	--	0.0440	--	MA
B4a4	--	--	0.0741	MA
B4b1	--	0.0818	--	MA
F3b1a	--	0.0189	--	MA
M12	--	0.0377	--	MA
M20	0.1579	0.0440	--	MA
M71a2	0.1579	--	--	MA
M73	--	0.0566	0.0370	MA
M74b1	--	0.1069	--	MA
N22	--	0.0189	--	MA
N9a6a	--	0.0252	--	MA
R22	--	--	0.0370	MA
R9b1a1a	0.1053	--	--	MA
X	--	--	0.0370	MA
Q1	--	--	0.0741	NG
B4a1a1	--	--	0.0370	Taiw
D4s	--	0.0629	--	Taiw
E1a	0.2105	--	0.0741	Taiw
F1a	--	0.0377	0.0370	Taiw
F1a1a	--	0.0377	--	Taiw
F1a3	--	--	0.0741	Taiw
F1a4	--	0.0252	--	Taiw
M7b1a1i	--	--	0.1481	Taiw
M7b1a2	--	0.0126	0.0741	Taiw
M7c1a4a	--	0.0126	0.1111	Taiw
M2	--	0.0063	--	WE
M35a	--	0.0189	--	WE
M5a4	--	0.0063	--	WE

Note.
MA: Mainland Asian origin; NG: New Guinea origin; Taiw: Out of Taiwan origin; WE: Western Eurasian origin

Paternal lineage proximity to Malagasy

Shared lineages

Among the haplogroups shared between Malagasy and Indonesians (Annex C: Table S5), four originated in Island Southeast Asia (C, O1a, O1a2, O2a1*), while six have western Eurasian origins (J1, J2, J2b, T*, L* and R1a). The Ma'anyan and five other Indonesian groups, all located around the Sulawesi sea (east Kalimantan Dayak, Java, Bali, Mandar and Sumba), share four of these Island Southeast Asian haplogroups. Importantly, Malagasy uniquely share just one subhaplogroup (O2a1a1-M88) with Ma'anyan, and this lineage has not been discovered in other regions of Indonesia. O2a1a1 may therefore be a marker of male genetic contributions from southern Borneo to Madagascar.

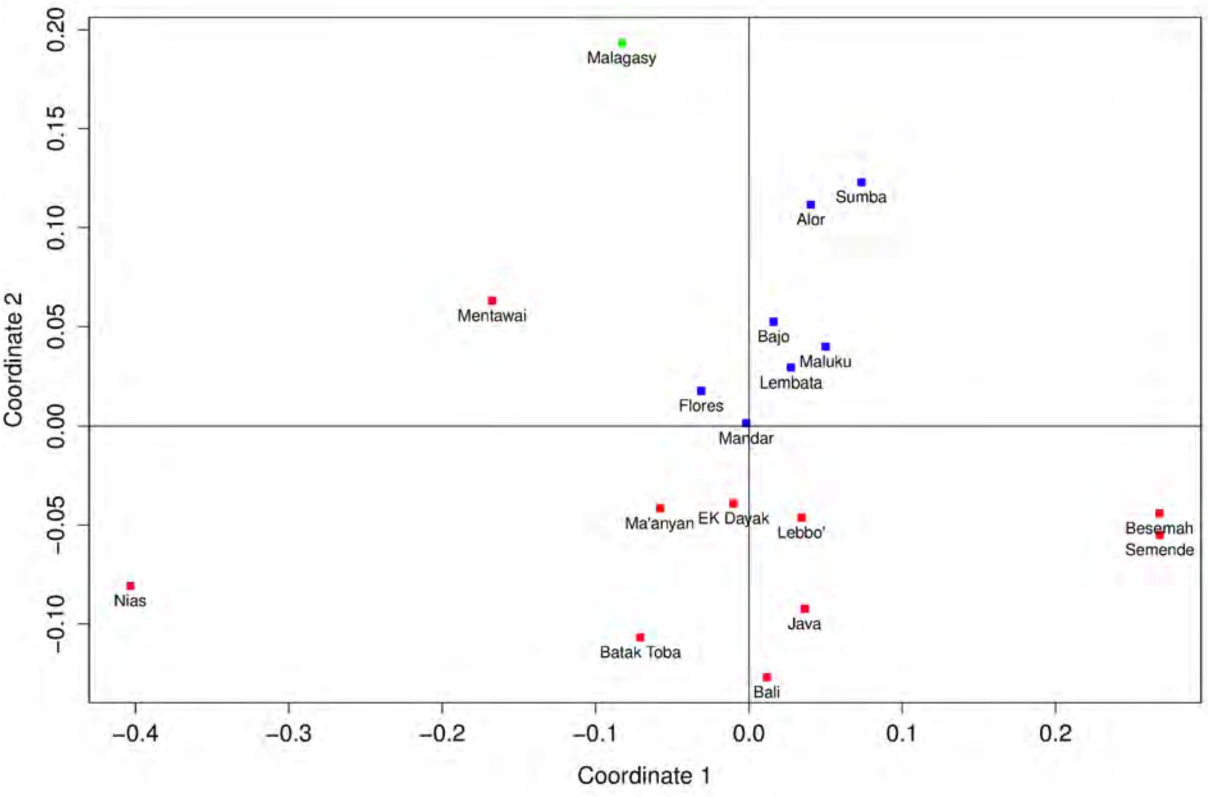


Figure 1. MDS plot showing F_{ST} values between Indonesian and Malagasy populations based on Y chromosome haplogroup frequencies (Kruskal stress: 0.149). Red: western Indonesians; blue: eastern Indonesians.

Shared Y chromosome haplogroups with a west Eurasian origin (J, T, L and R1a) (Annex C: Table S5) are also present in Indonesian populations, but only in the south and west of the Sulawesi sea. They occur at low frequency (<0.1%) in Java, Bali, Mandar and Bajo, but R1a is the only west Eurasian haplogroup identified in southern Borneo (Ma'anyan and east Kalimantan Dayak). However, west Eurasian lineages in Indonesia and Madagascar may result from independent dispersal events. Indeed, Indian and Arab traders have been active on both side of the Indian Ocean within the last three and one thousand years, respectively

(Ardika and Bellwood, 1991; Ardika et al., 1997; Calo, 2014; Donkin, 2004; Ferrand, 1891; Karafet et al., 2005). Therefore, west Eurasian haplogroups shared between Malagasy and Indonesians may have originated from Indonesia, or alternatively, they may have been obtained directly from southwestern Eurasia (the Middle East or India).

Population cross-comparisons

F_{ST} values based on Y chromosome haplogroup frequencies (Annex C: Table S6) were visualized on a multidimensional scaling (MDS) plot (Figure 1). Due to their statistically supported genetic homogeneity, Malagasy groups were pooled. The MDS plot (Figure 1) shows that Malagasy Y chromosome lineages are an outlier to Indonesian populations, in a similar way to certain Indonesian population outliers (Mentawai, Nias, Besemah, Semende). Y chromosome F_{ST} values between Malagasy and Indonesians are relatively high ($F_{ST} > 0.2$; Annex C: Table S6), mostly driven by the substantial African component of Malagasy (~65% of the paternal gene pool, but only ~30% of mtDNA (Capredon et al., 2013; Razafindrazaka et al., 2010)). No significant differences were observed to suggest specific genetic connections between Malagasy and eastern versus western Indonesians (Mann-Whitney U-test: $P = 0.06$). The Ma'anyan and other populations from Borneo cluster together with western Indonesian groups, including several population outliers (Mentawai, Nias, Besemah, Semende). The Bajo cluster with eastern Indonesian groups, consistent with the well-documented genetic division between western and eastern Indonesia broadly along the Wallace line.

When F_{ST} values are visualized with Surfer (Figure 2), the Indonesian populations with closest affinity to Malagasy (F_{ST} in the lower quartile of the range) are from regions near Wallace's line in the west and south of the Sulawesi sea (southern Sulawesi, eastern Borneo and the Lesser Sunda islands). Populations with highest affinity to Malagasy are Mandar (Sulawesi), Flores (Lesser Sunda), Bajo (Sulawesi), and east Kalimantan Dayak and Lebbo' (Borneo) (Annex C: Table S6). These results are supported by a linear optimization method, which aims to find the combination of Indonesian populations that most closely resembles the observed haplogroup diversity in Malagasy. This algorithm highlights two populations from the west and south of the Sulawesi sea, the Mandar (Sulawesi) and Lebbo' (Borneo), as populations that produce a Y chromosome genetic profile most closely resembling the observed pattern, which still accounting for the predominantly African genetic background found in Malagasy (Annex C: Figure S1).

These geographical regions comprised part of the trading sphere of the Srivijaya empire, including several Javanese kingdoms that played a crucial role in the region: Heluodan (5th century), Tarumanagara (5th century), Walaing (Chinese Heling, 7th-8th centuries), Kahuripan/Kediri (11th century), Singasari (13th century) and Majapahit (13th -15th centuries) (Beaujard, 2012b). This region also hosted several houseboat nomad groups (such as the Bajo), which had ample opportunities to incorporate men from a wider regional watershed. The Ma'anyan from southern Borneo do not show any privileged link with Malagasy (indeed, they have a relatively high F_{ST} value showing genetic differentiation), despite being the only Indonesian population that shares Y haplogroup O2a1a with Malagasy. This may indicate that the genetic contribution of Ma'anyan was limited, either due to the recent arrival of this lineage in Ma'anyan, or perhaps O2a1a has since been lost or is still undetected in other Indonesian populations.

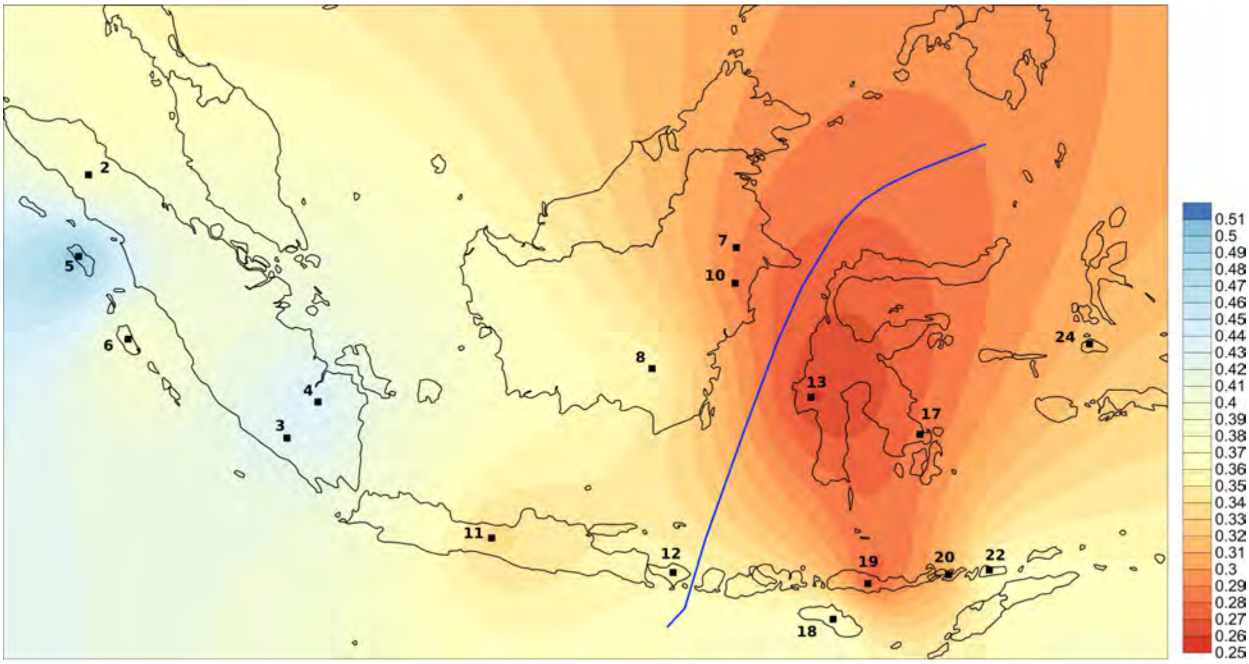


Figure 2. Map of Y chromosome F_{ST} values obtained by pairwise comparison between Malagasy and Indonesian populations. Dark red shading corresponds to lower pairwise FST values between Malagasy and Indonesian populations (represented by black squares), and dark blue to higher FST values. Note: 2. Batak Toba, 3. Besemah, 4. Semende, 5. Nias, 6. Mentawai, 7. Lebbo', 8. Ma'anyan, 10. EK Dayak, 11. Java, 12. Bali, 13. Mandar, 17. Bajo, 18. Sumba, 19. Flores, 20. Lembata, 22. Alor, 24. Maluku

Maternal lineage proximity to Malagasy

Shared lineages

Malagasy and Indonesians share mitochondrial haplogroups B4a1a, B4a1a1 (Polynesian motif), E1a1a, F3b, M7c1a4a, M32c and Q1 (Annex C: Table S7). Of these, B4a1a1, E1a1a and Q1 are found exclusively in eastern Indonesia. Conversely, F3b, B4a1a and M7c1a4a occur ubiquitously across both eastern and western Indonesia, and M32c has been observed in only one Javanese individual.

The Polynesian motif (B4a1a1) is considered strong evidence of Indonesian gene flow into Madagascar, where a variant is found at moderate frequency (11-50%). With the exception of Bali (Annex C: Table S4), B4a1a1 only occurs in eastern Indonesia. For the three new populations studied here, only the Bajo carry this Polynesian motif (just one of 27 individuals), and importantly, it was not found in any of our populations from Borneo. Furthermore, the specific Malagasy motif has not been found in Indonesia at all, including the new populations screened here. Considering the restricted geographic distribution of the Polynesian motif, it is most likely that this lineage from Madagascar traces back to eastern rather than western Indonesia.

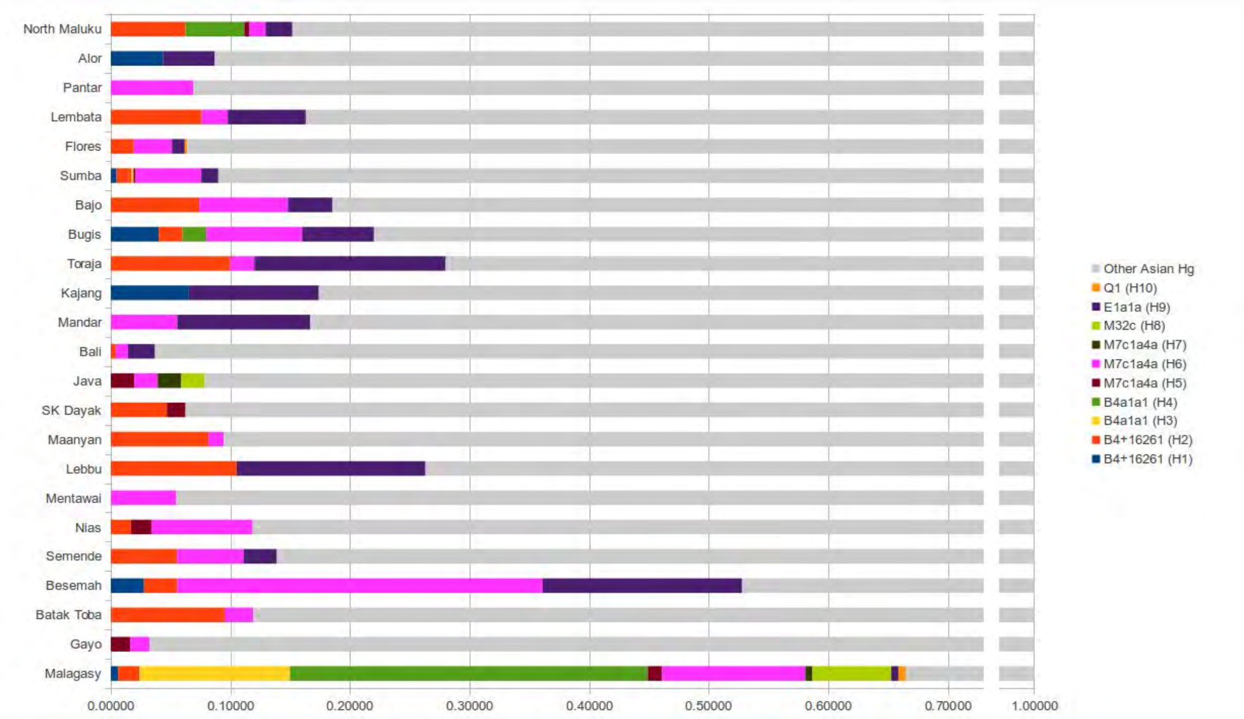


Figure 3. Asian mtDNA haplotypes shared between Malagasy and Indonesian populations.

Malagasy and Indonesians share ten haplotypes in seven haplogroups (Annex C: Table S8): two haplotypes each in B4a1a and B4a1a1; three haplotypes in M7c1a4a; and one haplotype in each of the other shared haplogroups. As shown in Figure 3, eastern Indonesian populations tend to share more haplotypes with Malagasy than western Indonesian groups. Populations from Sumba share the highest number of haplotypes (n = 6), followed by North Maluku and Sulawesi Bugis (n = 5). In a recurring theme, Ma'anyan exhibit limited sharing with only two haplotypes in common.

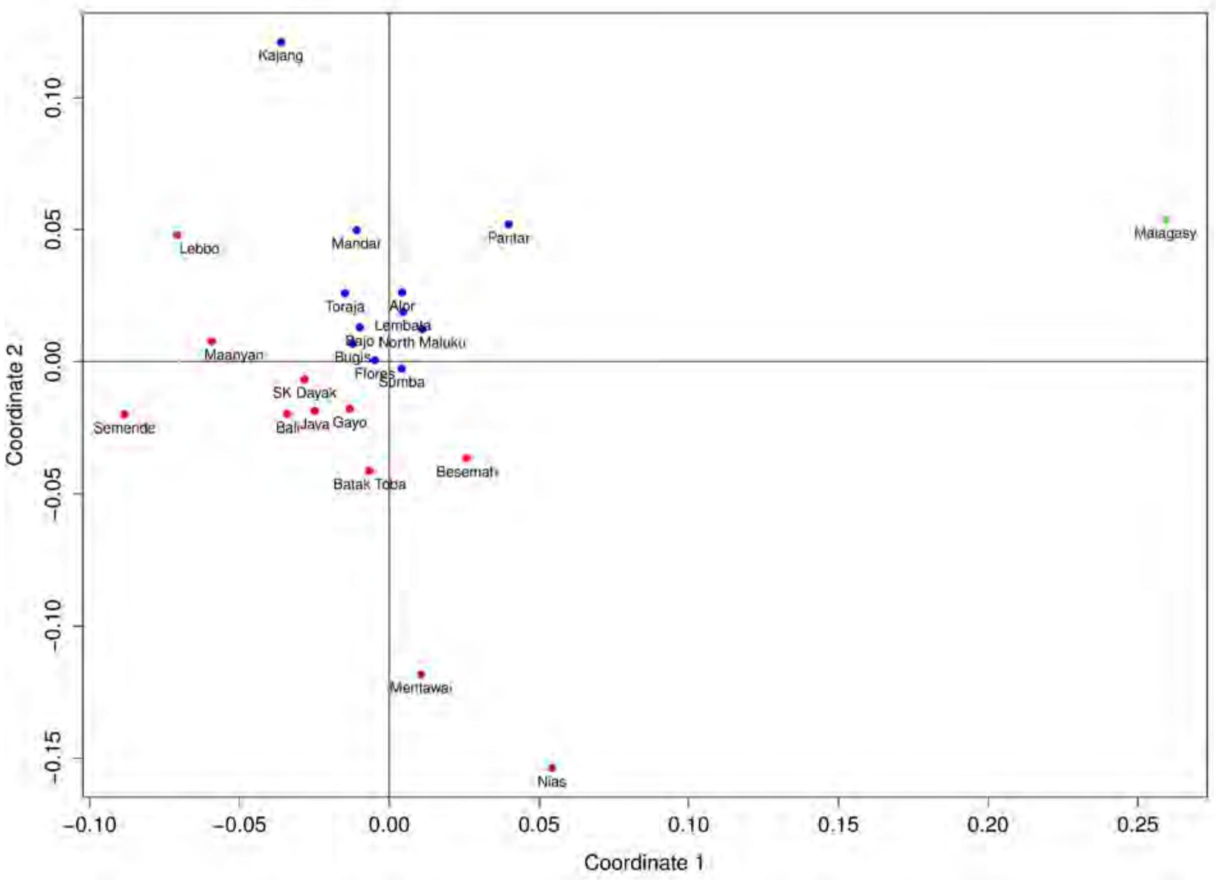


Figure 4. MDS plot showing F_{ST} values between Indonesian and Malagasy populations based on mtDNA haplogroup frequencies (Kruskal stress: 0.143). Red: western Indonesians; blue: eastern Indonesians.

Population cross-comparisons

The MDS plot (Figure 4, Malagasy groups again pooled), which is based on F_{ST} values from mtDNA haplogroup frequencies (Annex C: Table S4), shows that Malagasy maternal lineages differ markedly from those of Indonesians, while paternal lineages appear relatively closer (Figure 1). Unlike the Y chromosome data, which favors both eastern and western Indonesian sources, Malagasy are closer to the mtDNA diversity of eastern rather than western Indonesians (Mann-Whitney U test: $P < 0.01$). F_{ST} values (Annex C: Table S9) visualized in Surfer (Figure 5) show that the regions with closest affinity occur to the south and east of

Sulawesi, and support an eastern Indonesian connection. Populations with higher affinity to Malagasy (F_{ST} in the lower quartile of the range) include Sumba and Flores, the Maluku islands and the Bugis of south Sulawesi. As seen on the Y chromosome, the Bajo cluster with eastern Indonesian groups, while the Ma'anyan and other populations from Borneo cluster with western Indonesian groups. Linear optimization results broadly agree with the F_{ST} results: Malagasy are most likely derived from a combination of eastern Indonesian groups, such as North Maluku, Bugis and Bajo (Annex C: Figure S2).

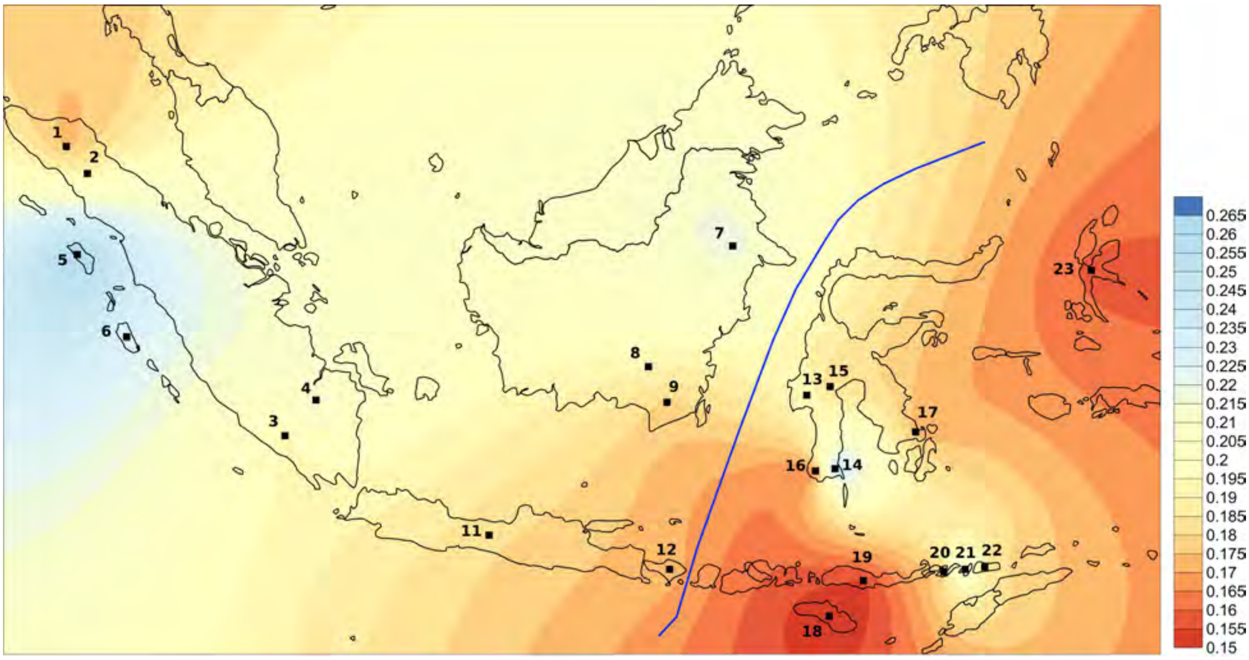


Figure 5. Map of mitochondrial DNA F_{ST} values obtained by pairwise comparison between Malagasy and Indonesian populations. Dark red shading corresponds to lower pairwise F_{ST} values between Malagasy and Indonesian populations (represented by black squares), and dark blue to higher F_{ST} values. Note: 1. Gayo, 2. Batak Toba, 3. Besemah, 4. Semende, 5. Nias, 6. Mentawai, 7. Lebbo', 8. Ma'anyan, 9. SK Dayak, 11. Java, 12. Bali, 13. Mandar, 14. Kajang, 15. Toraja, 16. Bugis, 17. Bajo, 18. Sumba, 19. Flores, 20. Lembata, 21. Pantar, 22. Alor, 23. North Maluku

These multiple lines of genetic evidence suggest that Malagasy women may have originated predominantly from eastern Indonesia. Key pieces of evidence include the restricted distribution of the Polynesian motif in eastern Indonesia and patterns of shared maternal lineages. This is consistent with the hypothesis that Austronesians borrowed the Polynesian motif – which perhaps arose in the Bismarck archipelago – from indigenous sources in eastern Indonesia (Oppenheimer and Richards, 2001; Soares et al., 2011). It is worth noting that eastern Indonesian influences occur in Madagascar: for cultivated plants (i.e., myths of origins, rituals for yams, the ancient importance of taro) (Beaujard, 2012a, 2011), and through the influence of the Orang Laut language on some Malagasy dialects (particularly the Vezo) (Dahl, 1991). Eastern Indonesia had no recorded sea-faring cultures involved in long-distance trading, except for the Bugis and Bajo. However, maritime foraging and trade were likely more common in earlier millennia as networks of long-distance sea-based interactions have been in place since at least the early Holocene (Bulbeck, 2008).

3.1.1.5. Discussion

The Southeast Barito language subgroup includes two languages spoken by populations separated by the 7,500 km expanse of the Indian Ocean: the Malagasy of Madagascar, off the east coast of Africa, and the Ma'anyan of Borneo, an island in western Indonesia. Knowing this linguistic connection, we investigated genetic linkages between these two populations. Our results suggest that few genetic connections exist, either on the paternal Y chromosome or the maternal mtDNA. These results suggest that 1) the Ma'anyan groups sampled here are not directly related to the individuals who settled Madagascar, 2) subsequent demographic events have erased any genetic affinity between them, or 3) the Ma'anyan were just one population of many that contributed to the settlement of Madagascar (a possibility suggested by the exclusive sharing of Y chromosome haplogroup O2a1a). The first hypothesis might suggest that other Southeast Barito groups from southern Borneo (such as the Samihim or the Dusun Witu (Adelaar, 1995b)) were involved instead.

These conclusions are drawn from a suite of complementary analyses, including phylogeography, haplotype sharing and linear optimization approaches. In combination, they paint a picture of the genetic dynamics between Indonesia and Madagascar. Although the geographic distribution of Indonesian populations that most closely reflect Malagasy genetic diversity are remarkably convergent for both male and female lineages, it is noticeable that the regions suggested by these two systems do not overlap perfectly. Our analyses suggest that populations from the south and west of the Sulawesi Sea (east Borneo, south Sulawesi and the Lesser Sunda islands) best explain Y chromosome diversity, while populations from the south and east of Sulawesi (south Sulawesi, the Lesser Sunda islands and the Maluku islands), all in eastern Indonesia, best explain mtDNA diversity. We emphasize that parts of these two regions overlap, thus potentially explaining both paternal and maternal affinity.

For instance, the distribution of shared lineages favors different source populations for maternal and paternal loci. Y chromosome haplogroup O2a1a is found only in Ma'anyan, while mtDNA haplogroups B4a1a1 and Q1 are found exclusively in eastern Indonesia. (Suggesting yet more connections, M32c has only been found in Java). F_{ST} and linear optimization results also highlight different source regions for the Y chromosome and mtDNA. Together, these patterns suggest that multiple regions may have contributed to the settlement of Madagascar, perhaps via one or a few admixed groups.

Sea nomads have been active traders along the eastern coast of Borneo, southern Sulawesi, the Lesser Sunda islands and the Maluku islands for at least the last few hundred years (Sather, 1997; Verheijen, 1986). These mobile populations linked western and eastern Indonesia, and absorbed individuals from different regions. Sea nomads traveled with their families, even on long distance journeys (Nuraini, 2008; Stacey, 2007). Moreover, languages of the Sama-Bajaw group, as spoken by the sea nomad Bajo, form a subgroup of the Barito languages of southeast Kalimantan (Blust, 2007), although not the closest language subgroup to Malagasy. Their patterns of genetic diversity and lifestyle make them possible contenders for the Indonesian populations who helped enact the settlement of Madagascar, although a definite assignment remains elusive.

3.1.1.6. Conclusion

We propose that the settlement of Madagascar had an Indonesian source location around southern Sulawesi, the Lesser Sunda islands and eastern Borneo. The populations involved may be related to modern sea nomad groups and the ancient Malay Srivijaya trading network. The Indonesian ancestors of Malagasy certainly carried maternal lineages with greater putative contributions from eastern Indonesia, and paternal lineages from both eastern and western Indonesia. The absence of any clear genetic connection between Malagasy and at least some populations speaking their most closely related language, Ma’anyan, raises important questions about the link between genes and language in the Indonesian dispersal across the Indian Ocean. Studying other Southeast Barito and sea nomad groups (such as the Orang Laut, who played a crucial role in the Srivijaya expansion) and the application of genome-wide genotyping technologies to provide additional genetic resolution promises to bring new insight to bear on these questions.

3.1.1.7. Miscellaneous

Abbreviations

SEB, Southeast Barito; SNP, Single Nucleotide Polymorphism; mtDNA, mitochondrial DNA; MDS, multidimensional scaling; SK Dayak, South Kalimantan Dayak; EK Dayak, East Kalimantan Dayak; lsei, Least Squares with Equalities and Inequalities.

Authors' contributions

P.K, M.P.C, and F-X.R designed the research, analyzed the data, and wrote the manuscript; P.K, H.L.S, H.S, and F-X.R collected the samples, P.K and L.T carried out the molecular labwork, D.P, H.R, N.B, and T.L helped draft the manuscript. All authors gave final approval for publication.

Acknowledgments

We thank Stephanie Schiavinato and Julie Bosc for technical assistance, and to Alexander Adelaar for helpful discussion. We also thank the Eijkman Institute for Molecular Biology (Jakarta, Indonesia) for accommodating the sampling program; Phillipe Grangé, Charles Illouz, and Chandra Nuraini (Université de La Rochelle) for facilitating access to the Bajo population; Budi Amuranto (Dinas Kebudayaan dan Pariwisata, East Kutai Sangata, Indonesia) for facilitating access to the Lebbo' population; Bambang Sugiyanto (Balai Arkeologi Banjarmasin, Indonesia); Adhi Agus Oktaviana (National Research Center for Archaeology, Indonesia); Antonio Guerreiro (CNRS UMR-6571, Université Aix-Marseille); and especially, all our Indonesian participants. This research was supported by the French ANR grant number ANR-14-CE31-0013-01 (grant OceoAdapto to F-X.R), the French ANR-12-PDOC-0037-01 (grant GENOMIX to D.P), the Region Aquitane of France (grant MAGE to D.P), the French Ministry of Foreign and European Affairs (French Archaeological Mission in Borneo (MAFBO) to F-X.R), the European Research Council (ERC 295918 MesAndLin(g)k to N.B.), and the French Embassy in Indonesia through its Cultural and Cooperation Services (Institut Français en Indonésie).

3.1.1.8. References

Adelaar, K.A., 2009. Loanwords in Malagasy, in: Haspelmath, M., Tadmor, U. (Eds.), Loanwords in the World's Languages: A Comparative Handbook. De Gruyter Mouton, Berlin, Germany, pp. 717–746.

Adelaar, K.A., 2006. The Indonesian migrations to Madagascar: making sense of the multidisciplinary evidence, in: Austronesian Diaspora and the Ethnogenesis of People in Indonesian Archipelago: Proceedings of the International Symposium. Presented at the Austronesian diaspora and the ethnogenesis of people in Indonesian archipelago, LIPI Press, Jakarta, Indonesia, pp. 205–232.

Adelaar, K.A., 2005. The Austronesian languages of Asia and Madagascar: a historical perspective, in: The Austronesian Languages of Asia and Madagascar. Routledge, New York, USA, pp. 1–42.

Adelaar, K.A., 1995a. Borneo as a cross-roads for comparative Austronesian linguistics, in: The Austronesians: Historical and Comparative Perspectives. ANU E Press, Canberra, Australia, pp. 81–102.

Adelaar, K.A., 1995b. L'importance du samihim (Bornéo du Sud) pour l'étymologie malgache, in: Champion, B. (Ed.), L'étrange intime. Mélanges offerts à Paul Ottino. Université de la Réunion, France.

Adelaar, K.A., 1989. Malay influence on Malagasy: linguistic and culture-historical implications. Ocean. Linguist. 28, 1–46. doi:10.2307/3622973

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., Howell, N., 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. 23, 147–147. doi:10.1038/13779

Ardika, I., Bellwood, P., 1991. Sembiran: the beginnings of Indian contact with Bali. Antiquity 65, 221–232.

Ardika, I.W., Bellwood, P., Sutaba, I.M., Yuliati, K.C., 1997. Sembiran and the first Indian contacts with Bali: an update. Antiquity 71, 193–194.

Beaujard, P., 2012a. Les mondes de l'océan Indien. Vol. 1 : De la formation de l'État au premier système-monde afro-eurasien (4e millénaire av. J.-C.-6e siècle apr. J.-C.). Armand Collin, Paris, France.

Beaujard, P., 2012b. Les mondes de l'ocean indien. Vol. 2 : L'océan Indien, au cœur des globalisations de l'Ancien Monde (7e-15e siècles). Armand Collin, Paris, France.

Beaujard, P., 2011. The first migrants to Madagascar and their introduction of plants: linguistic and ethnological evidence. Azania Archaeol. Res. Afr. 46, 169–189. doi:10.1080/0067270X.2011.580142

Beaujard, P., 2003. Les arrivées austronésiennes à Madagascar: vagues ou continuum? Études Océan Indien 59–147.

Bellwood, P., Fox, J.J., Tryon, D., 1995. The Austronesians in history: Common origins and diverse transformations, in: The Austronesians: Historical and Comparative Perspectives. ANU E Press, Canberra, Australia, pp. 1–16.

Blench, R., 2010. Evidence for the Austronesian voyages in the Indian Ocean, in: Anderson, A., Barrett, J.H., Boyle, K.V. (Eds.), The global origins and development of seafaring. McDonald Institute for Archeological Research, Cambridge, UK, pp. 239–248.

Blench, R.M., 2006. The Pleistocene settlement of the rim of the Indian Ocean. Presented at the 18th Congress of the Indo-Pacific Prehistory Association, Manila, Philippines.

Blust, R., 2011. Austronesian: A sleeping giant? Lang. Linguist. Compass 5, 538–550. doi:10.1111/j.1749-818X.2011.00297.x

Blust, R., 2007. The linguistic position of Sama-Bajaw. Stud. Philipp. Lang. Cult. 15, 73–114.

Bulbeck, D., 2008. An integrated perspective on the Austronesian diaspora: The switch from cereal agriculture to maritime foraging in the colonisation of Island Southeast Asia. Aust. Archaeol. 67, 31–52.

Burney, D.A., Burney, L.P., Godfrey, L.R., Jungers, W.L., Goodman, S.M., Wright, H.T., Jull, A.J.T., 2004. A chronology for late prehistoric Madagascar. J. Hum. Evol. 47, 25–63. doi:10.1016/j.jhevol.2004.05.005

Calo, A., 2014. Ancient trade between India and Indonesia. Science 345, 1255–1255. doi:10.1126/science.345.6202.1255

Capredon, M., Brucato, N., Tonasso, L., Choesmel-Cadamuro, V., Ricaut, F.-X., Razafindrazaka, H., Rakotondrabe, A.B., Ratolojanahary, M.A., Randriamarolaza, L.-P., Champion, B., Dugoujon, J.-M., 2013. Tracing arab-islamic inheritance in Madagascar: Study of the Y-chromosome and mitochondrial DNA in the Antemoro. PLoS ONE 8, e80932. doi:10.1371/journal.pone.0080932

Cox, M.P., Karafet, T.M., Lansing, J.S., Sudoyo, H., Hammer, M.F., 2010. Autosomal and X-linked single nucleotide polymorphisms reveal a steep Asian–Melanesian ancestry cline in eastern Indonesia and a sex bias in admixture rates. Proc. R. Soc. B Biol. Sci. 277, 1589–1596. doi:10.1098/rspb.2009.2041

Cox, M.P., Nelson, M.G., Tumonggor, M.K., Ricaut, F.-X., Sudoyo, H., 2012. A small cohort of Island Southeast Asian women founded Madagascar. Proc. R. Soc. B Biol. Sci. 279, 2761–2768. doi:10.1098/rspb.2012.0012

Cruciani, F., La Fratta, R., Santolamazza, P., Sellitto, D., Pascone, R., Moral, P., Watson, E., Guida, V., Colomb, E.B., Zaharova, B., Lavinha, J., Vona, G., Aman, R., Calì, F., Akar, N., Richards, M., Torroni, A., Novelletto, A., Scozzari, R., 2004. Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. Am. J. Hum. Genet. 74, 1014–1022. doi:10.1086/386294

Dahl, O.C., 1991. Migration from Kalimantan to Madagascar. Norwegian University Press : Institute for Comparative Research in Human Culture, Oslo, Norway.

Dahl, O.C., 1977. La subdivision de la famille Barito et la place du malgache. Acta Orient. 38, 77–134.

Dahl, O.C., 1951. Malgache et maanjan: une comparaison linguistique. Edege-Intituttet, Oslo, Norway.

Dewar, R.E., Wright, H.T., 1993. The culture history of Madagascar. J. World Prehistory 7, 417–466. doi:10.1007/BF00997802

Donkin, R.A., 2004. Between east and west: The Moluccas and the traffic in spices up to the arrival of Europeans. Amer Philosophical Society, Philadelphia, USA.

Donohue, M., Denham, T., 2010. Farming and language in Island Southeast Asia: Reframing Austronesian history. Curr. Anthropol. 51, 223–256. doi:10.1086/650991

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567. doi:10.1111/j.1755-0998.2010.02847.x

Ferrand, G., 1891. Les musulmans à Madagascar et aux iles Comores. E. Leroux, Paris, France.

Friedlaender, J.S., Friedlaender, F.R., Hodgson, J.A., Stoltz, M., Koki, G., Horvat, G., Zhadanov, S., Schurr, T.G., Merriwether, D.A., 2007. Melanesian mtDNA complexity. PLoS ONE 2, e248. doi:10.1371/journal.pone.0000248

Gunnarsdóttir, E.D., Nandineni, M.R., Li, M., Myles, S., Gil, D., Pakendorf, B., Stoneking, M., 2011. Larger mitochondrial DNA than Y-chromosome differences between matrilocal and patrilocal groups from Sumatra. Nat. Commun. 2, 228. doi:10.1038/ncomms1235

Hammer, M.F., Zegura, S.L., 2002. The human Y chromosome haplogroup tree: Nomenclature and phylogeography of its major divisions. Annu. Rev. Anthropol. 31, 303–321. doi:10.1146/annurev.anthro.31.040402.085413

Harich, N., Costa, M.D., Fernandes, V., Kandil, M., Pereira, J.B., Silva, N.M., Pereira, L., 2010. The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages. BMC Evol. Biol. 10, 138. doi:10.1186/1471-2148-10-138

Hurles, M.E., Sykes, B.C., Jobling, M.A., Forster, P., 2005. The dual origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from maternal and paternal lineages. Am. J. Hum. Genet. 76, 894–901. doi:10.1086/430051

Jordan, F.M., Gray, R.D., Greenhill, S.J., Mace, R., 2009. Matrilocal residence is ancestral in Austronesian societies. Proc. R. Soc. B Biol. Sci. 276, 1957–1964. doi:10.1098/rspb.2009.0088

Karafet, T., Lansing, J., Redd, A., Reznikova, S., 2005. Balinese Y-chromosome perspective on the peopling of Indonesia: Genetic contributions from pre-neolithic hunter- gatherers, Austronesian farmers, and Indian traders. Hum. Biol. 77, 93–114.

Karafet, T.M., Hallmark, B., Cox, M.P., Sudoyo, H., Downey, S., Lansing, J.S., Hammer, M.F., 2010. Major east–west division underlies Y chromosome stratification across Indonesia. Mol. Biol. Evol. 27, 1833–1844. doi:10.1093/molbev/msq063

Karafet, T.M., Mendez, F.L., Meilerman, M.B., Underhill, P.A., Zegura, S.L., Hammer, M.F., 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res. 18, 830–838. doi:10.1101/gr.7172008

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. doi:10.1093/molbev/mst010

Kayser, M., Brauer, S., Weiss, G., Schiefenhövel, W., Underhill, P., Shen, P., Oefner, P., Tommaseo-Ponzetta, M., Stoneking, M., 2003. Reduced Y-Chromosome, but not mitochondrial DNA: Diversity in human populations from West New Guinea. Am. J. Hum. Genet. 72, 281–302. doi:10.1086/346065

Kumar, A., 2011. “The single most astonishing fact of human geography”: Indonesia’s far west colony. Indonesia 92, 59–96.

Lansing, J.S., Cox, M.P., de Vet, T.A., Downey, S.S., Hallmark, B., Sudoyo, H., 2011. An ongoing Austronesian expansion in Island Southeast Asia. J. Anthropol. Archaeol. 30, 262–272. doi:10.1016/j.jaa.2011.06.004

Lewis, M.P., Simons, G.F., Fennig, C.D. (Eds.), 2016. Ethnologue: Languages of the world, Nineteenth edition. ed. SIL International, Texas.

Marck, J., 2008. Proto Oceanic society was matrilineal. J. Polyn. Soc. 117, 345–382.

Mendez, F.L., Karafet, T.M., Krahn, T., Ostrer, H., Soodyall, H., Hammer, M.F., 2011. Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa. Hum. Biol. 83, 39–53. doi:10.3378/027.083.0103

Mona, S., Grunz, K.E., Brauer, S., Pakendorf, B., Castrì, L., Sudoyo, H., Marzuki, S., Barnes, R.H., Schmidtke, J., Stoneking, M., Kayser, M., 2009. Genetic admixture history of eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. Mol. Biol. Evol. 26, 1865–1877. doi:10.1093/molbev/msp097

Nuraini, C., 2008. Langue et production de récits d’une communauté Bajo des îles Kangean (Indonésie) (Ph.D. Thesis). Univ. de La Rochelle.

Oppenheimer, S., Richards, M., 2001. Fast trains, slow boats, and the ancestry of the Polynesian islanders. Sci. Prog. 84, 157–181.

Pamjav, H., Fehér, T., Németh, E., Pádár, Z., 2012. Brief communication: New Y-chromosome binary markers improve phylogenetic resolution within haplogroup R1a1. Am. J. Phys. Anthropol. 149, 611–615. doi:10.1002/ajpa.22167

Pelras, C., 1997. The Bugis, 1st edition. ed. Wiley-Blackwell, Oxford, UK.

Peng, M.-S., Quang, H.H., Dang, K.P., Trieu, A.V., Wang, H.-W., Yao, Y.-G., Kong, Q.-P., Zhang, Y.-P., 2010. Tracing the Austronesian footprint in Mainland Southeast Asia: A perspective from mitochondrial DNA. Mol. Biol. Evol. 27, 2417–2430. doi:10.1093/molbev/msq131

Pierron, D., Razafindrazaka, H., Pagani, L., Ricaut, F.-X., Antao, T., Capredon, M., Sambo, C., Radimilahy, C., Rakotoarisoa, J.-A., Blench, R.M., Letellier, T., Kivisild, T., 2014. Genome-wide evidence of Austronesian–Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. Proc. Natl. Acad. Sci. 111, 936–941. doi:10.1073/pnas.1321860111

Razafindrazaka, H., 2010. Le peuplement humain de Madagascar : Anthropologie génétique de trois groupes traditionnels (Ph.D. Thesis). Université Paul Sabatier - Tolouse III, Toulouse, France.

Razafindrazaka, H., Ricaut, F.-X., Cox, M.P., Mormina, M., Dugoujon, J.-M., Randriamarolaza, L.P., Guitard, E., Tonasso, L., Ludes, B., Crubezy, E., 2010. Complete mitochondrial DNA sequences provide new insights into the Polynesian motif and the peopling of Madagascar. Eur. J. Hum. Genet. 18, 575–581. doi:10.1038/ejhg.2009.222

Rosa, A., Ornelas, C., Brehm, A., Villems, R., 2006. Population data on 11 Y-chromosome STRs from Guiné-Bissau. Forensic Sci. Int. 157, 210–217. doi:10.1016/j.forsciint.2005.04.005

Salas, A., Richards, M., De la Fe, T., Lareu, M.-V., Sobrino, B., Sánchez-Diz, P., Macaulay, V., Carracedo, Á., 2002. The making of the African mtDNA landscape. Am. J. Hum. Genet. 71, 1082–1111. doi:10.1086/344348

Salas, A., Richards, M., Lareu, M.-V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V., Carracedo, Á., 2004. The African diaspora: Mitochondrial DNA and the Atlantic slave trade. Am. J. Hum. Genet. 74, 454–465. doi:10.1086/382194

Sather, C., 1997. The Bajau Laut: adaptation, history, and fate in a maritime fishing society of south-eastern Sabah. Oxford University Press, Oxford, UK.

Serva, M., Petroni, F., Volchenkov, D., Wichmann, S., 2012. Malagasy dialects and the peopling of Madagascar. J. R. Soc. Interface 9, 54–67. doi:10.1098/rsif.2011.0228

Soares, P., Rito, T., Trejaut, J., Mormina, M., Hill, C., Tinkler-Hundal, E., Braid, M., Clarke, D.J., Loo, J.-H., Thomson, N., Denham, T., Donohue, M., Macaulay, V., Lin, M., Oppenheimer, S., Richards, M.B., 2011. Ancient voyaging and Polynesian origins. Am. J. Hum. Genet. 88, 239–247. doi:10.1016/j.ajhg.2011.01.009

Soodyall, H., Jenkins, T., Stoneking, M., 1995. “Polynesian” mtDNA in the Malagasy. Nat. Genet. 10, 377–378. doi:10.1038/ng0895-377

Stacey, N., 2007. Boats to burn: Bajo fishing activity in the Australian fishing zone. ANU E Press, Canberra, Australia.

Tishkoff, S.A., Gonder, M.K., Henn, B.M., Mortensen, H., Knight, A., Gignoux, C., Fernandopulle, N., Lema, G., Nyambo, T.B., Ramakrishnan, U., Reed, F.A., Mountain, J.L., 2007. History of click-speaking populations of Africa inferred from mtDNA and Y Chromosome genetic variation. Mol. Biol. Evol. 24, 2180–2195. doi:10.1093/molbev/msm155

Tofanelli, S., Bertoncini, S., Castrì, L., Luiselli, D., Calafell, F., Donati, G., Paoli, G., 2009. On the origins and admixture of Malagasy: New evidence from high-resolution analyses of paternal and maternal lineages. Mol. Biol. Evol. 26, 2109–2124. doi:10.1093/molbev/msp120

Tumonggor, M.K., Karafet, T.M., Hallmark, B., Lansing, J.S., Sudoyo, H., Hammer, M.F., Cox, M.P., 2013. The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific. J. Hum. Genet. 58, 165–173. doi:10.1038/jhg.2012.154

Underhill, P.A., Myres, N.M., Rootsi, S., Metspalu, M., Zhivotovsky, L.A., King, R.J., Lin, A.A., Chow, C.-E.T., Semino, O., Battaglia, V., Kutuev, I., Järve, M., Chaubey, G., Ayub, Q., Mohyuddin, A., Mehdi, S.Q., Sengupta, S., Rogaev, E.I., Khusnutdinova, E.K., Pshenichnov, A., Balanovsky, O., Balanovska, E., Jeran, N., Augustin, D.H., Baldovic, M., Herrera, R.J., Thangaraj, K., Singh, V., Singh, L., Majumder, P., Rudan, P., Primorac, D., Villems, R., Kivisild, T., 2010. Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. Eur. J. Hum. Genet. 18, 479–484. doi:10.1038/ejhg.2009.194

Van den Meersche, K., Soetaert, K., Middelburg, J.J., 2008. A Bayesian Compositional Estimator for microbial taxonomy based on biomarkers. Limnol. Oceanogr. Methods 6, 190–199. doi:10.4319/lom.2008.6.190

van Oven, M., Kayser, M., 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. 30, E386–E394. doi:10.1002/humu.20921

Verheijen, J.A.J., 1986. The Sama/Bajau language in the Lesser Sunda islands. Department of Linguistics, Research School of Pacific Studies, Australian National University, Canberra, Australia.

Wood, E.T., Stover, D.A., Ehret, C., Destro-Bisol, G., Spedini, G., McLeod, H., Louie, L., Bamshad, M., Strassmann, B.I., Soodyall, H., Hammer, M.F., 2005. Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. Eur. J. Hum. Genet. 13, 867–876. doi:10.1038/sj.ejhg.5201408

Xu, S., Pugach, I., Stoneking, M., Kayser, M., Jin, L., 2012. Genetic dating indicates that the Asian–Papuan admixture through Eastern Indonesia corresponds to the Austronesian expansion. Proc. Natl. Acad. Sci. 109, 4574–4579. doi:10.1073/pnas.1118892109

3.1.2. West Eurasian genetic influences in the Indonesian archipelago

(Published in Quaternary International, 416, September 19th 2016. doi:10.1016/j.quaint. 2015.06.048)

Pradiptajati Kusuma[a,b], Murray P. Cox[c], Nicolas Brucato[d], Herawati Sudoyo[b], Thierry Letellier[a], François-Xavier Ricaut[a*]

[a] Laboratoire d'Anthropologie Moléculaire et Imagérie de Synthèse UMR-5288, Université de Toulouse, Toulouse, France
[b] Laboratory of Human Genetic Diversity, Eijkman Institute for Molecular Biology, Jakarta, Indonesia
[c] Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
[d] Center for Linguistics, University of Leiden, Leiden, Netherlands

3.1.2.1. Abstract

Western Eurasia, notably the Near East and South Asia (Indian sub-continent), has interacted with Indonesia through Indian Ocean trade (the Maritime Silk Route) for more than 2,000 years. The Indianization, and later Islamization, of Indonesia was enacted largely through trading activities, but also spread with help from the many Indianized and Islamic kingdoms that reigned over parts of the Indonesian archipelago during this time. Western Eurasian interaction left behind not only imported trade goods and cultural features, but also genetic traces. To locate the primary areas of Western Eurasian genetic influence in Indonesia, we have assembled published uniparental genetic data from ~2,900 Indonesian individuals. Frequency distributions show that Western Eurasian paternal lineages are found more commonly than maternal lineages. Furthermore, the origins of Western Eurasian paternal lineages are more diverse than maternal lineages, predominantly tracing back to South West and South Asia, and the Indian sub-continent, respectively. Indianized kingdoms in the Indonesian archipelago likely played a major role in dispersing Western Eurasian lineages, as these kingdoms overlap geographically with the current distribution of individuals carrying Western Eurasian genetic markers. Our data highlight the important role of these Western Eurasian migrants in contributing to the complexity of genetic diversity across the Indonesian archipelago today.

Keywords: Western Eurasia; India; Arabia; Indonesia; mitochondrial DNA; Y chromosome.

3.1.2.2. Introduction

Western Eurasia (WE), notably the Near East and South Asia, had established very close contact with Southeast Asia since 4th B.C.E (Bellina & Glover, 2004), and spread to Island South East Asia (particularly Indonesia) through cultural and trading networks in Maritime Silk Route as early as the 1st century B.C.E. (Ardika and Bellwood, 1991; Ardika et al., 1997; Calo, 2014; Lawler, 2014). Indianization and Islamization in Indonesia led to the development of Hindu, and later Islamic, kingdoms between the 5th and 15th centuries, which reinforced interactions and influence from Western Eurasia to all the regions in western and central Indonesia spanned by these kingdoms (Beaujard, 2012a, 2012b; Gonda, 1975; Kanchan, 1990). These interactions brought not only new ideas and trade goods into the Indonesian archipelago, but were also accompanied to some degree by gene flow, which contributed to a number of modern Indonesian populations.

Considering the historical events of the last two millennia, it has been variously postulated that WE genetic influx into Indonesia could have originated from India (Lansing, 1983; Lukas, 2003; Mabbett, 1977), the Near East (Jacobsen, 2009) or Europe (Dutch colonialism) (Taylor, 2009). The genetic evidence so far suggests that the primary contribution came from the Indian subcontinent (HUGO Pan-Asian SNP Consortium et al., 2009; Karafet et al., 2005, 2010), but there is still debate regarding the importance of Indian gene flow, and its geographic diffusion into and impact on the Indonesian archipelago, with some scholars proposing colonization by Indian exiles (Majumdar, 1963), large scale migration establishing Indian colonies in South East Asia (Mabbett, 1977), or more punctuated contact by Indian traders (Lansing, 1983; Sandhu and Mani, 1993). Historical evidence mentions the “constant flux of people from Gujarat and neighboring regions” (in the Indian sub-continent) to Java (Beaujard, 2012b, 2012a). Indeed, genetic traces of paternal Indian ancestry have been detected at low frequency in Indonesia (<10%) (Karafet et al., 2005, 2010) across several islands (Sumatra, Java, Borneo and Bali), but in comparison, no Indian maternal lineages have been detected (Tumonggor et al., 2013). The influx of Indian ancestry has also been observed from genome-wide analysis using a limited number of nuclear SNPs (HUGO Pan-Asian SNP Consortium et al., 2009) in just a few populations from western Indonesia (Sumatra). However, Indian markers in autosomal data were not detected in large-scale genotyping of eleven populations across Island Southeast Asia (Pugach et al., 2013).

Our study is based on the largest dataset of maternal (mtDNA) and paternal (Y chromosome) lineages assembled for the Indonesian archipelago to date, and aims to map West Eurasian genetic influences in this region. We propose that genetic contributions from West Eurasia (such as India) are mostly restricted to the areas where Indianized kingdoms were most powerful. We also show that genetic contact with both Indian men and women was involved.

3.1.2.3. Material and methods

DNA samples and data sources

Indonesian samples used in this study have been described previously (Gunnarsdóttir et al., 2011; Karafet et al., 2010; Kusuma et al., 2015; Tumonggor et al., 2013). In brief, the mtDNA dataset includes 2,841 individuals, and the Y chromosome dataset includes 2,095 individuals, from 7 major Indonesian islands and representing 22 and 17 different ethnic groups, respectively (Table 1). These two datasets represent the largest Indonesian genetic assemblage studied to date.

Table 1. List of populations used in this study

Group	Islands	Code	Population	N mtDNA	N Y chr	Ref.
Western Indonesia	Sumatra	1	Gayo	62	*NA*	Eijkman Institute archived samples
		2	Batak Toba	42	37	Karafet et al., 2010; Tumonggor et al., 2013
		3	Besemah	36	38	Gunnarsdottir et al. 2011
		4	Semende	36	37	Gunnarsdottir et al. 2011
		5	Nias	59	60	Karafet et al., 2010; Tumonggor et al., 2013
		6	Mentawai	128	74	Karafet et al., 2010; Tumonggor et al., 2013
	Borneo	7	Lebbo'	19	15	Kusuma et. al., 2014.
		8	Ma'anyan	159	90	Kusuma et. al., 2014.
		9	SK Dayak*	64	*NA*	Eijkman Institute archived samples
		10	EK Dayak*	*NA*	85	Karafet et al., 2010
	Java	11	Java	51	61	Karafet et al., 2010; Tumonggor et al., 2013
	Bali	12	Bali	487	634	Karafet et al., 2010; Tumonggor et al., 2013
Eastern Indonesia	Sulawesi	13	Mandar	54	54	Karafet et al., 2010; Tumonggor et al., 2013
		14	Kajang	46	*NA*	Tumonggor et al., 2013
		15	Toraja	50	*NA*	Tumonggor et al., 2013
		16	Bugis	50	*NA*	Tumonggor et al., 2013
		17	Bajo	27	27	Kusuma et. al., 2014.
	Lesser Sunda	18	Sumba	634	349	Karafet et al., 2010; Tumonggor et al., 2013
		19	Flores	469	388	Karafet et al., 2010; Tumonggor et al., 2013
		20	Lembata	92	89	Karafet et al., 2010; Tumonggor et al., 2013
		21	Pantar	29	*NA*	Tumonggor et al., 2013
		22	Alor	23	27	Karafet et al., 2010; Tumonggor et al., 2013
	Maluku	23	North Maluku	224	*NA*	Eijkman Institute archived samples
		24	Maluku (Hiri and Ternate)	*NA*	30	Karafet et al., 2010

Note : NA = Not Available for analysis

* Samples of SK Dayak were collected from various Dayak ethnics in South Kalimantan, while the EK Dayak were collected from various Dayak et

Haplogroup assignment

Mitochondrial DNA hypervariable region I sequences were compiled from published sources (Table 1) and aligned against the revised Cambridge Reference Sequence (rCRS) (Andrews et al., 1999) using the MAFFT aligner v.7 (Katoh and Standley, 2013). Mitochondrial haplogroups were determined with the Haplogrep program (http://haplogrep.uibk.ac.at) based on Phylotree v.16 (van Oven and Kayser, 2009). Y chromosome haplogroups were assigned from SNPs data based on the updated ISOGG Y-DNA haplogroup tree (International Society of Genetic Genealogy, 2014) and the Y-Phylotree (van Oven et al., 2014).

Statistical analysis

Mitochondrial DNA and Y chromosome West Eurasian haplogroup frequencies in Indonesian populations were plotted geographically with Surfer v.12.0 using the Kriging method. Mitochondrial DNA and Y chromosome haplogroup diversity was calculated using Nei's method (Nei, 1987) as implemented in Arlequin v.3.5 (Excoffier and Lischer, 2010) with 5,040 permutations for statistical testing. To observe Western Eurasian and Asian components in both the paternal and maternal datasets, Principal Component Analysis (PCA) based on mtDNA and Y chromosome haplogroup frequencies (Annex D: Table S1 and S2) in populations grouped by island was performed using R, and visualized with a biplot graph implemented in the ggbiplot package.

3.1.2.4. Results

We observed 13 West Eurasian mtDNA and 15 West Eurasian Y chromosome haplogroups in 9 (out of 22) and 7 (out of 17) Indonesian populations, respectively (Annex D: Table S1 and S2). West Eurasian Y chromosome haplogroups are more frequent (4.49%) and diverse (0.0222-0.439) than West Eurasian mtDNA haplogroups (1.55% and 0.00320-0.242) (Annex D: Table S3). This is in agreement with the putative geographical origins of the WE haplogroups found in Indonesia. Indeed the source locations of WE mtDNA and Y haplogroups found in Indonesian populations do not overlap perfectly (Annex D: Table S1 and S2). The 13 WE mtDNA haplogroups are primarily derived from South Asia (mainly the Indian sub-continent). Conversely, the 15 WE Y chromosome haplogroups in Indonesia exhibit more diverse origins, yet South Asia remains the most frequent point of origin. The most common Y chromosome haplogroup is R1a1a (M17), which was observed in 5 of the 7 Indonesian populations that carry WE haplogroups, followed by R2a in 4 populations. Outside Indonesia, R1a1a and R2a (M124) have similar distributions with relatively high frequencies in South Asia (Sengupta et al., 2006; Underhill et al., 2010). For mtDNA, M5 is the most frequent haplogroup found in Indonesia, observed in 4 of 9 populations that carry WE haplogroups. Outside Indonesia, M5 is mainly found in India (Sahoo and Kashyap, 2006; Sun et al., 2006; Thangaraj et al., 2006).

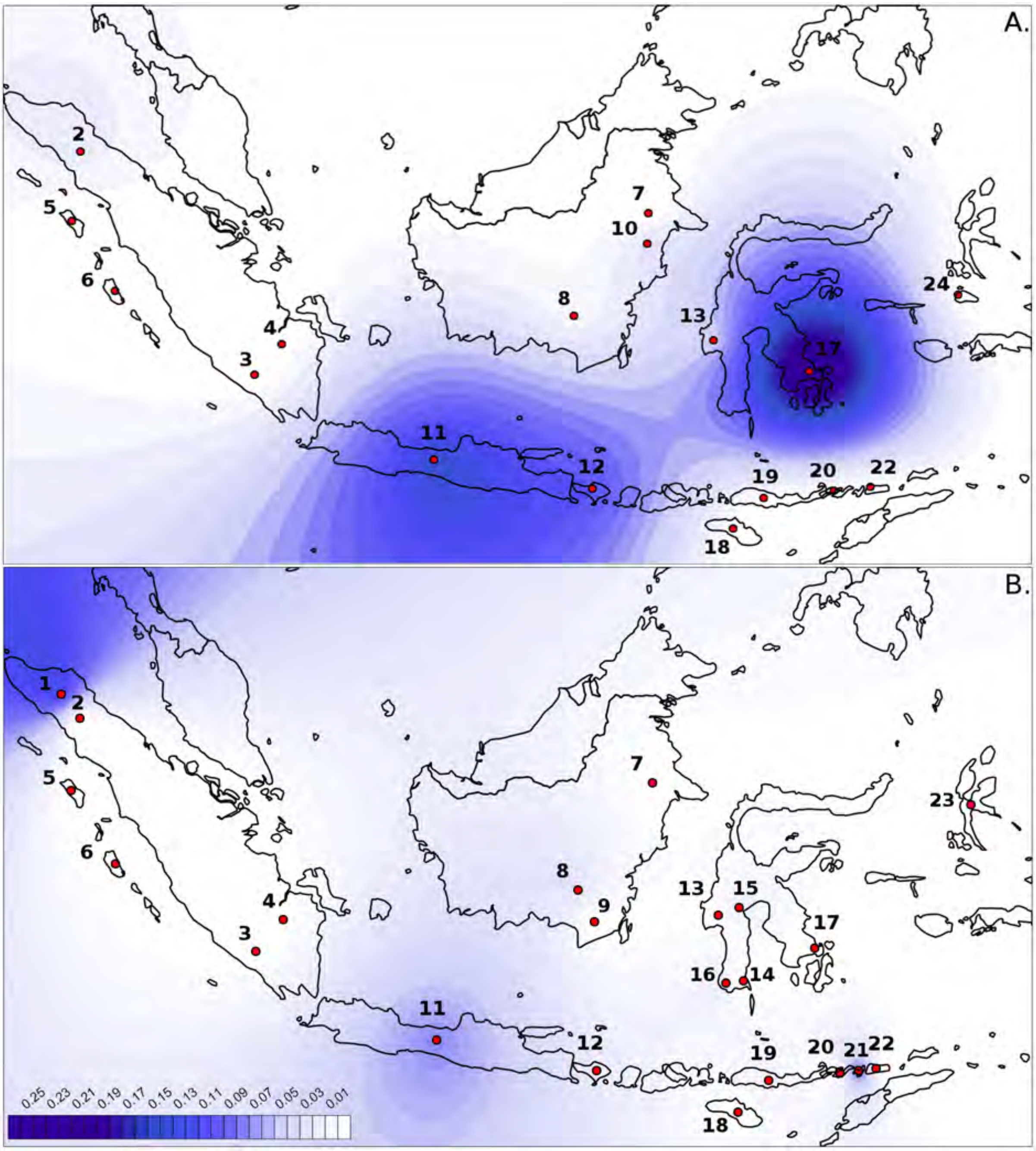


Figure 1. Frequency of West Eurasian A) Y chromosome and B) mtDNA haplogroups among Indonesian populations (indicated by red dots). Dark shaded areas have higher frequencies. Data and populations details are provided in Tables 1 and A1 and A2. Note: 1. Gayo, 2. Batak Toba, 3. Besemah, 4. Semende, 5. Nias, 6. Mentawai, 7. Lebbo’, 8. Ma’anyan, 9. SK Dayak, 10. EK Dayak, 11. Java, 12. Bali, 13. Mandar, 14. Kajang, 15. Toraja, 16. Bugis, 17. Bajo, 18. Sumba, 19. Flores, 20. Lembata, 21. Pantar, 22. Alor, 23. North Maluku, 24. Maluku

The distribution of WE haplogroups across Indonesia is depicted in Figure 1. West Eurasian Y haplogroups are clustered in central Indonesia, and geographically restricted to populations to the south of the Celebes sea, with the exception of one individual in northern Sumatra. WE Y chromosome haplogroups reach highest frequency in Bajo from Sulawesi (25.9%; 7 of 27 individuals), Java (14.7%; 9 of 61 individuals), and Bali (11.0%; 70 of 634 individuals). A broadly similar pattern can be observed from the distribution of WE mtDNA haplogroups, which are also detected in populations in south of the Celebes sea, with the exception of a population from Gayo in north Sumatra. WE mtDNA haplogroups reach highest frequency in Gayo (12.9%; 8 of 62 individuals), followed by Pantar (10.3%; 3 of 29 individuals), and Java (7.8%; 4 of 51 individuals).

A PCA plot of the distribution of Western Eurasian versus Asian lineages among Indonesia populations (grouped by island) also supports these results (Figure 2). The strongest West Eurasian influence on the paternal gene pool can be seen in populations from Java, Bali, and Sulawesi. The strongest WE influences on the maternal gene pool are found in Java, Sumatra, and the Lesser Sunda islands. Java is the only population that received a strong WE contribution for both paternal and maternal lineages (14.8% and 7.84%, respectively).

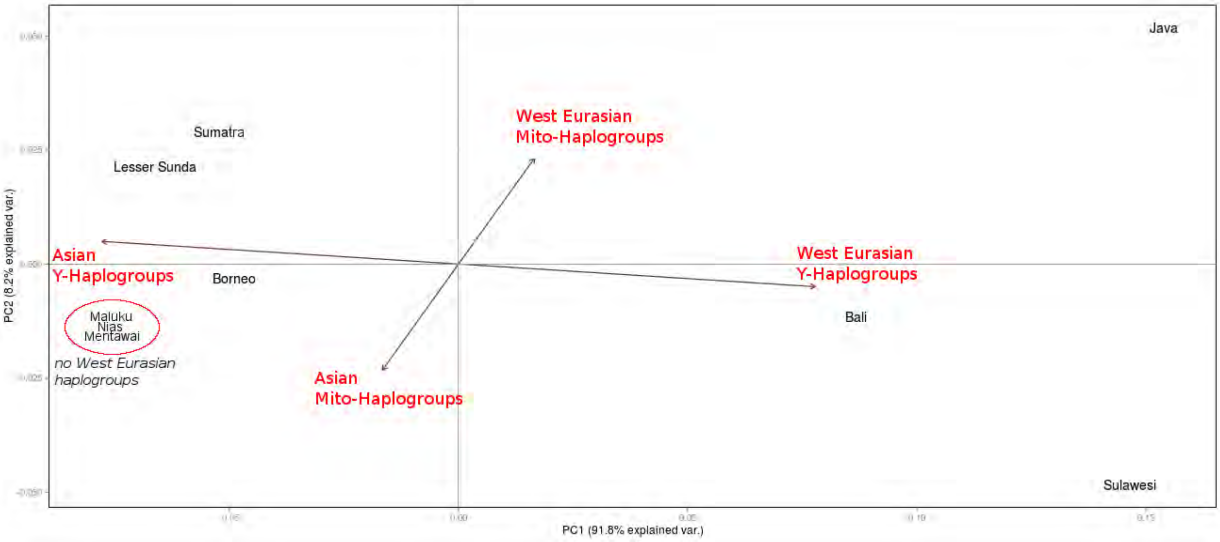


Figure 2. PCA plot showing the 1st and 2nd principal components computed from West Eurasian mtDNA and Y chromosome haplogroup frequencies in Indonesia (grouped by island).

3.1.2.5. Discussion

New archaeological evidence shows that a network of interactions moved goods and ideas across the Indian Ocean at least 2,000 years B.C.E (Lawler, 2014). Island South East Asia, and particularly the Indonesian archipelago, joined this network later in the 1st century B.C.E., with connections initially mostly with India (Ardika and Bellwood, 1991; Ardika et al., 1997; Calo, 2014). This promoted West Eurasian influences in Indonesia, and Indianized cultures became widespread in Indonesia shortly afterwards. Typical influences included old writings in Sanskrit, shadow puppets (wayang) and dances, sculptures, vast Hindu temples in Java and Bali, and modified socio-cultural structures (such as the caste system in Bali) (Coedes, 1968; Covarrubias and Vickers, 2008; Dumarcay, 1986; Guy, 2014; Kanchan, 1990; Mellema, 1954; Pigeaud, 1960; Sumarsam, 1995). Indian influences were accepted and frequently assimilated by Indonesians, and spread through many Hindu and Islamic kingdoms in Indonesia, such as Heluodan (5th century), Tarumanagara (5th century), Walaing (7th-8th centuries), Kahuripan/Kediri (11th century), Singhasari (13th century), and Majapahit (13th-15th centuries) (Beaujard, 2012a, 2012b). In parallel to cultural and trading influences, movements of people from South and Southwest Asia to Indonesia also likely occurred, but their impact on the gene pool of modern Indonesia has been debated and required further investigation.

Our study shows that a West Eurasian genetic contribution exists at non-negligible levels. Frequency distributions of West Eurasian maternal and paternal lineages in Indonesian populations suggest that 1) about the same number of Indonesian populations (~40%) carry WE paternal and/or maternal haplogroups, 2) populations with the highest WE contributions are found in central Indonesia, to the south of the Celebes sea, 3) WE input is more frequent in the paternal (4.49%) than maternal (1.55%) Indonesian gene pool, and 4) the geographical origin of WE maternal and paternal haplogroups suggest a preponderance of haplogroup sources in South Asia (the Indian sub-continent).

WE paternal lineages are more frequent than maternal lineages (4.49% versus 1.55%), but also seem to be sourced from across a broader geographical area (South West and South Asia) (Annex D: Table A1 and A2) than maternal lineages, which are mainly derived from the Indian sub-continent. This may be due to a male bias in trading activities, which are presumed to have been dominated by men, and also fits the hypothesis that Indian influence was not restricted only to trading/cultural activities, but involved a substantial movement of men, and importantly women, from the Indian sub-continent into Island South East Asia (Beaujard, 2012a, 2012b; Mabbett, 1977).

Interestingly, traces of Western Eurasian genetic lineages are mostly observed in regions where ancient Hindu and Islamic kingdoms were located. These kingdoms governed their regions in separate, but often overlapping, time frames from the 7th to 16th centuries. For example, Srivijaya was based in Sumatra; Majapahit, Singasari and Kahuripan/Kediri in Java; Kutai in Kalimantan and Sulawesi; and Buleleng in Bali and neighboring islands in the Lesser Sunda chain (Beaujard, 2012a, 2012b; Kanchan, 1990; Pigeaud, 1960; Wolters, 1967). Later, Islamic kingdoms arose in northern Sumatra around the 14th century, influenced culturally by Arab (Near East) or Gujarat traders (India) (Ali, 1975; Ricklefs, 1993). In this study, WE traces are found in East Kalimantan, Sulawesi, Java, Bali, and northern Sumatra for both paternal and maternal lineages, and in the Lesser Sunda islands for maternal lineages. This indicates that the influence of Indianized/Islamic kingdoms in Indonesia was not restricted simply to cultural assimilation, trading and economical or political power, but was also associated to some degree with gene flow into Indonesia from cultural and/or trading centers in South and South West Asia.

3.1.2.6. Conclusion

West Eurasian influences into Island South East Asia helped shape the genetic background of modern Indonesian populations. Despite their relatively small genetic impact, aside from its huge cultural impact, Western Eurasian lineages are found in both the paternal and maternal Indonesian gene pools suggesting that contact extended beyond cultural and trading activities to at least some small-scale population movements coupled with long-term settlement. The exact small-scale geographical sources of these West Eurasian contributions still need to be clarified, but genetic evidence indicates slightly different origins for paternal and maternal Western Eurasian lineages, while emphasizing the major role played by the Indian sub-continent. Further genetic studies will need to be performed using larger sample sizes and a broader array of genetic markers (including genome-wide genotyping) in order to provide additional information on the exact sources, timing, and modality of the West Eurasian contribution into Indonesia.

3.1.2.7. Miscellaneous

Acknowledgements

This research was supported by the French ANR grant number ANR-14-CE31-0013-01 (grant OceoAdapto to F-X.R), the French Ministry of Foreign and European Affairs (French Archaeological Mission in Borneo (MAFBO) to F-X.R), the European Research Council (ERC 295918 grant MesAndLin(g)k to N.B), and the French Embassy in Indonesia through its Cultural and Cooperation Services (Institut Français en Indonésie).

3.1.2.8. References

Ali, A.M., 1975. Islam in Indonesia, in: Kähler, H. (Ed.), Handbook of Oriental Studies. E.J. Brill, Leiden, Netherlands.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., Howell, N., 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. 23, 147–147. doi:10.1038/13779

Ardika, I., Bellwood, P., 1991. Sembiran: the beginnings of Indian contact with Bali. Antiquity 65, 221–232.

Ardika, I.W., Bellwood, P., Sutaba, I.M., Yuliati, K.C., 1997. Sembiran and the first Indian contacts with Bali: an update. Antiquity 71, 193–194.

Beaujard, P., 2012a. Les mondes de l'ocean indien. Vol. 2 : L'océan Indien, au cœur des globalisations de l'Ancien Monde (7e-15e siècles). Armand Collin, Paris, France.

Beaujard, P., 2012b. Les mondes de l'océan Indien. Vol. 1 : De la formation de l'État au premier système-monde afro-eurasien (4e millénaire av. J.-C.-6e siècle apr. J.-C.). Armand Collin, Paris, France.

Calo, A., 2014. Ancient trade between India and Indonesia. Science 345, 1255–1255. doi:10.1126/science.345.6202.1255

Coedes, G., 1968. The Indianized states of Southeast Asia. University of Hawaii Press, Honolulu, USA.

Covarrubias, M., Vickers, A., 2008. Island of Bali. Periplus Editions, Singapore.

Dumarcay, J., 1986. The temples of Java. Oxford University Press Australia and New Zealand, Melbourne, Australia.

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567. doi:10.1111/j.1755-0998.2010.02847.x

Gonda, J., 1975. The Indian religions in pre-Islamic Indonesia and their survival in Bali, in: Kähler, H. (Ed.), Handbook of Oriental Studies. E.J. Brill, Leiden, Netherlands.

Gunnarsdóttir, E.D., Nandineni, M.R., Li, M., Myles, S., Gil, D., Pakendorf, B., Stoneking, M., 2011. Larger mitochondrial DNA than Y-chromosome differences between matrilocal and patrilocal groups from Sumatra. Nat. Commun. 2, 228. doi:10.1038/ncomms1235

Guy, J., 2014. Lost Kingdoms: Hindu-Buddhist sculpture of early Southeast Asia. Metropolitan Museum of Art, New York, USA.

HUGO Pan-Asian SNP Consortium, Abdulla, M.A., Ahmed, I., Assawamakin, A., Bhak, J., Brahmachari, S.K., Calacal, G.C., Chaurasia, A., Chen, C.-H., Chen, J., Chen, Y.-T., Chu, J., Cutiongco-de la Paz, E.M.C., De Ungria, M.C.A., Delfin, F.C., Edo, J., Fuchareon, S., Ghang, H., Gojobori, T., Han, J., Ho, S.-F., Hoh, B.P., Huang, W., Inoko, H., Jha, P., Jinam, T.A., Jin, L., Jung, J., Kangwanpong, D., Kampuansai, J., Kennedy, G.C., Khurana, P., Kim, H.-L., Kim, K., Kim, S., Kim, W.-Y., Kimm, K., Kimura, R., Koike, T., Kulawonganunchai, S., Kumar, V., Lai, P.S., Lee, J.-Y., Lee, S., Liu, E.T., Majumder, P.P., Mandapati, K.K., Marzuki, S., Mitchell, W., Mukerji, M., Naritomi, K., Ngamphiw, C., Niikawa, N., Nishida, N., Oh, B., Oh, S., Ohashi, J., Oka, A., Ong, R., Padilla, C.D., Palittapongarnpim, P., Perdigon, H.B., Phipps, M.E., Png, E., Sakaki, Y., Salvador, J.M., Sandraling, Y., Scaria, V., Seielstad, M., Sidek, M.R., Sinha, A., Srikummool, M., Sudoyo, H., Sugano, S., Suryadi, H., Suzuki, Y., Tabbada, K.A., Tan, A., Tokunaga, K., Tongsima, S., Villamor, L.P., Wang, E., Wang, Y., Wang, H., Wu, J.-Y., Xiao, H., Xu, S., Yang, J.O., Shugart, Y.Y., Yoo, H.-S., Yuan, W., Zhao, G., Zilfalil, B.A., Indian Genome Variation Consortium, 2009. Mapping human genetic diversity in Asia. Science 326, 1541–1545. doi:10.1126/science.1177074

International Society of Genetic Genealogy, 2014. ISOGG 2014 Y-DNA Haplogroup Tree, Version: 9.70 [WWW Document]. URL http://www.isogg.org/tree/ (accessed 6.3.14).

Jacobsen, F.F., 2009. Hadrami Arabs in Present-day Indonesia: An Indonesia-oriented group with an Arab signature. Routledge, New York, USA.

Kanchan, R.K., 1990. Hindu kingdoms of South-East Asia, Indian historical researches ; Cosmo Publications, New Delhi, India.

Karafet, T., Lansing, J., Redd, A., Reznikova, S., 2005. Balinese Y-chromosome perspective on the peopling of Indonesia: Genetic contributions from pre-neolithic hunter- gatherers, Austronesian farmers, and Indian traders. Hum. Biol. 77, 93–114.

Karafet, T.M., Hallmark, B., Cox, M.P., Sudoyo, H., Downey, S., Lansing, J.S., Hammer, M.F., 2010. Major east–west division underlies Y chromosome stratification across Indonesia. Mol. Biol. Evol. 27, 1833–1844. doi:10.1093/molbev/msq063

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. doi:10.1093/molbev/mst010

Kusuma, P., Cox, M.P., Pierron, D., Razafindrazaka, H., Brucato, N., Tonasso, L., Suryadi, H.L., Letellier, T., Sudoyo, H., Ricaut, F.-X., 2015. Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations. BMC Genomics 16, 191. doi:10.1186/s12864-015-1394-7

Lansing, J.S., 1983. The “Indianization” of Bali. J. Southeast Asian Stud. 14, 409–421.

Lawler, A., 2014. Sailing Sinbad’s seas. Science 344, 1440–1445. doi:10.1126/science.344.6191.1440

Lukas, H., 2003. Theories of Indianization exemplified by selected case studies from Indonesia (Insular Southeast Asia), in: Sanskrit in Southeast Asia: The harmonizing factor of cultures. Presented at the International Sanskrit Conference, Mahachulalongkornrajavidyalaya Press, Bangkok, Thailand, pp. 82–107.

Mabbett, I.W., 1977. The “Indianization” of Southeast Asia: Reflections on the Historical Sources. J. Southeast Asian Stud. 8, 143–161.

Majumdar, R.C., 1963. Ancient Indian colonization in South-east Asia. B. J. Sandesara, Calcutta, India.

Mellema, R.L., 1954. Wayang puppets. carving, colouring and symbolism, First. ed. Koninklijk Instituut voor de Tropen, Amsterdam, Netherlands.

Nei, M., 1987. Molecular evolutionary genetics. Columbia University Press, New York, USA.

Pigeaud, T.G., 1960. Java in the 14th century: a study in cultural history : the Nāgara-Kĕrtāgama by Rakawi Prapañca of Majapahit, 1365 A.D. Martinus Nijhoff, Hague, Netherlands.

Pugach, I., Delfin, F., Gunnarsdóttir, E., Kayser, M., Stoneking, M., 2013. Genome-wide data substantiate Holocene gene flow from India to Australia. Proc. Natl. Acad. Sci. 110, 1803–1808. doi:10.1073/pnas.1211927110

Ricklefs, M.C., 1993. A History of modern Indonesia since C. 1300. Stanford University Press, California, USA.

Sahoo, S., Kashyap, V. k., 2006. Phylogeography of mitochondrial DNA and Y-Chromosome haplogroups reveal asymmetric gene flow in populations of Eastern India. Am. J. Phys. Anthropol. 131, 84–97. doi:10.1002/ajpa.20399

Sandhu, K.S., Mani, A., 1993. Indian communities in Southeast Asia. Institute of Southeast Asian Studies, Singapore.

Sengupta, S., Zhivotovsky, L.A., King, R., Mehdi, S.Q., Edmonds, C.A., Chow, C.-E.T., Lin, A.A., Mitra, M., Sil, S.K., Ramesh, A., Usha Rani, M.V., Thakur, C.M., Cavalli-Sforza, L.L., Majumder, P.P., Underhill, P.A., 2006. Polarity and temporality of high-resolution Y-Chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. Am. J. Hum. Genet. 78, 202–221.

Sumarsam, 1995. Gamelan: Cultural interaction and musical development in Central Java. University of Chicago Press, Chicago, USA.

Sun, C., Kong, Q.-P., Palanichamy, M. gounder, Agrawal, S., Bandelt, H.-J., Yao, Y.-G., Khan, F., Zhu, C.-L., Chaudhuri, T.K., Zhang, Y.-P., 2006. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. Mol. Biol. Evol. 23, 683–690. doi:10.1093/molbev/msj078

Taylor, J.G., 2009. The social world of Batavia: Europeans and Eurasians in colonial Indonesia, 2nd ed. University of Wisconsin Press, Wisconsin, USA.

Thangaraj, K., Chaubey, G., Singh, V.K., Vanniarajan, A., Thanseem, I., Reddy, A.G., Singh, L., 2006. In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup “M” in India. BMC Genomics 7, 151. doi:10.1186/1471-2164-7-151

Tumonggor, M.K., Karafet, T.M., Hallmark, B., Lansing, J.S., Sudoyo, H., Hammer, M.F., Cox, M.P., 2013. The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific. J. Hum. Genet. 58, 165–173. doi:10.1038/jhg.2012.154

Underhill, P.A., Myres, N.M., Rootsi, S., Metspalu, M., Zhivotovsky, L.A., King, R.J., Lin, A.A., Chow, C.-E.T., Semino, O., Battaglia, V., Kutuev, I., Järve, M., Chaubey, G., Ayub, Q., Mohyuddin, A., Mehdi, S.Q., Sengupta, S., Rogaev, E.I., Khusnutdinova, E.K., Pshenichnov, A., Balanovsky, O., Balanovska, E., Jeran, N., Augustin, D.H., Baldovic, M., Herrera, R.J., Thangaraj, K., Singh, V., Singh, L., Majumder, P., Rudan, P., Primorac, D., Villems, R., Kivisild, T., 2010. Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. Eur. J. Hum. Genet. 18, 479–484. doi:10.1038/ejhg.2009.194

van Oven, M., Kayser, M., 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. 30, E386–E394. doi:10.1002/humu.20921

van Oven, M., Van Geystelen, A., Kayser, M., Decorte, R., Larmuseau, M.H., 2014. Seeing the wood for the trees: A minimal reference phylogeny for the human Y chromosome. Hum. Mutat. 35, 187–191. doi:10.1002/humu.22468

Wolters, O.W., 1967. Early Indonesian commerce: A study of the origins of Srivijaya. Cornell University Press, New York, USA.

3.2. Autosomal DNA analyses

In the previous sub-chapter, the analyses of uniparental markers generated robust results regarding the two main aims of this thesis – the origin of Indonesian genetic diversity and its relationship to the genetic background of the Malagasy population. However, it was not possible to accurately determine the Asian group(s) that settled Madagascar, or undertake a deep investigation of Indonesian genetic diversity. Thus, in this sub-chapter, I present my research on the analyses of autosomal genome-wide SNP data from new Indonesian populations, and bring key new elements to the aims. I published three articles on this topic. The first was published in the journal **Scientific Reports**, entitled "*Contrasting linguistic and genetic influences during the Austronesian settlement of Madagascar*", the second is *in press*. in the journal **European Journal of Human Genetics**, entitled "*The last sea nomads of the Indonesian archipelago: Genomic origins and dispersal*", and the third was published in the journal **Molecular Biology and Evolution**, entitled "*Malagasy genetic ancestry comes from an historical Malay trading post in Southeast Borneo*". The first article focuses on questions 1 and 3 (see p13), and explores the autosomal genetic background of the Ma'anyan, the linguistic-sister population of the Malagasy; it assesses whether the Ma'anyan have a close genetic affinity to the Malagasy. The second article investigates the genetic history of the Indonesian Bajo sea-nomads, and how their way of life shapes the genetic diversity of central Indonesia populations. Later, I also tested their affinity with the Malagasy to determine if the sea nomads contributed directly to the westward migration into Madagascar. This article gives additional insight into questions 1, 3, and 4 (see p13). The third article examines the genetic affinities of other Indonesian populations living in the vicinity of the Ma'anyan in South Borneo, especially Banjar and Ngaju populations, to the Malagasy, and finally manages to identify the Asian ancestor of the Malagasy, hence reconciling linguistic, historical and genetic evidence. This article gives strong answers to the general question addressed throughout this thesis on the genetic connection between Indonesia and Madagascar.

3.2.1. Contrasting linguistic and genetic influences during the Austronesian settlement of Madagascar

(Published in Scientific Reports, 6:26066, May 18th 2016. doi:10.1038/srep26066)

Pradiptajati Kusuma[1,2], Nicolas Brucato[1], Murray P. Cox[3], Denis Pierron[1], Harilanto Razafindrazaka[1], Alexander Adelaar[4], Herawati Sudoyo[2,5], Thierry Letellier[1], and François-Xavier Ricaut[1,*]

[1] Evolutionary Medicine Group, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse UMR-5288, Université de Toulouse, Toulouse, France
[2] Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular Biology, Jakarta, Indonesia
[3] Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
[4] Asia Institute, University of Melbourne, Melbourne, Australia
[5] Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia

3.2.1.1. Abstract

The Austronesian expansion, one of the last major human migrations, influenced regions as distant as tropical Asia, Remote Oceania and Madagascar, off the east coast of Africa. The identity of the Asian groups that settled Madagascar is particularly mysterious. While language connects Madagascar to the Ma’anyan of southern Borneo, haploid genetic data are more ambiguous. Here, we screened genome-wide diversity in 211 individuals from the Ma’anyan and surrounding groups in southern Borneo. Surprisingly, the Ma’anyan are characterized by a distinct, high frequency genomic component that is not found in Malagasy. This novel genetic layer occurs at low levels across Island Southeast Asia and hints at a more complex model for the Austronesian expansion in this region. In contrast, Malagasy show genomic links to a range of Island Southeast Asian groups, particularly from southern Borneo, but do not have a clear genetic connection with the Ma’anyan despite the obvious linguistic association.

3.2.1.2. Introduction

The Austronesian expansion was a major human migration in Southeast Asia, triggered by the spread of agricultural populations approximately 5,000 years ago (Bellwood et al., 1995; Bellwood, 2007; Oppenheimer and Richards, 2001).Thought to have originated in Taiwan, its influence spread through Philippines and Indonesian archipelago, ultimately impacting a wide geographical area ranging from Remote Oceania in the east, to Madagascar and the eastern coast of Africa in the west (Bellwood, 2007; Blench, 2006; Soares et al., 2011). This expansion had outsized cultural and genetic impact on these territories, but the populations caught up in the dispersal were regionally different and diverse across the Indo-Pacific. This created a diverse modern range of Austronesian populations with their own cultural traits and genetic heritage, among which Madagascar is a unique case.

Despite clear evidence, based on biological (Hurles et al., 2005; Kusuma et al., 2015; Pierron et al., 2014; Soodyall et al., 1995; Tofanelli et al., 2009) and linguistic data (Adelaar, 2009; Serva et al., 2012) of Malagasy's mixed ancestry with both African and Southeast Asian groups, identifying the parental populations of Malagasy and clarifying the process of settling Madagascar around the middle of the first millennium AD (Burney, 1987; Burney et al., 2004; MacPhee and Burney, 1991) has remained complex. Language studies have identified many linguistic characters that relate Malagasy to languages spoken in Borneo, notably in the Southeast Barito region. This includes much vocabulary and structural linguistic agreement shared between Malagasy and Southeast Barito languages, which form a subgroup of West Malayo-Polynesian languages in the Austronesian language family (Adelaar, 1989, 2005, 2006, 2009, Dahl, 1951, 1977, 1991). Among the communities speaking Southeast Barito languages, the Ma'anyan show linguistic characteristics that place them as the closest known Asian parental population to Malagasy (Adelaar, 1995; Dahl, 1951, 1977, 1991; Lewis et al., 2016). Curiously, the Ma'anyan are an indigenous ethnic group representing approximately 70,000 individuals, who live in remote inland areas of central and southeastern Kalimantan (the Indonesian part of the island of Borneo). Today, the Ma'anyan are largely agricultural, cultivating dry rice on shifting fields, but also gathering forest products (Hudson, 1967). They do not exhibit any particular mastery of seafaring technologies or navigational knowledge (Adelaar, 1995), raising questions about how a closely related language travelled across the vast Indian Ocean and came to be spoken in Madagascar. However, in historical times, the south Borneo coastline was split by a gulf that may have extended 200 kilometres into the interior (Adelaar, 2017; Van Bemmelen, 1949), thus potentially placing Ma'anyan communities that are firmly inland today in what was then a formerly coastal environment.

Several genetic studies have sought to detect Indonesian genetic connections in the Malagasy genome (including mitochondrial DNA, Y chromosome and autosomal markers) (Hurles et al., 2005; Kusuma et al., 2015; Pierron et al., 2014; Soodyall et al., 1995; Tofanelli et al., 2009), but no clear parental groups in Southeast Asia have yet been identified. The limited geographical coverage of Indonesian populations in these studies (including the absence of key populations such as the Ma'anyan) has often prevented precise conclusions. The possibility that the Ma'anyan are the Asian parental source of Malagasy was first explored genetically using uniparental markers (mitochondrial DNA and the Y chromosome) only in 2015 (Kusuma et al., 2015). This preliminary study, which covered a range of Southeast Asian groups, linked the origins of the Asian genetic components in Malagasy to modern populations located between Sulawesi (eastern Indonesia) and eastern Borneo (western Indonesia), thus confirming the general results of earlier studies (Tofanelli et al., 2009). However, surprisingly, the Ma'anyan shared few mtDNA or Y chromosome lineages with Malagasy. Given this apparent contradiction between linguistic evidence and genetic analyses of uniparental markers, and to overcome the potential bias of this lineage-based approach (which is more sensitive to genetic drift), a genome-wide analysis of Southeast Borneo individuals was deemed necessary to better explore the link between Madagascar and Borneo.

Here, we perform that genome-wide analysis in the Ma'anyan and other groups from southern Borneo to determine the genetic background and potential Asian sources of the Malagasy. Using Illumina HumanOmniExpress Bead Chips, we genotyped over 700,000 genomic markers in 169 Ma'anyan individuals, together with a further 42 individuals from Dayak ethnic groups across southern Borneo. The aims of this study were dual: i) to examine the genetic diversity of populations in southeastern Borneo (focusing on the Ma'anyan and other indigenous Dayak groups), and thereby determine their place in the wider genetic diversity of Island Southeast Asia; and ii) to identify whether the clear linguistic relationship between the Ma'anyan and Malagasy is also reflected in a shared genetic inheritance.

3.2.1.2. Methods

Sample collection and ethics

A total of 211 DNA samples were analysed from two groups in Borneo: The Ma'anyan ethnic group (169 individuals), and the South Kalimantan Dayak, which comprises a mixed assemblage of diverse Dayak ethnic groups (42 individuals) (Figure 1 and Annex E: Table S1). The samples used in this study have been described previously (Kusuma et al., 2015). Briefly, blood samples were collected from healthy adult donors, all of whom provided written informed consent. DNA was extracted using a standard salting-out procedure. All participants were surveyed for language affiliation, current residence, familial birthplaces, and a genealogy of four generations to establish ancestry. This study was approved by the Research Ethics Commission of the Eijkman Institute for Molecular Biology (Jakarta, Indonesia), and the methods were carried out in accordance with the approved guidelines. Genome-wide SNP genotypes for the two groups were generated using the Illumina Human Omni Express-24 v1.0 Bead Chip (Illumina Inc., San Diego, CA), which surveys 730,525 single nucleotide markers regularly spaced across the genome. Genotyping data are available upon request.

Dataset integration

Two datasets were compiled from previous published data to fulfill key analytical criteria: i) the low density dataset has wide geographical coverage, but includes relatively few SNPs; while ii) the high density dataset has greatly increased SNP density, but includes fewer populations. This approach, which is necessitated by the wide range of DNA genotyping chip technologies used by the scientific community (Supplementary Table S5), allows us to address the widest range of questions.

Filtering and quality controls were performed using PLINK v1.9 (Chang et al., 2015): i) to avoid close relatives, relatedness was measured between all pairs of individuals within each population using an Identity-by-Descent (IBD) estimation with upper threshold of 0.25 (second degree relatives); ii) SNPs that failed the Hardy-Weinberg exact (HWE) test ($P < 10^{-6}$) were excluded; iii) samples with an overall call rate <0.99 and individual SNPs with missing rates >0.05 across all samples in each population were excluded; and iv) variants in high linkage disequilibrium ($r^2 > 0.5$; 50 SNP sliding windows) were also removed for the low density dataset.

The final low density dataset contained 9,743 SNPs in 1,817 individuals from 73 populations, after excluding 7 Ma'anyan and 2 South Kalimantan Dayak individuals for reasons of low data quality. This low density dataset includes East and Southeast Asian populations (Mörseburg et al., 2016), Indonesian populations including the Lebbo' and Bajo (Pierron et al., 2014), and groups from Sumba (Cox et al., 2016), together with CEPH-HGDP data (Li et al., 2008), HUGO Pan-Asian SNP data (HUGO Pan-Asian SNP Consortium et al., 2009) and data for three Malagasy populations (Mikea, Vezo and Temoro) (Pierron et al., 2014) (Annex E: Table S1). The final high density dataset comprises a subset of the populations in the low density dataset, specifically covering 311,871 SNPs in 820 individuals from 28 populations.

Population structure analysis

The low density dataset was analysed using the following approaches. Genetic diversity was described using pairwise FST distance calculations and Principal Components Analysis using the 'smartpca' algorithm of EIGENSOFT v6.0.1 (Patterson et al., 2006). The Runs of Homozygosity (ROH) analysis was performed in PLINK v1.9 from the linkage-disequilibrium-pruned dataset. ADMIXTURE v1.23 (Alexander et al., 2009) was used to estimate the profile of individual genomic ancestries using maximum likelihood for components $K = 2$ to $K = 20$. Ten replicates were run at each value of K with different random seeds, then merged and assessed for clustering quality using CLUMPP (Jakobsson and Rosenberg, 2007), and the cross-validation value was calculated to determine the optimal number of genomic components (here, $K = 14$). ADMIXTURE and PCA plots were generated with Genesis and the results were confirmed using the high density dataset, to avoid any misinterpretation due to a potential bias driven by the density of SNPs. Gene flow between populations was investigated using two different approaches: i) SNP frequencies using TreeMix v1.12 (Pickrell and Pritchard, 2012), with blocks of 200 SNPs to account for linkage disequilibrium and migration edges added sequentially until the model explained 99% of the variance (the TreeMix outputs in Newick format were visualized with MEGA6 (Tamura et al., 2013)); and three-population (f3) statistics (Patterson et al., 2012), defining the African Yoruba population as an outgroup for the low density dataset; and ii) haplotype sharing using the Refined IBD algorithm of Beagle v.4.0 (Browning and Browning, 2007) visualized with Cytoscape v.3.2.1 (Shannon et al., 2003) using the high density dataset to estimate the total number of shared genetic fragments (logarithm of odds ratio>3) between each pair of individuals.

To characterize the Island Southeast Asian ancestry in Malagasy individuals, we discarded estimated African components using PCAdmix (Brisbin et al., 2012). First, genome-wide SNP data from Malagasy, Yoruba and Asian samples (represented by the Ma'anyan, the Igorot and the Bajo to cover a range of Asian diversity) of the high density dataset were phased using Beagle v4.0. The Yoruba and Asian samples comprised 100 randomly selected individuals, and were defined as 'parental' populations compared to the Malagasy 'daughter' population for the purposes of the PCAdmix software. The ancestry of each defined linkage disequilibrium window was estimated by the Viterbi algorithm for each individual and used to mask all potential African SNPs. The masked Malagasy dataset was merged with the high density dataset, trimmed to 17,043 overlapping SNPs, and used to find the closest Indonesian populations that match the Malagasy Asian component using FST distances, an ancestry-specific PCA in EIGENSOFT v6.0.1 and a TreeMix analysis.

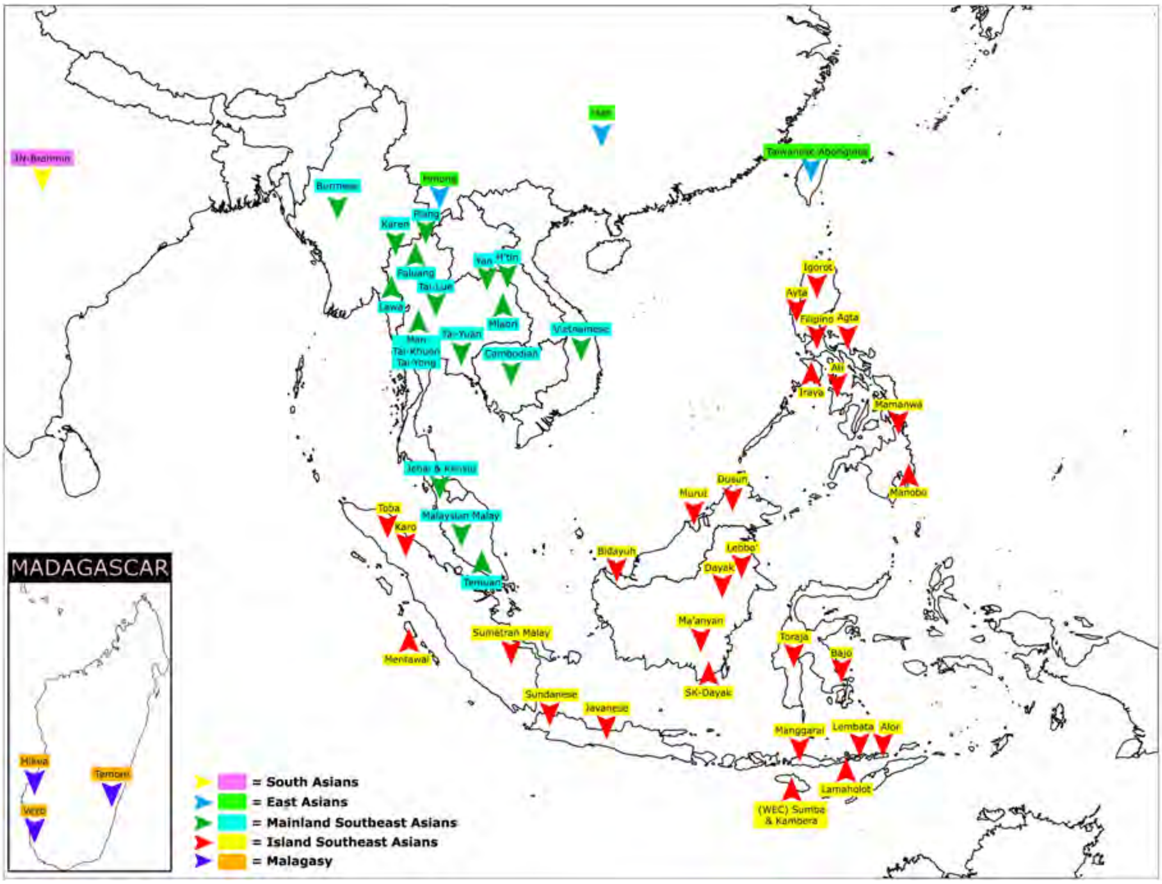


Figure 1. Map showing the location of each population group studied in this work. The map is generated using Global Mapper v.15 software (http://www.bluemarblegeo.com/ products/global-mapper.php)

3.2.1.3. Results

The unique Austronesian origin of the Ma'anyan

Following quality control, we obtained genotypes for 701,211 SNPs in a new set of 202 individuals from Borneo: 162 Ma'anyan and 40 South Kalimantan Dayak (SK-Dayak). To characterize the Ma'anyan and SK-Dayak gene pool within an Asian context, we focused our analyses on Island Southeast Asian, East Asian and Mainland Southeast Asian populations (Fig. 1). In a Principal Component Analysis (PCA) using a subset of the SNPs that intersect with published data from an extensive range of regional populations (the low density dataset) (Annex E: Figure S1), the first principal component (explaining 19.3% of the variance) separates Island Southeast Asian populations from East Asian and Mainland Southeast Asian groups, while the second principal component (explaining 17.5% of the variance) splits the Igorot on the positive axis and the Ma'anyan on the negative axis, with other Austronesian-speaking populations falling in between, such as Taiwanese aborigines, Filipinos, Borneo populations (Murut, Dusun, Lebbo' and South Kalimantan Dayak) and Sumatran populations (Sumatran Malay and Karo). Other Austronesian-speaking groups, like the Bidayuh, Javanese and Malaysians cluster towards mainland Southeast Asia, likely due to the historical influence of that region on these groups. Interestingly, the Ma'anyan form their own pole on the plot and do not cluster closely with other populations from Borneo, although the genetically closest population is still the South Kalimantan Dayak group, which is also geographically the nearest neighbour to the Ma'anyan. A similar population clustering pattern is observed with both the low- and high density SNP datasets (Annex E: Figure S2). This observation also agrees with F_{ST} values calculated on the low density dataset (Annex E: Table S2).

This unique genetic placement of the Ma'anyan is supported by admixture estimates, also performed on the low density dataset (Figure 2), especially at K=14 where it achieves its lowest cross-validation value (Annex E: Figure S3). The main ancestral components observed in Southeast Asian populations are: i) an Austronesian Igorot and indigenous Taiwanese component (C3; light green), ii) a Mainland Southeast Asian (MSEA) component (C11; light brown); and iii) a Papuan component (C2; light blue). However, our analysis reveals a major new component in Island Southeast Asia, representing 80% to 95% of the ancestry in Ma'anyan individuals (C8; dark blue). This Ma'anyan component is also found using an ADMIXTURE analysis on our high density SNP dataset (Annex E: Figure S4a and S4b). The remaining ancestry components in the Ma'anyan also occur in most of the other Indonesian populations, and may result from shared history and/or limited gene flow between the Ma'anyan and neighbouring populations. In return, the new C8 component identified in the Ma'anyan is also found at much lower frequencies in many other Indonesian groups, reaching its highest frequency in surrounding populations of Ma’anyan in Borneo (~40%), but also appearing in some mainland Southeast Asian populations.

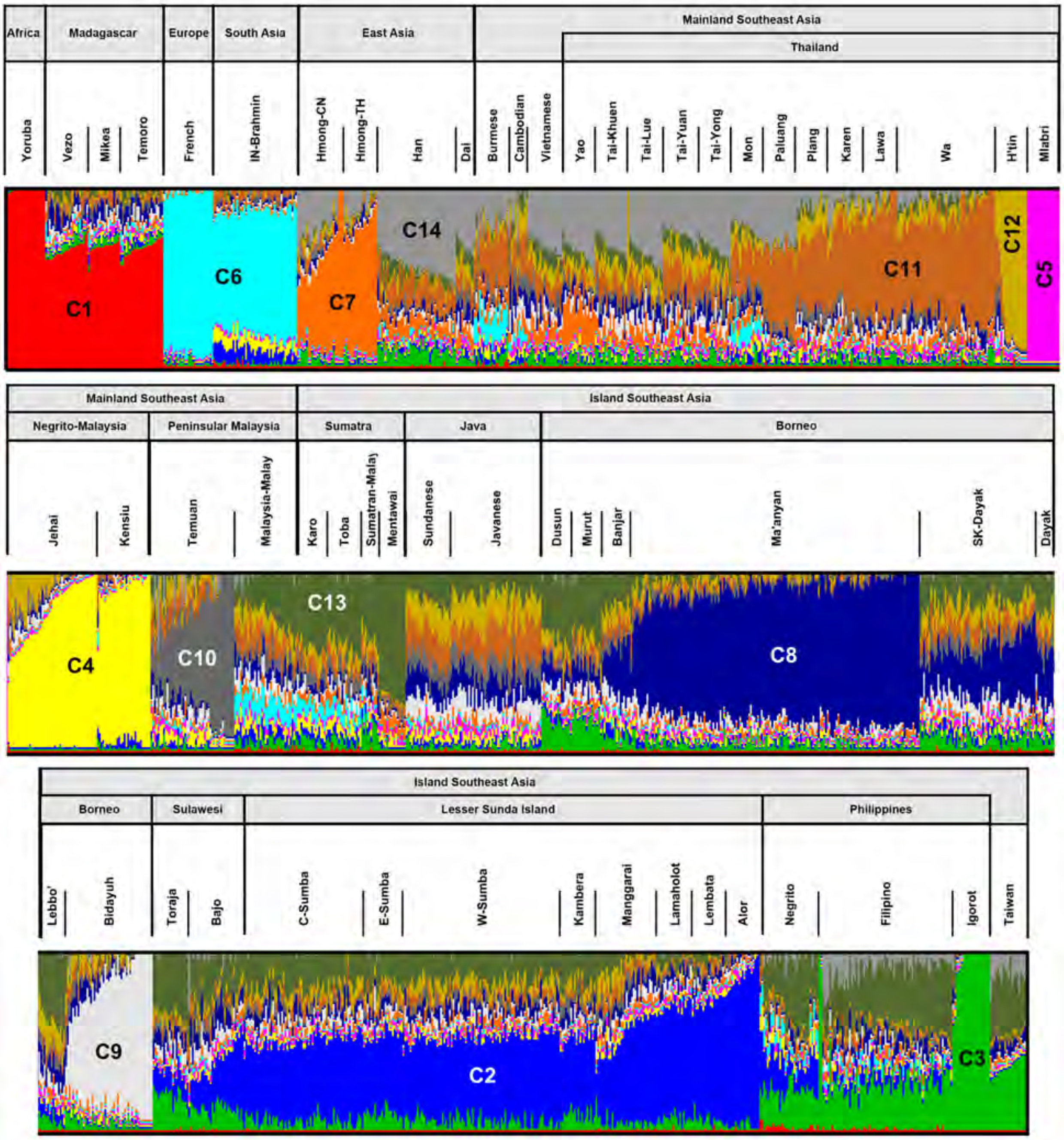


Figure 2. ADMIXTURE plot using the low density database with K = 14 (the optimum determined by cross-validation). Each component is identified by a specific color and a C label which corresponds to its order of appearance from K = 2 to K = 14.

To determine whether this distinct and homogenous genetic component in the Ma’anyan results from genetic drift (due to geographic isolation and/or endogamy), we inferred the extent of ‘Runs of Homozygosity’ (ROH) in the full high density dataset. Homozygosity in the Ma’anyan is similar to that of other Borneo populations (Annex E: Figure S5), even though these show much higher levels of admixture (Figure 2). However, homozygosity in the Ma’anyan is lower than in the Igorot, an isolated, indigenous Austronesian-speaking population living in the Philippine highlands. This suggests that the unusual homogeneity and unique ancestry component found in the Ma’anyan reflects the population’s migration history, rather than simply resulting from high levels of genetic drift. Genetic drift has also potentially occurred in the Igorot, and other isolated ethnic populations that exhibit low genetic diversity and have small population size (such as the Mlabri) (Oota et al., 2005; Xu et al., 2010), or that show a high level of consanguinity (such as the Malay Negritos) (Aghakhanian et al., 2015).

An f3-statistics analysis reveals more clearly that the Ma’anyan is not an admixed population (Annex E: Table S3). Defining the Ma’anyan as the daughter group, all possible combinations of populations in the low density dataset returned positive f3 statistics with Z-scores > -2, indicating no significant gene flow. In addition, a TreeMix analysis supported eight migration events, none of which involved gene flow to or from the Ma’anyan (Annex E: Figure S6). In contrast, a migration event was supported from the basal cluster of MSEA Austroasiatic-speaking H’tin and Mlabri to the Bidayuh, a population in northwest Borneo. This suggests that MSEA gene flows reached the west of Borneo, but not the east.

To test whether the Ma’anyan gene pool has drifted from its original Austronesian or MSEA ancestry, we performed an f3-outgroup statistics analysis (Figure 3a). All Island Southeast Asian populations, except the Bidayuh, Javanese and Sundanese, were pulled to the Austronesian side (as defined by the Taiwanese aborigines). Conversely, mainland Southeast Asian groups were pulled to the MSEA side (as defined by the H’tin). The Ma’anyan fall in the upper left diagonal of the plot, indicative of genetic similarity with Austronesian rather than MSEA groups. To determine the closest population to the Ma’anyan, the configuration f3(Yoruba; Ma’anyan, *x*) was explored, where *x* represents all populations in turn in the low density dataset. The highest value was obtained when *x* was the Igorot from the Philippines or non-Ma’anyan Borneo populations (Figure 3b), a result that is also obtained when using the high density dataset. These results place the genetic diversity of the Ma’anyan within the broader Austronesian gene pool.

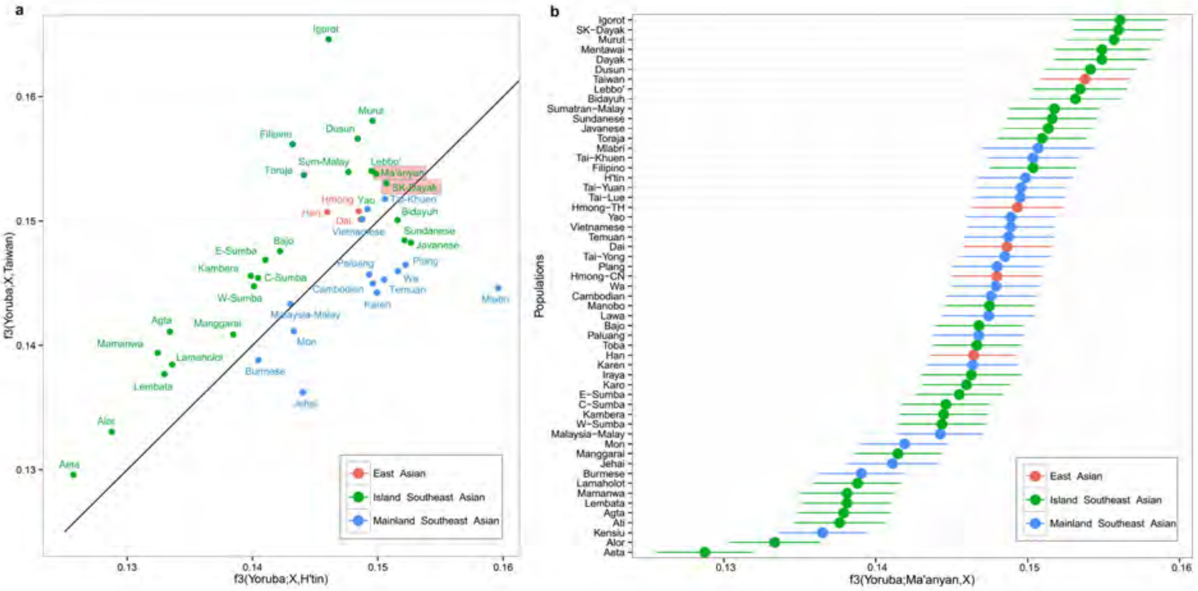


Figure 3. (a) An f3 outgroup statistics analysis showing shared genetic history with Austronesian groups (represented by indigenous Taiwanese) compared to Mainland Southeast Asian groups (represented by the H'tin). (b) Genetic similarity between Ma'anyan and other Asian populations measured using f3 outgroup statistics. Error bars show the standard error of the f3 statistics. Red dots represent East Asian groups; blue dots represent Island Southeast Asian groups; green dots represent Mainland Southeast Asian groups.

This Austronesian connection is also highly supported by an Identity-by-Descent (IBD) analysis performed with Refined IBD on the high density dataset. The Ma'anyan share more haplotypes with surrounding Borneo populations and the Igorot than with Mainland Southeast Asian groups (e.g. Cambodians, Burmese and Vietnamese) (Figure 4 and Annex E: Figure S7). When filtered for a total shared haplotype length greater than 20 cM (~20 Mb) between two individuals, links were still observed between the Ma'anyan and Mainland Southeast Asian groups, as well as other Indonesian populations. However, the links with Mainland Southeast Asian groups disappear with larger haplotype lengths, while connections with Austronesian groups (including the Igorot) are maintained up to a threshold of 40 cM, indicating more recent common ancestry (the hypothesis of recent gene flow can be discarded from earlier analyses). At higher thresholds (i.e. longer shared haplotypes), only connections within Borneo remain. Together, these analyses (ADMIXTURE, PCA, Runs of Homozygosity, f3 statistics, TreeMix and IBD) suggest that the unique Ma'anyan genetic component is an undetected part of the broader Austronesian genetic diversity. The Ma'anyan harbour a unique Austronesian genetic component, thus allowing us to raise the question: did the Ma'anyan gene pool contribute strongly to Malagasy, as suggested by linguistic evidence?

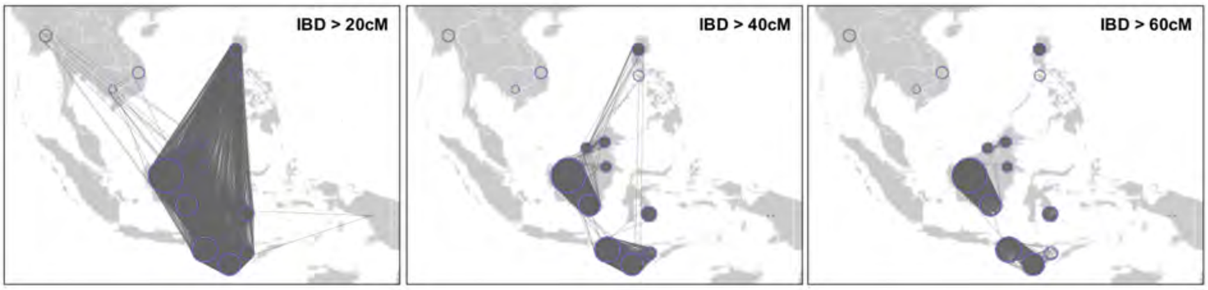


Figure 4. Shared Identity-By-Descent fragments between pairs of individuals in Southeast Asia, filtering for shared IBD >20 cM, 40 cM and 60 cM. Each individual is represented as a blue dot. Each individual is represented as a blue dot. Populations are represented by a circle of dots. Shared IBD fragments are represented by a black line. The maps were generated using Global Mapper v.15 software (http://www.bluemarblegeo.com/products/global-mapper.php). The networks lines were generated using Cytoscape v.3.2.152 software.

The Island Southeast Asian ancestries of the Malagasy

We performed PCA using the low density dataset, finding that the first two components described 54.6% of the observed variance (Annex E: Figure S8). The first component (PC1; explaining 39.1% of the variance) largely separated the continental groups of Africa, Europe, South Asia, and East and Southeast Asia. The second component (PC2; explaining 15.5% of the variance) differentiated the Malagasy, and separated the East and Southeast Asians into a broad north-to-south gradient. The Ma'anyan and South Kalimantan Dayak populations fall within the Asian cluster. The three previously published Malagasy groups (Temoro, Vezo and Mikea) are located at an intermediate position between the African and Asian clusters, reflecting their mixture of African and Asian ancestries (Pierron et al., 2014). Overall, Malagasy appear to contain more African ancestry than Asian.

Explicit admixture analysis on the low density dataset confirms this assessment, showing that the three Malagasy populations have ~70% African ancestry (red) versus ~30% Asian ancestry (mixed colours; Figure 2). These two main components appear consistently, in similar proportions, in plots from K = 2 to K =14 (Annex E: Figure S9). The Asian ancestry of Malagasy individuals is diverse, with no component (or set of components) pointing to a specific Asian population as the source of Malagasy. The Asian components found in Malagasy instead occur across Island Southeast Asia, including the South Kalimantan Dayak, Dusun, Murut, Javanese and the Ma'anyan. However, as described in the previous section, the Ma'anyan carry a particular component (C8) at very high frequency (50 to 95%), but this is much less frequent in other Western Indonesian populations (<50% in the South Kalimantan

Dayak) and in the Malagasy (2-15%), which instead exhibit a balanced range of other Asian components. The PCA and ADMIXTURE analyses confirm potential connections between Malagasy and western and central Indonesian populations (particularly Java, Borneo and Sulawesi), but do not pinpoint a primary source. These results are also consistent with the general nature of Island Southeast Asian gene flow into Malagasy, as determined by TreeMix (Annex E: Figure S10).

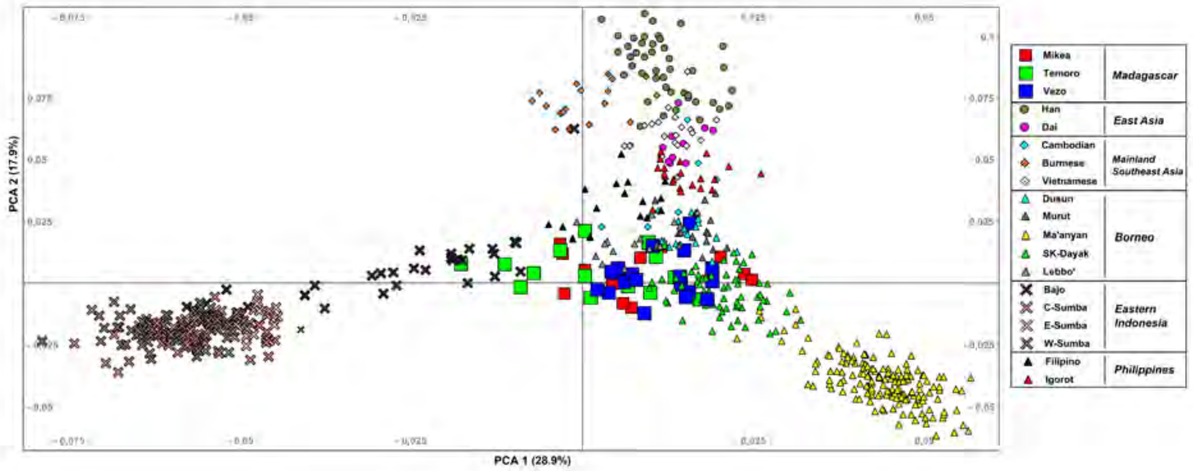


Figure 5. Ancestry-specific Principal Component Analysis based on masked SNPs from the high density dataset obtained after PCAdmix analysis.

Since the African ancestry in Malagasy may hinder the precise identification of Asian parental sources, we performed a PCAdmix analysis on the full high density dataset to mask African variants in the Malagasy data. An Asian ancestry-specific PCA, run on the filtered set of 17,043 SNPs, explained 46.8% of the observed variance in the dataset (Figure 5). The first principal component separated eastern Indonesians from western Indonesians and mainland East Asians. The second principal component separated the mainland Asian groups from those in Island Southeast Asia. As observed previously (Annex E: Figure S1), the Ma’anyan are positioned away from the other Island Southeast Asian groups and form their own pole on the graph. In this more refined analysis, the Asian markers found in the three Malagasy populations overlap closely with those from coastal Borneo (South Kalimantan Dayak, Murut and Dusun), although they do not obviously show any specific affinity with the Ma’anyan. Additionally, some Malagasy individuals are closely clustered with Bajo individuals, which may indicate that sea-nomads are relevant factors in the migrations to Madagascar, as suggested earlier (Kusuma et al., 2015). Despite this general link between Malagasy Asian ancestry and Borneo groups, an F_{ST} analysis using the same dataset highlights that the South Kalimantan Dayak still have the lowest genetic distance to the three Malagasy groups (average $F_{ST} = 0.022$) (Annex E: Table S4), thus suggesting that this is a likely Asian source population.

Together, these analyses confirm that Malagasy are a mixture of African and Island Southeast Asian populations, as suggested by much previous research (Capredon et al., 2013; Hurles et al., 2005; Pierron et al., 2014; Razafindrazaka et al., 2010; Soodyall et al., 1995; Tofanelli et al., 2009). However, this study provides the new information that the Island Southeast Asian populations with closest genetic affinity to the Malagasy are located along the coasts of Borneo, although exact source populations still cannot be clearly identified. Surprisingly, the Ma'anyan, despite speaking the closest sister language to Malagasy, do not share any particularly strong genetic links with Malagasy (Figure 2 and Figure 5). This lack of convergence between the genetic and linguistic evidence suggests that a more complex model is needed for the settlement of Madagascar. On the other hand, the uniqueness of the genetic diversity observed in the Ma'anyan opens an unexpected window for studying the complex history of the Austronesian expansion in Island Southeast Asia.

3.2.1.4. Discussion

A more complex picture of Austronesian genetic diversity

A genome-wide analysis of 211 individuals from Borneo reveals the unique genetic diversity of the Ma'anyan, opening an unexpected viewpoint into Southeast Asian prehistory. Our data reveal that the Ma'anyan are characterized by a specific genomic component that differentiates them from other Island Southeast Asian groups (Figure 2 and Annex E: Figure S1). This does not simply result from strong genetic drift (Annex E: Figure S5), but instead represents a homogenous genetic component that is largely uninfluenced by external gene flow. Although currently living in an isolated location, the Ma'anyan only settled there recently (see details below) (Hudson, 1967, 1972). This recent migration to isolated inland territories appears to have favoured the preservation of a unique genetic component, which is only rarely found in other Southeast Asian populations.

Recent studies have identified at least three broad genomic classes that dominate the gene pool of Southeast Asian individuals: Papuan ancestry, Mainland Southeast Asian ancestry, and Austronesian ancestry (Deng et al., 2015; Lipson et al., 2014). To relate these components to major episodes of human migration inferred from previous anthropological and archaeological studies, the Papuan ancestry likely tracks back to the initial settlement period (60-45 kya) (Karafet et al., 2010; Lipson et al., 2014; Tumonggor et al., 2013), the Mainland Southeast Asian ancestry probably to the very late Pleistocene (30-10 kya) (Blench, 2011; Deng et al., 2015; Jinam et al., 2012; Lipson et al., 2014), and the Austronesian ancestry to the mid-Holocene (5 kya) (Deng et al., 2015; Karafet et al., 2010; Lipson et al., 2014; Tumonggor et al., 2013). The discovery of a new ancestry component in the Ma'anyan is novel, although we show that it does occur at low levels in many populations across Island Southeast Asia (Figure 2). The presence of this component in these groups does not appear to be linked to any recent admixture events (Annex E: Figure S6 and Table S2), and therefore might instead be the signal of ancient shared ancestry. Nevertheless, this Ma'anyan component retains links to Austronesian diversity (Figure 3a and Figure 3b), with the Ma'anyan showing a particularly close genetic connection to the Igorot in the Philippines (Figure 3b and Figure 4). The Igorot, who also have strong Austronesian connections, live in remote areas of the Philippine highlands, which likely favoured the retention of their specific genetic signature. Shared connections between the Igorot and the Ma'anyan highlight a more complex picture of Austronesian genetic ancestry than has previously been presumed. We postulate that the ancestral diversity behind the Ma'anyan and Igorot genomic components emerged from some common unidentified source around East Asia or Taiwan, perhaps due to isolation-by-distance effects. The diffusion, and subsequent differentiation, of these two genetic components may find some support in the diffusion from Taiwan of two different cultural groups identified, respectively, by cord-marked and red-slipped pottery materials (Plutniak et al., 2014; Spriggs, 2007). However, the modality and timing of the spread of this ancestral Ma'anyan population and its relationship to the Austronesian expansion needs to be investigated further.

The Ma'anyan are not the primary biological ancestors of Malagasy

Despite strong linguistic affinities (Adelaar, 2005, 2009, Dahl, 1951, 1977), the Ma'anyan were not obviously the primary source population of the Malagasy. This confirms results obtained from uniparental markers, which show little sharing of genetic lineages between these two populations (Kusuma et al., 2015). As hinted previously (Pierron et al., 2014), the Asian ancestry of the Malagasy is instead diverse, and appears to relate to a range of Southeast Asian populations, albeit with especially close connections to groups in southern Borneo. It seems likely that the Asian individuals who settled Madagascar were already highly mixed, rather than coming from a wide range of Asian populations with later mixing in Madagascar, in agreement with the most likely scenario whereby only a small number of migrants were involved in the initial settlement of Madagascar (Cox et al., 2012). Looking across the Indonesian genetic landscape, the Ma'anyan carry a distinctive autosomal gene pool (dominated by the C8 component), which is not found in Malagasy (Figure 2 and Figure 5). This marked genomic difference between the Ma'anyan and the Asian component of Malagasy contradicts the hypothesis of a common origin inferred from the languages spoken by these two groups (Adelaar, 2005, 2009, Dahl, 1951, 1977). Hence, despite the strong affinity of Ma'anyan with the Malagasy language, the Ma'anyan people apparently did not contribute significantly to the Malagasy gene pool.

Other anthropological data may shed new light on the complex history of the Ma'anyan, perhaps reconciling this discrepancy between the linguistic and genetic data. Prior to their migration to Madagascar around 1,500 years ago, proto-Malagasy people had probably already developed a derived language that differed from Ma'anyan (Adelaar, 2017). This cultural process was likely driven by the growing influence of Malay and Javanese populations, which were trading intensively with groups in southeast Borneo (Adelaar, 2005, 2009). The only pre-colonial record from the region, the Hikayat Banjar (the 'Tale of Banjar') describes an old Malay settlement in southern Borneo, further inland than today's south Borneo coastline, that acted as a trading outpost of Malay Kingdom – such as the important Hindu kingdom of Srivijaya, which was dominant from the 7-13th centuries AD (Ras, 1968). This outpost was established because the coastline might have extended over 100 kilometres, and perhaps as much as 200 kilometres, further inland that at present (Adelaar, 2017; Van Bemmelen, 1949),and possibly lay near Tanjung-Amuntai region, where the Ma'anyan consider their original homeland (Hudson, 1967, 1972). It is conceivable that this settlement might then have provided sea contact to what are now land-bound Ma'anyan. As the coastline move southward, the trading post were also moved south and later formed the city of Banjarmasin: the dominant city, commercial state, and centre of activity in the trading network of this region. The inhabitants of Banjarmasin, the Banjar people, might then have constituted a mix of individuals from south Borneo under the cultural influence of the Malay Srivijaya kingdom. Based on this historical source, together with linguistic work on the ancestral states of the Malagasy language showing a substantial number of Malay loanwords (Adelaar, 2017, 2009), we postulate that the Asian source population of the Malagasy constituted admixed Ma'anyan individuals (best represented in our dataset by the South

Kalimantan Dayak), who lived in the Srivijaya area of influence, integrating Malay and Javanese cultural traits and favouring a large degree of gene flow, before migrating to Madagascar. Although the cause of their migration remains elusive, our data tend to favour an origin for the Malagasy in southern Borneo. Curiously, the group with the closest genetic affinity to Malagasy in our dataset is the South Kalimantan Dayak, a composite population of several ethnic groups located in southeast Borneo today (Annex E: Table S4). This suggests that an in-depth analysis of these ethnic groups, including the Banjar people and other southeast Borneo ethnic communities, might be a promising direction to better identify the (possibly mixed) genetic sources of the Malagasy and to determine the ultimate causes of the Malagasy expansion.

Our study shows that the Ma'anyan have genetic diversity that is unique in Southeast Asia, complicating existing scenarios of dispersal during the Austronesian expansion. Surprisingly, this component clearly shows that the Ma'anyan are not the primary source population of the Malagasy, as has long been supposed based on their common linguistic origin. The Asian parental population of the Malagasy instead appears to lie among the ethnic groups of the South East region of Borneo, potentially represented by the Banjar, or more generally, by the South Kalimantan Dayak people. This discrepancy between linguistic and genetic evidence may reflect the complex history of the south Borneo region, and more focused study of its peoples is needed to explore this hypothesis further.

3.2.1.5. Miscellaneous

Acknowledgements

We wish to acknowledge support from the GenoToul bioinformatics facility of the Genopole Toulouse Midi Pyrénées, France. This research was supported by the French ANR via grant ANR-14-CE31-0013-01 (OCEOADAPTO to F.-X.R.), grant ANR-12-PDOC-0037-01 (GENOMIX to D.P.), as well as the Region Aquitaine of France (MAGE to D.P.), the French Ministry of Foreign and European Affairs (French Archaeological Mission in Borneo (MAFBO) to F.-X.R.), a Rutherford Fellowship from the Royal Society of New Zealand (RDF-10-MAU-001 to M.P.C.) and the French Embassy in Indonesia through its Cultural and Cooperation Services (Institut Français en Indonésie).

Author contributions

All authors (P.K, N.B, M.P.C, F-X.R, H.S, H.R, T.L, D.P, and A.A) contributed to the design of the study. P.K and N.B performed the computational analyses. P.K, N.B, M.P.C, F-X.R wrote the manuscript based on the input from all authors (D.P, H.R, A.A, H.S, T.L).

3.2.1.6. References

Adelaar, K.A., 2017. Who were the first Malagasy, and what did they speak?, in: Acri, A., Landmann, A. (Eds.), Spirits and Ships: Cultural Transfer in Early Monsoon Asia. Institute of Southeast Asian Studies - Yusof Ishak Institute, Singapore, pp. 441–469.

Adelaar, K.A., 2009. Loanwords in Malagasy, in: Haspelmath, M., Tadmor, U. (Eds.), Loanwords in the World's Languages: A Comparative Handbook. De Gruyter Mouton, Berlin, Germany, pp. 717–746.

Adelaar, K.A., 2006. The Indonesian migrations to Madagascar: making sense of the multidisciplinary evidence, in: Austronesian Diaspora and the Ethnogenesis of People in Indonesian Archipelago: Proceedings of the International Symposium. Presented at the Austronesian diaspora and the ethnogenesis of people in Indonesian archipelago, LIPI Press, Jakarta, Indonesia, pp. 205–232.

Adelaar, K.A., 2005. The Austronesian languages of Asia and Madagascar: a historical perspective, in: The Austronesian languages of Asia and Madagascar. Routledge, New York, USA, pp. 1–42.

Adelaar, K.A., 1995. Borneo as a cross-roads for comparative Austronesian linguistics, in: The Austronesians: Historical and comparative perspectives. ANU E Press, Canberra, Australia, pp. 81–102.

Adelaar, K.A., 1989. Malay influence on Malagasy: linguistic and culture-historical implications. Ocean. Linguist. 28, 1–46. doi:10.2307/3622973

Aghakhanian, F., Yunus, Y., Naidu, R., Jinam, T., Manica, A., Hoh, B.P., Phipps, M.E., 2015. Unravelling the genetic history of Negritos and indigenous populations of Southeast Asia. Genome Biol. Evol. 7, 1206–1215. doi:10.1093/gbe/evv065

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664. doi:10.1101/gr.094052.109

Bellwood, P., 2007. Prehistory of the Indo-Malaysian archipelago. ANU E Press, Canberra.

Bellwood, P., Fox, J.J., Tryon, D., 1995. The Austronesians in history: Common origins and diverse transformations, in: The Austronesians: Historical and comparative perspectives. ANU E Press, Canberra, Australia, pp. 1–16.

Blench, R., 2011. Was there an Austroasiatic presence in Island Southeast Asia prior to the Austronesian expansion? Bull. Indo-Pac. Prehistory Assoc. 30, 133–144.

Blench, R.M., 2006. The Pleistocene settlement of the rim of the Indian Ocean. Presented at the 18th Congress of the Indo-Pacific Prehistory Association, Manila, Philippines.

Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., Reynolds, A., Ostrer, H., Mezey, J.G., Bustamante, C.D., 2012. PCAdmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. Hum. Biol. 84, 343–364. doi:10.3378/027.084.0401

Browning, S.R., Browning, B.L., 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81, 1084–1097. doi:10.1086/521987

Burney, D.A., 1987. Late Holocene vegetational change in central Madagascar. Quat. Res. 28, 130–143. doi:10.1016/0033-5894(87)90038-X

Burney, D.A., Burney, L.P., Godfrey, L.R., Jungers, W.L., Goodman, S.M., Wright, H.T., Jull, A.J.T., 2004. A chronology for late prehistoric Madagascar. J. Hum. Evol. 47, 25–63. doi:10.1016/j.jhevol.2004.05.005

Capredon, M., Brucato, N., Tonasso, L., Choesmel-Cadamuro, V., Ricaut, F.-X., Razafindrazaka, H., Rakotondrabe, A.B., Ratolojanahary, M.A., Randriamarolaza, L.-P., Champion, B., Dugoujon, J.-M., 2013. Tracing arab-islamic inheritance in Madagascar: Study of the Y-chromosome and mitochondrial DNA in the Antemoro. PLoS ONE 8, e80932. doi:10.1371/journal.pone.0080932

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7. doi:10.1186/s13742-015-0047-8

Cox, M.P., Hudjashov, G., Sim, A., Savina, O., Karafet, T.M., Sudoyo, H., Lansing, J.S., 2016. Small traditional human communities sustain genomic diversity over microgeographic scales despite linguistic isolation. Mol. Biol. Evol. 33, 2273–2284. doi:10.1093/molbev/msw099

Cox, M.P., Nelson, M.G., Tumonggor, M.K., Ricaut, F.-X., Sudoyo, H., 2012. A small cohort of Island Southeast Asian women founded Madagascar. Proc. R. Soc. B Biol. Sci. 279, 2761–2768. doi:10.1098/rspb.2012.0012

Dahl, O.C., 1991. Migration from Kalimantan to Madagascar. Norwegian University Press : Institute for Comparative Research in Human Culture, Oslo, Norway.

Dahl, O.C., 1977. La subdivision de la famille Barito et la place du malgache. Acta Orient. 38, 77–134.

Dahl, O.C., 1951. Malgache et maanjan: une comparaison linguistique. Edege-Intituttet, Oslo, Norway.

Deng, L., Hoh, B.-P., Lu, D., Saw, W.-Y., Twee-Hee Ong, R., Kasturiratne, A., Janaka de Silva, H., Zilfalil, B.A., Kato, N., Wickremasinghe, A.R., Teo, Y.-Y., Xu, S., 2015. Dissecting the genetic structure and admixture of four geographical Malay populations. Sci. Rep. 5, 14375. doi:10.1038/srep14375

Hudson, A.B., 1972. Padju Epat: The Maʼanyan of Indonesian Borneo. Irvington Publishers, New York, USA.

Hudson, A.B., 1967. The Padju Epat Ma'anjan Dajak in historical perspective. Indonesia 4, 8–42.

HUGO Pan-Asian SNP Consortium, Abdulla, M.A., Ahmed, I., Assawamakin, A., Bhak, J., Brahmachari, S.K., Calacal, G.C., Chaurasia, A., Chen, C.-H., Chen, J., Chen, Y.-T., Chu, J., Cutiongco-de la Paz, E.M.C., De Ungria, M.C.A., Delfin, F.C., Edo, J., Fuchareon, S., Ghang, H., Gojobori, T., Han, J., Ho, S.-F., Hoh, B.P., Huang, W., Inoko, H., Jha, P., Jinam, T.A., Jin, L., Jung, J., Kangwanpong, D., Kampuansai, J., Kennedy, G.C., Khurana, P., Kim, H.-L., Kim, K., Kim, S., Kim, W.-Y., Kimm, K., Kimura, R., Koike, T., Kulawonganunchai, S., Kumar, V., Lai, P.S., Lee, J.-Y., Lee, S., Liu, E.T., Majumder, P.P., Mandapati, K.K., Marzuki, S., Mitchell, W., Mukerji, M., Naritomi, K., Ngamphiw, C., Niikawa, N., Nishida, N., Oh, B., Oh, S., Ohashi, J., Oka, A., Ong, R., Padilla, C.D., Palittapongarnpim, P., Perdigon, H.B., Phipps, M.E., Png, E., Sakaki, Y., Salvador, J.M., Sandraling, Y., Scaria, V., Seielstad, M., Sidek, M.R., Sinha, A., Srikummool, M., Sudoyo, H., Sugano, S., Suryadi, H., Suzuki, Y., Tabbada, K.A., Tan, A., Tokunaga, K., Tongsima, S., Villamor, L.P., Wang, E., Wang, Y., Wang, H., Wu, J.-Y., Xiao, H., Xu, S., Yang, J.O., Shugart, Y.Y., Yoo, H.-S., Yuan, W., Zhao, G., Zilfalil, B.A., Indian Genome Variation Consortium, 2009. Mapping human genetic diversity in Asia. Science 326, 1541–1545. doi:10.1126/science.1177074

Hurles, M.E., Sykes, B.C., Jobling, M.A., Forster, P., 2005. The dual origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from maternal and paternal lineages. Am. J. Hum. Genet. 76, 894–901. doi:10.1086/430051

Jakobsson, M., Rosenberg, N.A., 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23, 1801–1806. doi:10.1093/bioinformatics/btm233

Jinam, T.A., Hong, L.-C., Phipps, M.E., Stoneking, M., Ameen, M., Edo, J., Consortium, H.P.-A.S., Saitou, N., 2012. Evolutionary history of continental Southeast Asians: "Early Train" hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. Mol. Biol. Evol. 29, 3513–3527. doi:10.1093/molbev/mss169

Karafet, T.M., Hallmark, B., Cox, M.P., Sudoyo, H., Downey, S., Lansing, J.S., Hammer, M.F., 2010. Major east–west division underlies Y chromosome stratification across Indonesia. Mol. Biol. Evol. 27, 1833–1844. doi:10.1093/molbev/msq063

Kusuma, P., Cox, M.P., Pierron, D., Razafindrazaka, H., Brucato, N., Tonasso, L., Suryadi, H.L., Letellier, T., Sudoyo, H., Ricaut, F.-X., 2015. Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations. BMC Genomics 16, 191. doi:10.1186/s12864-015-1394-7

Lewis, M.P., Simons, G.F., Fennig, C.D. (Eds.), 2016. Ethnologue: Languages of the world, Nineteenth edition. ed. SIL International, Texas.

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., Myers, R.M., 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319, 1100–1104. doi:10.1126/science.1153717

Lipson, M., Loh, P.-R., Patterson, N., Moorjani, P., Ko, Y.-C., Stoneking, M., Berger, B., Reich, D., 2014. Reconstructing Austronesian population history in Island Southeast Asia. Nat. Commun. 5. doi:10.1038/ncomms5689

MacPhee, R.D.E., Burney, D.A., 1991. Dating of modified femora of extinct dwarf Hippopotamus from Southern Madagascar: Implications for constraining human colonization and vertebrate extinction events. J. Archaeol. Sci. 18, 695–706. doi:10.1016/0305-4403(91)90030-S

Mörseburg, A., Pagani, L., Ricaut, F.-X., Yngvadottir, B., Harney, E., Castillo, C., Hoogervorst, T., Antao, T., Kusuma, P., Brucato, N., Cardona, A., Pierron, D., Letellier, T., Wee, J., Abdullah, S., Metspalu, M., Kivisild, T., 2016. Multi-layered population structure in Island Southeast Asians. Eur. J. Hum. Genet. 24, 1605–1611. doi:10.1038/ejhg.2016.60

Oota, H., Pakendorf, B., Weiss, G., Haeseler, A. von, Pookajorn, S., Settheetham-Ishida, W., Tiwawech, D., Ishida, T., Stoneking, M., 2005. Recent origin and cultural reversion of a Hunter–Gatherer group. PLOS Biol. 3, e71. doi:10.1371/journal.pbio.0030071

Oppenheimer, S., Richards, M., 2001. Fast trains, slow boats, and the ancestry of the Polynesian islanders. Sci. Prog. 84, 157–181.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012. Ancient Admixture in Human History. Genetics 192, 1065–1093. doi:10.1534/genetics.112.145037

Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. PLoS Genet 2, e190. doi:10.1371/journal.pgen.0020190

Pickrell, J.K., Pritchard, J.K., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8, e1002967. doi:10.1371/journal.pgen.1002967

Pierron, D., Razafindrazaka, H., Pagani, L., Ricaut, F.-X., Antao, T., Capredon, M., Sambo, C., Radimilahy, C., Rakotoarisoa, J.-A., Blench, R.M., Letellier, T., Kivisild, T., 2014. Genome-wide evidence of Austronesian–Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. Proc. Natl. Acad. Sci. 111, 936–941. doi:10.1073/pnas.1321860111

Plutniak, S., Oktaviana, A.A., Sugiyanto, B., Chazine, J.M., Ricaut, F.X., 2014. New ceramic data from East Kalimantan: Pottery chronology and the cord-marked and red-slipped sherds of Liang Abu’s layer 2. J. Pac. Archaeol. 5, 90–99.

Ras, J.J., 1968. Hikajat Banjar: a study in Malay historiography. Martinus Nijhoff, The Hague.

Razafindrazaka, H., Ricaut, F.-X., Cox, M.P., Mormina, M., Dugoujon, J.-M., Randriamarolaza, L.P., Guitard, E., Tonasso, L., Ludes, B., Crubezy, E., 2010. Complete mitochondrial DNA sequences provide new insights into the Polynesian motif and the peopling of Madagascar. Eur. J. Hum. Genet. 18, 575–581. doi:10.1038/ejhg.2009.222

Serva, M., Petroni, F., Volchenkov, D., Wichmann, S., 2012. Malagasy dialects and the peopling of Madagascar. J. R. Soc. Interface 9, 54–67. doi:10.1098/rsif.2011.0228

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. doi:10.1101/gr.1239303

Soares, P., Rito, T., Trejaut, J., Mormina, M., Hill, C., Tinkler-Hundal, E., Braid, M., Clarke, D.J., Loo, J.-H., Thomson, N., Denham, T., Donohue, M., Macaulay, V., Lin, M., Oppenheimer, S., Richards, M.B., 2011. Ancient voyaging and Polynesian origins. Am. J. Hum. Genet. 88, 239–247. doi:10.1016/j.ajhg.2011.01.009

Soodyall, H., Jenkins, T., Stoneking, M., 1995. “Polynesian” mtDNA in the Malagasy. Nat. Genet. 10, 377–378. doi:10.1038/ng0895-377

Spriggs, M., 2007. The Neolithic and Austronesian expansion within Island Southeast Asia and into the Pacific, in: Chiu, S., Sand, C. (Eds.), From Southeast Asia to the Pacific: Archaeological perspectives on the Austronesian expansion and the Lapita cultural complex. Academia Sinica, Taiwan, pp. 104–125.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30, 2725–2729. doi:10.1093/molbev/mst197

Tofanelli, S., Bertoncini, S., Castrì, L., Luiselli, D., Calafell, F., Donati, G., Paoli, G., 2009. On the origins and admixture of Malagasy: New evidence from high-resolution analyses of paternal and maternal lineages. Mol. Biol. Evol. 26, 2109–2124. doi:10.1093/molbev/msp120

Tumonggor, M.K., Karafet, T.M., Hallmark, B., Lansing, J.S., Sudoyo, H., Hammer, M.F., Cox, M.P., 2013. The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific. J. Hum. Genet. 58, 165–173. doi:10.1038/jhg.2012.154

Van Bemmelen, R.W., 1949. The Geology of Indonesia. Government Printing Office; Martinus Nijhoff, The Hague.

Xu, S., Kangwanpong, D., Seielstad, M., Srikummool, M., Kampuansai, J., Jin, L., 2010. Genetic evidence supports linguistic affinity of Mlabri - a hunter-gatherer group in Thailand. BMC Genet. 11, 18. doi:10.1186/1471-2156-11-18

3.2.2. The last sea nomads of the Indonesian archipelago: Genomic origins and dispersal

(Submitted to European Journal of Human Genetics, accepted on 13th April 2017)

Pradiptajati Kusuma[1,2], Nicolas Brucato[1], Murray P Cox[3], Thierry Letellier[1], Abdul Manan[4], Chandra Nuraini[5], Philippe Grangé[5], Herawati Sudoyo[2,6], François-Xavier Ricaut[1*]

Affiliations:
[1]Equipe de Médecine Evolutive, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse UMR-5288, Université de Toulouse, Toulouse, France
[2]Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular Biology, Jakarta, Indonesia
[3]Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
[4]University of Halu Oleo, Kendari, Indonesia
[5]UFR des Lettres, Langues, Arts et Sciences Humaines, Université de La Rochelle, La Rochelle, France
[6]Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia

3.2.2.1. Abstract

The Bajo, the world's largest remaining sea nomad group, are scattered across hundreds of recently settled communities in Island Southeast Asia, along the coasts of Indonesia, Malaysia and the Philippines. With a significant role in historical trading, the Bajo lived until recently as nomads, spending their entire lives on houseboats while moving long distances to fish and trade. Along the routes they traveled, the Bajo settled and intermarried with local land-based groups, leading to 'maritime creolization', a process whereby Bajo communities retained their culture, but assimilated – and frequently married into – local groups. The origins of the Bajo have remained unclear despite several hypotheses from oral tradition, culture and language, all currently without supporting genetic evidence. Here, we report genome-wide SNP analyses on 73 Bajo individuals from three communities across Indonesia – the Derawan of Northeast Borneo, the Kotabaru of Southeast Borneo, and the Kendari of Southeast Sulawesi, with 87 new samples from three populations surrounding the area where these Bajo peoples live. The Bajo likely share a common connection with Southern Sulawesi, but crucially, each Bajo community also exhibits unique genetic contributions from neighboring populations.

Key words: Indonesia; Bajo; Sama-Bajau; sea nomads; genome-wide SNP array

3.2.2.2. Introduction

Rapid advances in sea faring technologies in Island Southeast Asia (ISEA) around 5,000 years ago created an intricate network of maritime interactions, the leading example being the well-known expansion of Austronesian peoples (Bellwood, 2007; Bulbeck, 2008; Solheim et al., 2006). Triggering inter-continental maritime connections linking ISEA with East Africa and Remote Oceania (Beaujard, 2012a, 2012b; Bellwood, 2007; Kayser et al., 2008; Oppenheimer and Richards, 2001; Skoglund et al., 2016), these contacts drove exchanges of goods, ideas, cultures and people around the Indo-Pacific region (Chaudhuri, 1985; Lawler, 2014). Sea-orientated populations, including sea nomads, emerged from this milieu, dominating trade within ISEA for centuries and helping to structure population interactions across Indonesia and beyond. Today, the Indonesian archipelago hosts ~600 ethnic groups (Lewis et al., 2016), of which only a handful are known for their sea-based lifestyles. Some, like the Bugis and Makassar of Southern Sulawesi (Pelras, 1997), are maritime inter-regional traders that arose within the framework of regional empires, such as Malay/Hindu Śrīvijaya and Majapahit. However, these groups still have homeland territories on land. Far more extreme are ethnic groups that subsist entirely detached from the land, living their whole lives aboard small boats, and as recently as 40 years ago, living as nomadic seafarers (Hogan and Pattemore, 1988; Larish, 2007; Sopher, 1977).

The biggest group, the Bajo (also Bajaw, Bajau or Sama-Bajau (Sather, 1997; Stacey, 2007)), number approximately one million people, who today live in numerous scattered hamlets and villages recently created by the Indonesian government along the coasts of the Indonesian archipelago, as well as Sabah in Malaysia, and the Sulu archipelago and South-Western Mindanao in the Southern Philippines (Nuraini, 2016; Pallesen, 1985; Sopher, 1977; Verheijen, 1986). The geographical distribution of Bajo communities overlaps large parts of the coral triangle, which contains one of the highest rates of marine biodiversity in the world, thus underpinning the Bajo economy based on exploiting marine resources including fish, tortoise shell and sea cucumber. Within Indonesia, the Bajo presence extends over a wide geographical area (Figure 1). Historically, Bajo were frequently associated with Sulawesi Bugis traders and ship owners (Pelras, 1997; Stacey, 2007) and were well known for traveling with their families, even for long distance journeys reaching as far as New Guinea and Australia (Macknight, 1976). The Bajo may have mediated westward dispersals into the Indian Ocean, perhaps even playing a role in the Indonesian settlement of Madagascar (Kusuma et al., 2015).

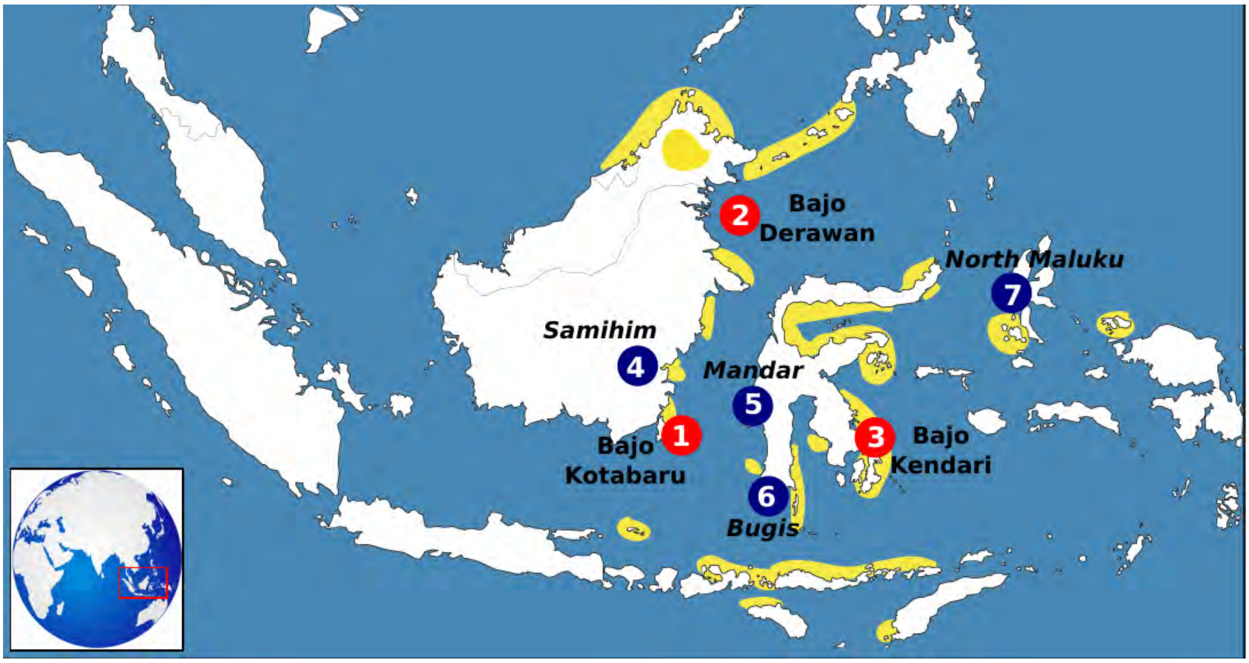


Figure 1. Map showing the distribution of Bajo communities across Island Southeast Asia (yellow), together with the location of sampled Bajo villages (red dots) and sampled historically related communities (blue dots).

Although some Bajo communities live far apart today, they still have similar social and cultural features, including shared shipbuilding and fishing culture, traditions and myths (Nagatsu, 2013; Nuraini, 2008; Stacey, 2007). Their languages belong to a single sub-family, the Sama-Bajau sub-group on the West Malayo-Polynesian branch of the Austronesian language family (Blust, 2007; Noorduyn, 1991). This sub-group includes at least nine languages (Lewis et al., 2016; Pallesen, 1985), with its highest diversity in Sabah (North Borneo) and the Southern Philippines (Pallesen, 1985). However, the Sama-Bajau languages of Indonesia are poorly documented, and an ongoing survey has identified at least three unrecognized languages (Grangé, forthcoming). Some of these languages are mutually unintelligible, suggesting that the Bajo diaspora started centuries ago, fitting with oral tradition. Numerous loanwords indicate that the languages spoken by the Bajo were influenced by neighboring ethnic groups with whom the Bajo interacted and socialized, in a process called 'maritime creolization' (Nagatsu, 2013). However, the extent of these social interactions on the genetic composition of Bajo communities remains unknown.

The Bajo have no written history, instead relying on oral tradition, especially epic songs, which say little about their early history. Hypotheses about their origins have been drawn from this folklore, as well as linguistic studies and rare records from European sailors from the 16th century onward (Sather, 1997). The Bajo diaspora may have originated in Johor, Malaysia (Sopher, 1977), or even Arabia (Nuraini, 2008), according to oral tradition. Brunei and Southern Sulawesi have also been proposed based on other Bajo stories (Jubilado et al., 2011; Liebner, 1998; Nuraini, 2008). Linguistic surveys point towards the Sulu archipelago of the Philippines,18 and at an earlier stage, to Southeast Borneo (Blust, 2007). None of these hypotheses have been tested with genetic data.

Here, we undertake a genomic survey to help clarify the history of Bajo sea nomad populations. We present genome wide analyses from three Bajo communities (n = 73; Annex F: Table S1), representing different Sama-Bajaw dialects, together with comparative data from neighboring populations potentially connected historically with the Bajo. Using this large comparative dataset, we investigate the genetic origins and history of the Bajo, and characterize the genetic impact of their near-unique lifestyle as some of the world's last remaining sea nomads.

3.2.2.3. Materials and methods

Ethics

Biological sampling was conducted by the Eijkman Institute for Molecular Biology, with the assistance of Indonesian Public Health clinic staff, following protocols for the protection of human subjects established by the Eijkman Institute. All samples were collected with informed consent from unrelated individuals. Collection and use of these samples was approved by the Research Ethics Commission at the Eijkman Institute for Molecular Biology, Indonesia.

Samples, dataset integration and quality control

Subjects were surveyed for language affiliation, current residence, familial birthplaces, and a short genealogy of four generations to establish regional ancestry. A total of 47 saliva samples were collected using the Oragene saliva sampling kit (DNA Genotek Inc., ON, Canada) from two Bajo communities: Derawan (n = 18) in coastal Northeastern Borneo, Indonesia and Kotabaru (n = 23) in coastal Southeastern Borneo, Indonesia (Figure 1). DNA was extracted using the standard kit protocol. We also added DNA samples from the Samihim in Eastern Borneo (n = 25), the Bugis of Southern Sulawesi (n = 25), the Mandar of Southern Sulawesi (n = 23) and North Maluku individuals from various linguistic groups (n = 14) as comparative populations. This sampling strategy is relevant for the statistical tests that are described below, both on population structure and admixture. Genome wide SNP genotypes were generated using the Illumina Human Omni5 Bead Chip (Illumina Inc., CA, USA), which surveys 4,284,426 single nucleotide markers semi-regularly spaced across the genome. Genotype data from previously published Bajo individuals from the Kendari community of Southern Sulawesi were also included (n = 32) (Pierron et al., 2014). New genotyping data have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and CRG, under accession number EGAS00001002246.

A comparative dataset was built from 110 worldwide populations comprising an additional 2,256 individuals (Annex F: Table S1). Data quality controls were performed using PLINK v1.9 (Chang et al., 2015): i) to avoid close relatives, relatedness was measured between all pairs of individuals within each population using an Identity-by-Descent (IBD) estimation with upper threshold of 0.25 (second degree relatives); ii) SNPs that failed the Hardy-Weinberg exact (HWE) test ($P<10^{-6}$) were excluded; iii) samples with an overall call rate <0.99 and individual SNPs with missing rates >0.05 across all samples in each population were excluded. The final dataset contains 230,833 SNPs. Genotypes were then phased with SHAPEIT v2 (Delaneau et al., 2012) using the 1000 Genomes Project phased data (Delaneau and Marchini, 2014) as a reference panel and the HapMap phase II genetic map. For specific analyses mentioned below, variants in high linkage disequilibrium (LD) ($r^2> 0.5$; 50 SNP sliding windows) were also pruned, leaving a final dataset of 168,368 SNPs.

Population structure

Population structure was evaluated using a suite of different programs, each relying on specific algorithms and types of data, to obtain the most relevant and robust interpretations. A fineSTRUCTURE v2.07 (Lawson et al., 2012) analysis was performed using $2x10^6$ Markov Chain Monte Carlo (MCMC) iterations, discarding the first 10^6 iterations as 'burn in', and sampling from the posterior distribution every 10^4 iterations following the burn in. This analysis detects shared IBD fragments between each pair of individuals, without self-copying, as calculated with Chromopainter v2.0 (Lawson et al., 2012) to perform a model-based Bayesian clustering of genotypes. From the results, a co-ancestry heat map and dendrogram were built to visualize the number of statistically defined clusters that best describe the data. Principal Components Analysis was performed using the 'smartpca' algorithm of EIGENSOFT v6.0.1 (Patterson et al., 2006). The Runs of Homozygosity (ROH) and Inbreeding Coefficient (F_{IS}) analyses were performed in PLINK v1.9. F_{ST} distance calculations were calculated with EIGENSOFT v6.0.1. To ascertain the significance of each pairwise F_{ST} value, 10,000 bootstraps were conducted using StAMPP (Pembleton et al., 2013), from which probability values were determined.

Population admixture

Admixture scenarios are determined from statistically complex models that rely a priori on the algorithms, and their assumptions, implemented in each program. To compensate for the potential biases of individual methods, we based our interpretations on the convergence of results from multiple different programs and different types of data. ADMIXTURE v1.30 (Shringarpure et al., 2016) was used to estimate the genomic ancestry profile of individuals using maximum likelihood for components (K) from K = 2 to K = 20. Ten replicates were run at each value of K with different random seeds, then merged and assessed for clustering quality using CLUMPP (Jakobsson and Rosenberg, 2007), and the cross-validation value was calculated to determine the optimal number of genomic components. To determine the sex bias of admixture for all Bajo communities, unsupervised ADMIXTURE analysis were run on K = 2 using both autosomal and X-chromosome SNPs using Igorot and PNG highlanders as proxies for East Asian and Papuan ancestry, respectively. Significance tests of the proportion of the Papuan component between the autosomes and X chromosome for all Bajo communities were conducted using the one-tailed Wilcoxon test. Gene flow between populations was first investigated using TreeMix v1.12 (Pickrell and Pritchard, 2012), with blocks of 200 SNPs to account for linkage disequilibrium, and migration edges added sequentially until the model explained 99% of the variance. The three-population (f3) test was performed as implemented in ADMIXTOOLS v1.3 (Shringarpure et al., 2016). Haplotype sharing using the Refined IBD algorithm of Beagle v.4.0 (Browning and Browning, 2007) was computed to estimate the total number of shared genetic fragments (logarithm of odds ratio > 3) between each pair of individuals. Finally, we used Chromopainter v2 (Lawson et al., 2012) and GLOBETROTTER v1 (Hellenthal et al., 2014) to estimate the ratios and dates of potential admixture events. For all results presented here, we standardized each co-ancestry curve by a 'NULL' individual designed to eliminate any spurious linkage disequilibrium patterns not attributable to that expected under a genuine admixture event (Hellenthal et al., 2014), and consistency between each estimated parameter was checked, though we note that results were similar when not performing this standardization. The 'best-guess' scenario given by GLOBETROTTER was considered for each target population. Using the parental populations given by GLOBETROTTER, we ran 100 bootstrap iterations to estimate admixture dates, assuming a generation interval of 28 years for all analyses (Fenner, 2005).

With the parental populations given by GLOBETROTTER, dates of admixture were also estimated using MALDER v1.3 (Loh et al., 2013; Pickrell et al., 2014).

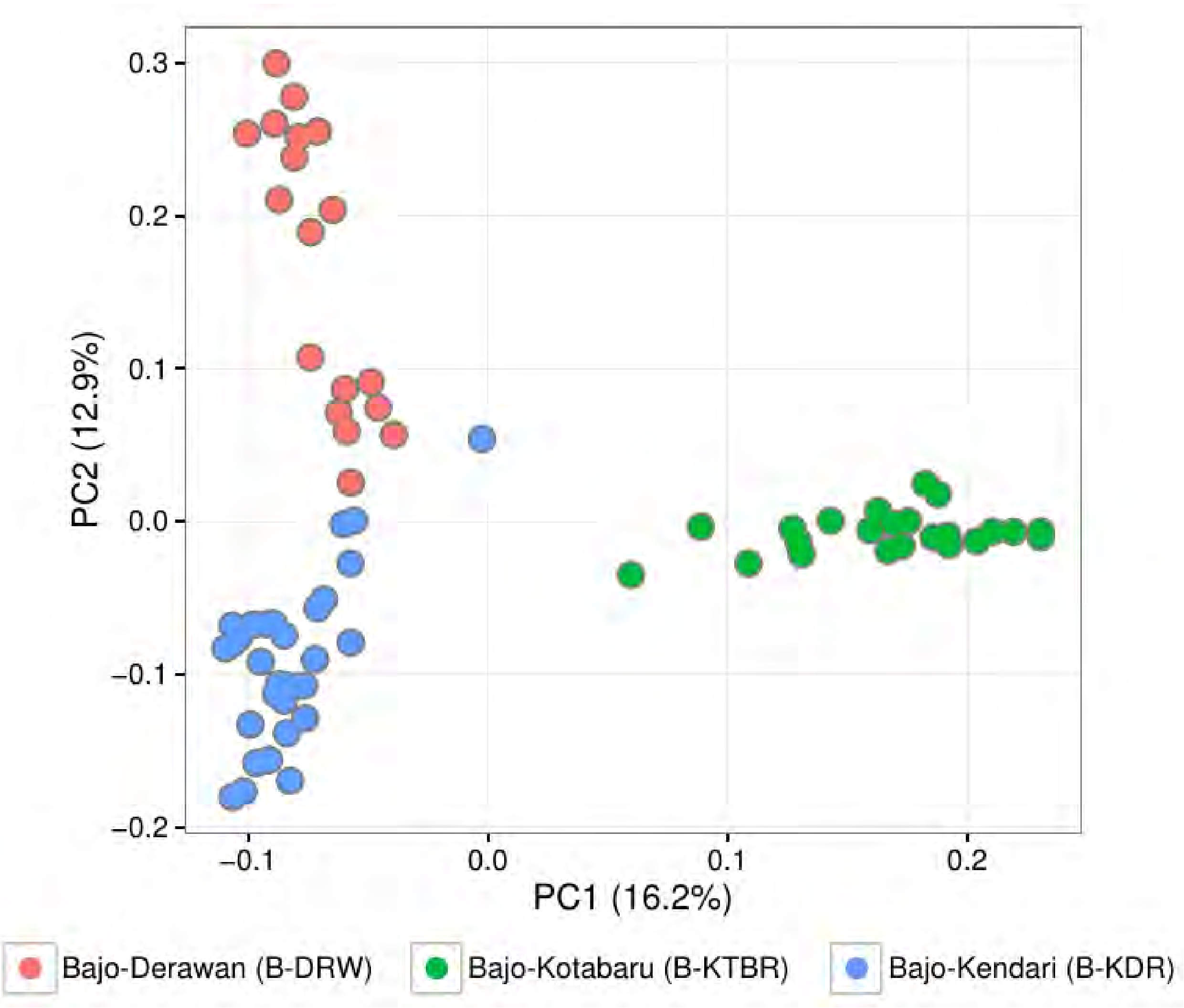


Figure 2. Principal Component Analyses of the three Bajo communities (Kendari, blue; Kotabaru, green; Derawan, red) based on 645,385 SNPs, showing independent clustering and limited overlap between individuals from different Bajo communities.

3.2.2.4. Results

We studied genetic variation in three Bajo communities spread across large parts of their geographical range: the Derawan of coastal Northeastern Borneo (n = 18, B-DRW), the Kotabaru of coastal Southeastern Borneo (n = 23, B-KTBR), and the Kendari of coastal Sulawesi (n = 32, B-KDR). To determine the population structure of these three Bajo communities, a principal component analysis (PCA) was performed using 645,385 overlapping SNPs in just the Bajo (Figure 2). Individuals from the three groups form distinguishable clusters. PC1 (16.2% variance explained) separates the Kotabaru Bajo from the two other groups, while PC2 (12.9% variance explained) differentiates the Kendari Bajo from the Derawan Bajo. Interestingly, there is no overlap between the groups presently living in Borneo.

The regional connections of the Bajo were determined from 230,833 overlapping SNPs in 116 surrounding populations. A clear division appears between East Asia/Mainland Southeast Asia and Island Southeast Asia, notably separating Papuan/Eastern Indonesian populations (PC1) from Western Indonesian populations (PC 2). All Bajo individuals fall within the Island Southeast Asia cluster, specifically with other Indonesian groups (Annex F: Figure S1). As before, all three Bajo communities still form their own clusters with limited overlap. Most Bajo individuals lie close to populations from Sulawesi, such as the Bugis and Mandar. The Derawan Bajo cluster close to Philippine populations; the Kotabaru Bajo cluster close to Borneo populations; while the Kendari Bajo have connections with eastern Indonesia, such as Sumba and North Maluku, and with Papuans.

The PCA results are consistent with fineSTRUCTURE clustering on phased genotype data (Annex F: Figure S2), which shows that all three Bajo communities form a single group, but trend towards their close geographic neighbors. Conversely, pairwise F_{ST} values ($P < 1 \times 10^{-4}$ for all F_{ST} pairs) suggest that all three Bajo communities have closer genetic ties to their surrounding populations than between themselves (Annex F: Table S2), thus hinting that genetic connections within the Bajo are correspondingly weaker. For instance, the Kendari Bajo have closest genetic distances with Sulawesi Bugis and Mandar; the Kotabaru Bajo with Borneo Banjar and Malay; and the Derawan Bajo with Philippine populations and Borneo Lebbo. Geography, and interactions with local groups, are therefore dominant features in the development of Bajo genetic diversity.

However, all Bajo individuals do share common patterns of genetic ancestry, as revealed by ADMIXTURE analysis (Figure 3, Annex F: Figure S3, Figure S4). The three Bajo communities have an admixed profile with two major Asian components and a Papuan component, but in varying proportions. The Kendari Bajo have more of the Papuan component (red) than the two Borneo Bajo groups (~20%), in keeping with their location further east. The Asian genetic ancestry is formed by similar components as for other Indonesian groups, with three main contributions: one East Asian (orange), two Austronesian components (pink and yellow), and an indigenous peninsular Malaysia component (cyan), cumulatively summing to 80-90%. The three Bajo communities only differ by relatively minor proportions of genomic ancestry that can be linked to their specific locations: minor Negrito Philippine (Aeta and Batak) components, with white and grey colors respectively, are observed in the Derawan Bajo (~1-2%); and an Indian component (green) is detected in both the Derawan Bajo and Kotabaru Bajo (~6%). Interestingly, this Indian component was not clearly detected in the Kendari Bajo, contra Mörseburg et al. (2016), probably because of its very low proportion. We detect sex biased admixture in Kotabaru Bajo and Derawan Bajo (one-tailed Wilcoxon test; $P < 0.01$), but not in Kendari Bajo nor in the Bugis (one-tailed Wilcoxon test; $P > 0.05$). A higher proportion of Papuan X chromosomes relative to the autosomal contribution is also observed (Annex F: Figure S5).

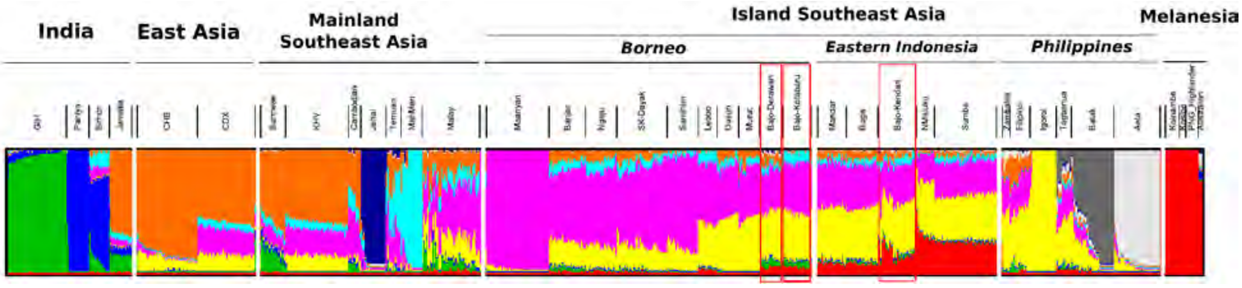


Figure 3. ADMIXTURE plot at K=10 depicting admixture of ancestral components in Derawan, Kotabaru and Kendari Bajo (red boxes), composed of East Asian, Austronesian, Papuan and minor Indian components.

The f3 statistics suggest that the Derawan and Kendari Bajo are admixed (Annex F: Table S3). However, defining the Kotabaru Bajo as a daughter population, all possible surrogate population combinations return positive f3 statistics with Z-scores > -2, indicating no significant gene flow, or recent bottlenecks, or founder effects (Patterson et al., 2012), as also suggested by the Admixture plot at $K = 20$ (Annex F: Figure S3). This is consistent with the Runs of Homozygosity (ROH) and Inbreeding Coefficient (F_{IS}) analyses, which show higher values compared to the other two Bajo groups (Annex F: Figure S6, Figure S7).

Identity-by-Descent (IBD) was used to measure haplotype sharing across the genome. All Bajo communities share longer fragments with each other than with other regional populations (Annex F: Figure S8), suggesting that the Bajo communities did intermarry until their recent land-based resettlement. The highest IBD sharing was observed between the Kendari and Kotabaru Bajo, then between the Kotabaru and Derawan Bajo, with much less between the Derawan and Kendari Bajo, again suggesting that genetic similar does not simply match current geographical location. As also shown by F_{ST} distances, high shared IBD between Bajo groups does not exclude sharing with non-Bajo neighbors. Nevertheless, IBD sharing between the Kendari Bajo and Bugis, two sea based communities currently settled in Sulawesi, is lower than IBD sharing between the Kendari Bajo and other Bajo groups.

Like other analyses, a TreeMix analysis situates the three Bajo communities with eastern Indonesian and Philippine populations (Annex F: Figure S9). The tree supports 16 migration nodes, many showing migration into the Bajo from Papuan clusters. Interestingly, there is Papuan migration into the two Bajo groups on Borneo, as well as the Kendari Bajo, where Papuan contributions were noted by Admixture. The most parsimonious hypothesis is multilayer admixture – from Papuan or Eastern Indonesian groups into the Kendari Bajo, and from there into the other Bajo groups. However, the data cannot exclude a more complex scenario with direct contact between Bajo groups in Borneo and Papuans.

We also inferred admixture scenarios for the three Bajo populations using GLOBETROTTER. This suggests that the Kendari Bajo mixed with surrogates of Sulawesi Bugis and Papuans multiple times. The oldest admixture event occurred around 62 generations ago (1,736 years ago, assuming a 28 year generation interval) with 90% and 10% contributions from Sulawesi Bugis and Papuans, respectively, and more recent admixture 6 generations ago (175 years ago), with admixture just from the Bugis (Figure 4, Annex F: Table S4). In contrast, the Kotabaru Bajo show one admixture event between Indian (5%), Sulawesi Bugis (70%) and Bornean Banjar (25%) sources around 33 generations ago (925 years ago), suggesting that Sulawesi played a major role in shaping the genomes of Kotabaru Bajo individuals. Local populations also contribute to the genomic make up, highlighting the neighboring Banjar of Borneo as another contributing group. The Derawan Bajo have genomic components from Indian (5%), Filipino (70%), and Malay (25%) sources, dating to around 24 generations ago (675 years ago).

These results were obtained with significant fit values by excluding other Bajo communities as potential surrogate populations for any given Bajo community. When we allowed all Bajo groups to act as potential surrogate populations, lower fit values were obtained, reflecting uncertain inference of admixture scenarios (Annex F: Table S4). Nonetheless, these runs confirm the earlier GLOBETROTTER results for the Kendari and Kotabaru Bajo, but in a new finding, the Bugis appears to be a surrogate population for the Derawan Bajo, in addition to Malay, Filipinos and Indians. Given the potential role of the Bugis on the genetic make-up of the Bajo, we therefore tested their admixture profile using GLOBETROTTER. The Bugis experienced a multiway admixture event around 57 generations ago (1,600 years ago) between Papuans (14%), Filipinos (41%) and Malay (45%), at around the same time as the admixture event with the Kendari Bajo. These admixture events were confirmed using MALDER (Annex F: Table S5).

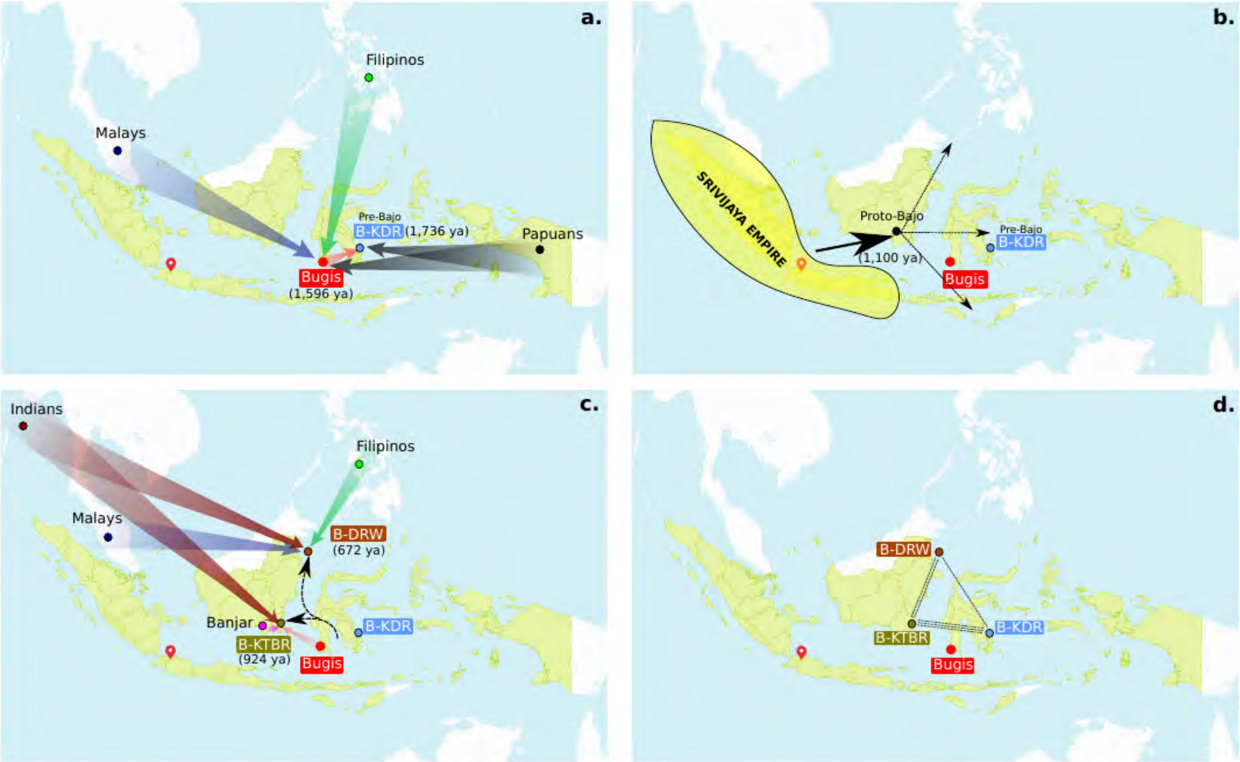


Figure 4. Admixture history of the three Bajo communities inferred with GLOBETROTTER. (a) Admixture of Bugis (South Sulawesi) with multiple populations, including Malay, Filipinos and Papuans, up to 1,600 years ago (ya), followed by subsequent admixture of pre-Bajo Kendari by Bugis and Papuans around 1,750 ya. (b) The expansion of the Śrīvijaya empire to Southeast Borneo triggered the dispersal of Bajo language, culture and people in many directions, including South Sulawesi and the Kendari, who assimilated them into its society. (c) South Sulawesi populations subsequently migrated westward to Southeast Borneo, forming the Kotabaru community by admixing with local Banjar populations, in addition to Indian influences through the reigning Malay empire around 925 ya. Northward migrations formed the Derawan community, which also admixed with local Malay and Filipino groups around 675 ya. The influence of South Sulawesi populations (dashed arrows) is observed in both the Kotabaru and Derawan Bajo. (d) Recent interactions between the three Bajo communities were maintained with different intensities (dashed lines).

3.2.2.5. Discussion

Even among the extraordinary diversity of human lifeways, the entirely sea-based lives of the Bajo – being born, growing up, marrying and dying on the sea – is special. This way of living is unique to Southeast Asia, with the Bajo, Urak Lawoi and Moken being well known examples (Ivanoff, 2002; Sopher, 1977). However, very little is understood about the genetic structure of these communities. Using genome-wide SNP data, we can reconstruct the genetic background and diversity of the Bajo across three communities with different dialects spanning their geographic range, thus helping to clarify where the Bajo originated and how their society interacted with other groups. Each Bajo community constitutes a homogenous genetic group, with surprisingly little overlap. A common theme is that genetic sharing is greater with neighboring populations than other Bajo groups, although there is a clear shared component of Bajo ancestry. Nevertheless, genetic contributions from these local populations were far from trivial, matching the maritime creolization process observed in their languages.

This admixture seems to have started early. Bajo were never the major ethnic group in the regions where they first lived, but instead seem to have attracted and assimilated people from nearby communities (Nagatsu, 2013). For example, in Kangean, a small archipelago in the Java Sea between Eastern Java and Southern Borneo, where the Bajo language and culture predominates today, ethnicities were historically more numerous (Nagatsu, 2013; Nuraini, 2008). Before Bajo migrants arrived, the main island was inhabited by indigenous Madura (East Javanese people), then several waves of migrants spread from Southern Sulawesi (including the Bajo, but also Bugis, Makasar and Mandar). However, the number of Bajo speakers then increased dramatically, quickly reaching one third of the total island population. Over time, non-Bajo speakers adopted Bajo languages and intermarried with the original Bajo. This also occurred elsewhere, with mixing between Bajo and neighboring 'land owners' being commonplace (Nuraini, 2016).

Consequently, all Bajo individuals share at least some common genetic background, suggesting that gene flow between these groups occurred until recently, and indeed, may still be ongoing today. Bajo communities maintained contact through sharing of goods, trading, fishing and marriage. Until recently, Bajo trading routes spanned Singapore in the west to New Guinea in the east, and Northeastern Borneo in the north to the Lesser Sunda Islands in the south (Nagatsu, 2013). Records note peaceful contact of Bugis and Bajo with Australian

Aborigines along the Northern coast of Australia, where the Bajo harvested trepang (sea cucumber) in shallow near-shore waters (Macknight, 1976; Pelras, 1997; Stacey, 2007), but no genetic contact is known. Strikingly, an established and stable Bajo sea trading route connected Southeastern Sulawesi with Southeastern Borneo (including Kotabaru Island), and from there, Kotabaru Island with Northeastern Borneo as far as Brunei, albeit with less intense activity (Nagatsu, 2013). This may explain the very recent admixture seen in the Bajo genomes, best illustrated by long shared IBD regions (Figure 4D).

Despite a complex genetic history involving creolization and multiple admixture events, the genomic data are suggestive of a single population origin for the Bajo, converging on Southern Sulawesi. The oldest estimated admixture event dates to the 4th century CE (Figure 4A, Annex F: Table S4) between ancestral Bugis (90%) and a Papuan group (10%). The two Bajo communities on Borneo appear to have emerged later, around the 12th century for the Kotabaru Bajo and the 14th century for Derawan Bajo, perhaps suggesting that Bajo communities lived in Southern Sulawesi for nearly eight hundred years before spreading west to Borneo. In contrast, the most recent linguistic studies support an origin of the Bajo language in the Southeast Borneo region, followed by a dispersal up the east coast of Borneo during the 11th century, only later spreading to the Southern Philippines and Northeast Borneo in the 13th-14th centuries (Blust, 2007). The linguistic and genetic evidence are therefore in broad agreement regarding the timing of the Bajo dispersal along the east coast of Borneo, but point to quite different locations for its origin: Southern Sulawesi for the gene pool and Southeast Borneo for the languages.

This apparent contradiction may be reconciled by aspects of recent history, as the expanding influence of the Malay kingdom of Śrīvijaya from the 7th century onward (7-13th centuries) (Beaujard, 2012a; Ras, 1968) heavily modified population structure and interactions in Southeast Borneo, triggering large population movements, such as the likely migration of the Banjar to Madagascar (Brucato et al., 2016; Kusuma et al., 2015). We postulate that similar causes may have also stimulated the dispersal of Bajo speakers from Southeast Borneo, again around the 11th century. The spread of the Bajo culture from Southeast Borneo possibly impacted pre-Bajo groups in Southern Sulawesi, leading to the emergence of the Kendari Bajo (Figure 4B). This Southern Sulawesi community with an incipient Bajo culture then unified the Bajo language and genome by settling other areas, creating communities such as the Kotabaru and Derawan Bajo (11th-14th centuries), likely with sex biased admixture between men from mainland Asia and women from the Bajo ancestral population (Annex F: Figure S5). This sex bias pattern is also observed in other sea-nomad populations along coastal Mainland Southeast Asia, such as the Moken sea-nomads, who exhibit lower female gene flow from mainland Asian populations (Dancause et al., 2009). Later, admixture with local groups occurred (Figure 4C), as well as ongoing contact between Bajo communities (Figure 4D). A similar process likely impacted other regions where the Bajo culture is common now, such as the Sulu archipelago in the Southern Philippines. Furthermore, Southern Sulawesi was long a center of trading activity during and after the Śrīvijaya Empire (Pelras, 1997), reaching its peak during the 16th century (Villiers, 1990). The main actors with significant role as traders are the Bugis, Makasar and Bajo, all with Southern Sulawesi connections (Sopher, 1977). Therefore, the presence of a Southern Sulawesi genetic background in all Bajo communities may also result from contact directly between these three sea trading groups.

Outside Indonesia, similar admixture behaviors, notably shared long-distance contact with local genetic contributions, has also been observed in other recent diasporas, such as the Romani and the Jewish in Europe. The Romani, who originated in Northwest India, later admixed with local European populations where they settled, yet with relatively modest genetic contributions (Mendizabal et al., 2012). Likewise, the Jewish diaspora has been traced back to the Levant, but local genetic admixture has been identified in each respective community (Behar et al., 2010). In both examples, all communities shared a common culture and genetic heritage, but like the Bajo, experienced gene flow from populations surrounding them.

The complexity of the Bajo genomic profile provides a striking reflection of their history, mediated by both migratory and local admixture events, and emphasizing their unique lifestyle played out across multiple geographical scales. Despite speaking Sama-Bajau languages, the Bajo prove to be diverse, encompassing rich genetic inputs from many groups, each distributed differently in the major Bajo communities, but homogenous for individuals within each community. It appears that contact between Bajo groups was a major feature of this history, but countered by strong regional contacts: Papuan influence in the Kendari Bajo; Banjar in the Kotabaru Bajo; and Filipino and Malay in the Derawan Bajo – all with an outsized influence from their Southern Sulawesi origin, possibly obtained by proxy from the Bugis. This genetic structure is in part due to a process of maritime creolization, exhibiting closer genetic connections with neighboring populations than distant Bajo groups. The sea oriented way of life of the Bajo and their prime role in the maritime trading network placed them in contact with surprisingly diverse populations, including South Asians and Papuans, whose contact left secondary traces in the genomes of the Bajo today. Studies of other Bajo communities, of which there are hundreds scattered across 1,300 km of Island Southeast Asia from east to west and 2,000 km from north to south, are likely to reveal more nuanced patterns of contact, as well as differential associations with means of subsistence, language, traditions and origin myths. In addition, this may well provide greater insight into the likely genetic histories of other nomadic populations that speak closely related languages, but span wide geographical areas.

3.2.2.6. Miscellaneous

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors thank Gludhug A. Purnomo, Isabella Apriyana, and Chelzie C. Darusalam from Eijkman Institute for Molecular Biology, Jakarta, Indonesia, for biological sampling, and Laure Tonasso and Stéphanie Schiavinato (University of Toulouse) for laboratory assistance. We acknowledge support from the GenoToul bioinformatics facility of the Genopole Toulouse Midi Pyrénées, France. This research was supported by the French ANR via grant ANR-14-CE31-0013-01 (OCEOADAPTO) to F.-X.R., the French Ministry of Foreign and European Affairs (French Archaeological Mission in Borneo; MAFBO) to F.-X.R., a Rutherford Fellowship from the Royal Society of New Zealand (RDF-10-MAU-001) to M.P.C., the French Embassy in Indonesia through its Cultural and Cooperation Services (Institut Français en Indonésie) to F.-X. R. and the Ministry of Education and Culture of Indonesia to P.K.

3.2.2.7. References

Beaujard, P., 2012a. Les mondes de l'océan Indien. Vol. 1 : De la formation de l'État au premier système-monde afro-eurasien (4e millénaire av. J.-C.-6e siècle apr. J.-C.). Armand Collin, Paris, France.

Beaujard, P., 2012b. Les mondes de l'ocean indien. Vol. 2 : L'océan Indien, au cœur des globalisations de l'Ancien Monde (7e-15e siècles). Armand Collin, Paris, France.

Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., Khusnutdinova, E.K., Balanovsky, O., Semino, O., Pereira, L., Comas, D., Gurwitz, D., Bonne-Tamir, B., Parfitt, T., Hammer, M.F., Skorecki, K., Villems, R., 2010. The genome-wide structure of the Jewish people. Nature 466, 238–242. doi:10.1038/nature09103

Bellwood, P., 2007. Prehistory of the Indo-Malaysian archipelago. ANU E Press, Canberra.

Blust, R., 2007. The linguistic position of Sama-Bajaw. Stud. Philipp. Lang. Cult. 15, 73–114.

Browning, S.R., Browning, B.L., 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81, 1084–1097. doi:10.1086/521987

Brucato, N., Kusuma, P., Cox, M.P., Pierron, D., Purnomo, G.A., Adelaar, A., Kivisild, T., Letellier, T., Sudoyo, H., Ricaut, F.-X., 2016. Malagasy genetic ancestry comes from an historical Malay trading post in southeast Borneo. Mol. Biol. Evol. 33, 2396–2400. doi:10.1093/molbev/msw117

Bulbeck, D., 2008. An integrated perspective on the Austronesian diaspora: The switch from cereal agriculture to maritime foraging in the colonisation of Island Southeast Asia. Aust. Archaeol. 67, 31–52.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7. doi:10.1186/s13742-015-0047-8

Chaudhuri, K.N., 1985. Trade and civilisation in the Indian Ocean: An economic history from the rise of Islam to 1750, First edition. ed. Cambridge University Press, Cambridge, UK.

Dancause, K.N., Chan, C.W., Arunotai, N.H., Lum, J.K., 2009. Origins of the Moken Sea Gypsies inferred from mitochondrial hypervariable region and whole genome sequences. J. Hum. Genet. 54, 86–93. doi:10.1038/jhg.2008.12

Delaneau, O., Marchini, J., 2014. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat. Commun. 5, 3934. doi:10.1038/ncomms4934

Delaneau, O., Marchini, J., Zagury, J.-F., 2012. A linear complexity phasing method for thousands of genomes. Nat. Methods 9, 179–181. doi:10.1038/nmeth.1785

Fenner, J.N., 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Am. J. Phys. Anthropol. 128, 415–423. doi:10.1002/ajpa.20188

Grangé, P. forthcoming. Langues et migrations des Bajos d'Indonésie.

Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., Myers, S., 2014. A genetic atlas of human admixture history. Science 343, 747–751. doi:10.1126/science.1243518

Hogan, D.W., Pattemore, S.W., 1988. Urak Lawoi': Basic structures and a dictionary. Dept. of Linguistics, Research School of Pacific Studies, Australian National University, Canberra, Australia.

Ivanoff, J., 2002. Rings of coral: Moken folktales, 1st edition. ed. White Lotus Co Ltd, Bangkok, Thailand.

Jakobsson, M., Rosenberg, N.A., 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23, 1801–1806. doi:10.1093/bioinformatics/btm233

Jubilado, R.C., Hussin, H., Manueli, M.K., 2011. The Sama-Bajaus of Sulu-Sulawesi seas: Perspectives from linguistics and culture. J. Southeast Asian Stud. 15.

Kayser, M., Lao, O., Saar, K., Brauer, S., Wang, X., Nürnberg, P., Trent, R.J., Stoneking, M., 2008. Genome-wide analysis iIndicates more Asian than Melanesian ancestry of Polynesians. Am. J. Hum. Genet. 82, 194–198. doi:10.1016/j.ajhg.2007.09.010

Kusuma, P., Cox, M.P., Pierron, D., Razafindrazaka, H., Brucato, N., Tonasso, L., Suryadi, H.L., Letellier, T., Sudoyo, H., Ricaut, F.-X., 2015. Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations. BMC Genomics 16, 191. doi:10.1186/s12864-015-1394-7

Larish, M.D., 2007. The position of Moken and Moklen within the Austronesian language family. University of Michigan, Ann Arbor.

Lawler, A., 2014. Sailing Sinbad's seas. Science 344, 1440–1445. doi:10.1126/science.344.6191.1440

Lawson, D.J., Hellenthal, G., Myers, S., Falush, D., 2012. Inference of population structure using dense haplotype data. PLOS Genet. 8, e1002453. doi:10.1371/journal.pgen.1002453

Lewis, M.P., Simons, G.F., Fennig, C.D. (Eds.), 2016. Ethnologue: Languages of the world, Nineteenth edition. ed. SIL International, Texas.

Liebner, H.H., 1998. Four oral versions of a story about the origin of the Bajo people of Southern Selayar, in: Robinson, K., Paeni, M. (Eds.), Living through histories: Culture, history, and social life in South Sulawesi. Department of Anthropology, Research School of Pacific and Asian Studies, Australian National University, Canberra, Australia.

Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., Berger, B., 2013. Inferring admixture histories of human populations using linkage disequilibrium. Genetics 193, 1233–1254. doi:10.1534/genetics.112.147330

Macknight, C.C., 1976. The voyage to Marege': Macassan trepangers in Northern Australia. Melbourne University Press, Melbourne, Australia.

Mendizabal, I., Lao, O., Marigorta, U.M., Wollstein, A., Gusmão, L., Ferak, V., Ioana, M., Jordanova, A., Kaneva, R., Kouvatsi, A., Kučinskas, V., Makukh, H., Metspalu, A., Netea, M.G., de Pablo, R., Pamjav, H., Radojkovic, D., Rolleston, S.J.H., Sertic, J., Macek Jr., M., Comas, D., Kayser, M., 2012. Reconstructing the population history of European Romani from genome-wide data. Curr. Biol. 22, 2342–2349. doi:10.1016/j.cub.2012.10.039

Nagatsu, K., 2013. Persisting maritime frontiers and multi-layered networks in Wallacea. Presented at the Asian CORE Program Seminar: "Interface, Negotiation, and Interaction in Southeast Asia," Kyoto, Japan.

Noorduyn, J., 1991. A critical survey of studies on the languages of Sulawesi. KITLV Press, Leiden, Netherlands.

Nuraini, C., 2016. The intangible legacy of the Indonesian Bajo. Wacana 17, 1–18. doi:10.17510/wacana.v17i1.427

Nuraini, C., 2008. Langue et production de récits d'une communauté Bajo des îles Kangean (Indonésie) (Ph.D. Thesis). Univ. de La Rochelle.

Oppenheimer, S., Richards, M., 2001. Fast trains, slow boats, and the ancestry of the Polynesian islanders. Sci. Prog. 84, 157–181.

Pallesen, A.K., 1985. Culture contact and language convergence, Special Monograph Issue 24. Linguistic Society of the Philippines, Manila, Philippines.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012. Ancient admixture in human history. Genetics 192, 1065–1093. doi:10.1534/genetics.112.145037

Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. PLoS Genet 2, e190. doi:10.1371/journal.pgen.0020190

Pelras, C., 1997. The Bugis, 1st edition. ed. Wiley-Blackwell, Oxford, UK.

Pembleton, L.W., Cogan, N.O.I., Forster, J.W., 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Mol. Ecol. Resour. 13, 946–952. doi:10.1111/1755-0998.12129

Pickrell, J.K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B., Reich, D., 2014. Ancient west Eurasian ancestry in southern and eastern Africa. Proc. Natl. Acad. Sci. 111, 2632–2637. doi:10.1073/pnas.1313787111

Pickrell, J.K., Pritchard, J.K., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8, e1002967. doi:10.1371/journal.pgen.1002967

Pierron, D., Razafindrazaka, H., Pagani, L., Ricaut, F.-X., Antao, T., Capredon, M., Sambo, C., Radimilahy, C., Rakotoarisoa, J.-A., Blench, R.M., Letellier, T., Kivisild, T., 2014. Genome-wide evidence of Austronesian–Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. Proc. Natl. Acad. Sci. 111, 936–941. doi:10.1073/pnas.1321860111

Ras, J.J., 1968. Hikajat Banjar: a study in Malay historiography. Martinus Nijhoff, The Hague.

Sather, C., 1997. The Bajau Laut: adaptation, history, and fate in a maritime fishing society of south-eastern Sabah. Oxford University Press, Oxford, UK.

Shringarpure, S.S., Bustamante, C.D., Lange, K., Alexander, D.H., 2016. Efficient analysis of large datasets and sex bias with ADMIXTURE. BMC Bioinformatics 17, 218. doi:10.1186/s12859-016-1082-x

Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., Clark, G.R., Reepmeyer, C., Petchey, F., Fernandes, D., Fu, Q., Harney, E., Lipson, M., Mallick, S., Novak, M., Rohland, N., Stewardson, K., Abdullah, S., Cox, M.P., Friedlaender, F.R., Friedlaender, J.S., Kivisild, T., Koki, G., Kusuma, P., Merriwether, D.A., Ricaut, F.-X., Wee, J.T.S., Patterson, N., Krause, J., Pinhasi, R., Reich, D., 2016. Genomic insights into the peopling of the Southwest Pacific. Nature 538, 510–513. doi:10.1038/nature19844

Solheim, W.G., II, Bulbeck, D., Flavel, A., 2006. Archaeology and culture in Southeast Asia: Unraveling the Nusantao. University of Philippines Press, Diliman, Quezon City.

Sopher, D.E., 1977. The Sea Nomads: A study of the maritime boat people of Southeast Asia. National Museum, Singapore.

Stacey, N., 2007. Boats to burn: Bajo fishing activity in the Australian fishing zone. ANU E Press, Canberra, Australia.

Verheijen, J.A.J., 1986. The Sama/Bajau language in the Lesser Sunda islands. Department of Linguistics, Research School of Pacific Studies, Australian National University, Canberra, Australia.

Villiers, J., 1990. Makassar: the rise and fall of an east Indonesian maritime trading state, 1512–1669, in: Kathririthamby-Wells, J., Villiers, J. (Eds.), The Southeast Asian Port and Polity: Rise and Demise. Singapore University Press, Singapore.

3.2.3. Malagasy genetic ancestry comes from an historical Malay trading post in Southeast Borneo

(Published in Molecular Biology and Evolution, 33 (9): 2396-2400, July 5th 2016.
doi:10.1093/molbev/msw117)

Nicolas Brucato[1,*], Pradiptajati Kusuma[1,2,*], Murray P. Cox[3,] Denis Pierron[1], Gludhug A. Purnomo[2], Alexander Adelaar[4], Toomas Kivisild[6,7], Thierry Letellier[1], Herawati Sudoyo[2,5] and François-Xavier Ricaut[1,†].

[1] Evolutionary Medicine Group, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse UMR 5288 CNRS, Université Toulouse III, Université de Toulouse. France, Toulouse, France
[2] Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular Biology, Jakarta, Indonesia
[3] Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
[4] Asia Institute, University of Melbourne, Melbourne, Australia
[5] Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia
[6] Department of Biological Anthropology, University of Cambridge, Cambridge, United Kingdom
[7] Estonian Biocentre, Tartu, Estonia

3.2.3.1. Abstract

Malagasy genetic diversity results from an exceptional proto-globalisation process that took place over a thousand years ago across the Indian Ocean. Previous efforts to locate the Asian origin of Malagasy highlighted Borneo broadly as a potential source, but so far no firm source populations were identified. Here, we have generated genome-wide data from two Southeast Borneo populations, the Banjar and the Ngaju, together with published data from populations across the Indian Ocean region. We find strong support for an origin of the Asian ancestry of Malagasy among the Banjar. This group emerged from the long-standing presence of a Malay Empire trading post in Southeast Borneo, which favoured admixture between the Malay and an autochthonous Borneo group, the Ma'anyan. Reconciling genetic, historical and linguistic data, we show that the Banjar, in Malay-led voyages, were the most probable Asian source among the analysed groups in the founding of the Malagasy gene pool.

3.2.3.2. Main manuscript

The Malagasy are the descendants of a unique prehistoric admixture event between African and Asian individuals, the outcome of one of the earliest proto-globalization processes that began approximately 4,000 years before present (BP) (Beaujard, 2012b, 2012a; Fuller et al., 2011; Lawler, 2014). Today, this island population in the western Indian Ocean shows strong biological (Fourquet et al., 1974; Pierron et al., 2014) and linguistic (Adelaar, 2009; Serva et al., 2012) connections to east coast African and Island Southeast Asian groups. Although this remarkable admixture event has long been studied (Fourquet et al., 1974; Hewitt et al., 1996; Hurles et al., 2005; Kusuma et al., 2015; Pierron et al., 2014; Soodyall et al., 1995; Tofanelli et al., 2009), the settlement process – specifically the geographic origin of the populations involved, dates of settlement and the nature of the population admixture event – are still routinely debated (Adelaar, 2009, 2017; Beaujard, 2012b; Kusuma et al., 2015; Pierron et al., 2014). Both linguistic and genetic studies suggest that the western and central regions of Indonesia (particularly Java, Borneo and Sulawesi) have the closest Asian genetic connections with modern Malagasy (Adelaar, 2017; Kusuma et al., 2015, 2016), and the island was probably settled by a small founding population (Cox et al., 2012). More specifically, recent anthropological research points to the Southeast region of Borneo as a potential source location for Malagasy (Adelaar, 2017; Kusuma et al., 2016). However, the Indonesian parental population of the Malagasy has so far not been clearly identified (Kusuma et al., 2015, 2016). To better reconstruct the settlement history of Madagascar and identify the Asian source populations, we conducted widespread sampling and genome-wide analysis of human groups across the Indian Ocean region. In particular, we present here new genome-wide data for two southeast Borneo populations, the Banjar (n=16) and the Ngaju (n=25) (Annex G: Figure S1), which have previously been noted as potential parental groups to the Malagasy (Adelaar, 2017). The newly generated data were analysed in the context of previously published data for a wide range of populations from the Indian Ocean rim (Annex G: Table S1).

The Malagasy Asian ancestry derives from Southeast Borneo.

To identify the most probable Asian parental groups of the Malagasy, we adopted a two-stage approach: first, we identified the most likely proxy populations using a dataset with wide geographical coverage but relatively low density of SNP, then followed by a higher density SNP dataset that gives increased statistical power. This allows us to reconstruct the admixture processes that led to the emergence of modern Malagasy. The admixture profile of our dataset

(2,183 individuals from 61 populations genotyped for 40,272 SNPs; Annex G: Figures S2 and Figure S3), based on ADMIXTURE analyses (Alexander et al., 2009), shows that the Malagasy genetic diversity is best described as a mixture of 68% African genomic components and 32% Asian components, corresponding well with the results of previous studie (Capredon et al., 2013; Pierron et al., 2014). While the African ancestry component in Malagasy appears to be broadly similar to that still present today in South African Bantu, the Asian ancestry presents a more complex pattern. This complexity is the key reason why previous studies have been unable to point firmly to a unique Asian source, making any subsequent anthropological inferences debatable (Pierron et al., 2014). This problem arose with a study of the Ma'anyan population of Borneo, whose language has long been identified as the closest to the Malagasy language (Dahl, 1951), but who surprisingly exhibit no clear genetic connection to Malagasy (Kusuma et al., 2016). Instead, the higher genetic complexity of the Asian ancestry component in Malagasy likely reflects the fact that more than a single source population was involved in its formation.

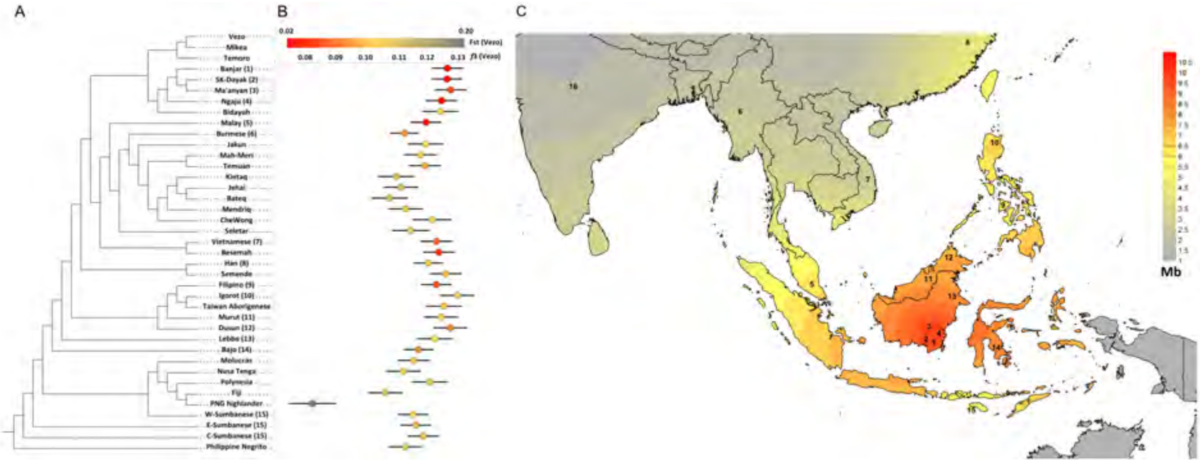


Figure 1. Localization of the Asian ancestry of Malagasy by (A) a TreeMix dendrogram, (B) F_{ST} distances and f3 statistics, all based on the Asian-SNP dataset, and (C) a shared IBD analysis based on haplotypes inferred from the high density of SNP dataset.(A) The TreeMix dendrogram was inferred imposing no a priori assumptions of migration, displaying only the tree topology. (B) Values of the f3(Asian-SNP Vezo, *X*; Yoruba) statistics are represented by dots with standard error bars. The color of each dot corresponds to the F_{ST} distances between the Asian ancestry of the Vezo and each Asian population using a grey-yellow-red color scale from the highest (0.196) to the lowest values (0.02). (C) The cumulative shared IBD (Mb) between pairs of Malagasy:Asian individuals were averaged to obtain one value of IBD sharing per Asian population. The obtained values are represented by a grey-yellow-red color scale. The numbers 1 to 15 correspond to the populations presented on the TreeMix dendrogram with the addition of 16 which stands for Brahmin Indian.

Affinities to the Malagasy Asian components are found at high frequency across several Island Southeast Asian groups, but notably in Malay, a dominant group of ancient seafaring traders (first millennium CE), and admixed groups from Borneo (i.e. Banjar, Ngaju, South Kalimantan Dayak, Lebbo, Murut, Dusun and Bidayuh; Annex G: Figure S2). The connection between Malagasy, and the Borneo and Malay populations, is supported by f3-statistics (*Z*-scores < -2; Annex G: Table S2) (Patterson et al., 2012) and TreeMix analyses (35% of Malay/Borneo gene flow to Malagasy; Annex G: Figure S4) (Pickrell and Pritchard, 2012). However, to more specifically identify the Asian ancestry of the Malagasy genome, we performed a Local Ancestry analysis with PCAdmix (Brisbin et al., 2012) using two proxy parental meta-populations comprising 100 individuals with African ancestry (randomly selected from Yoruba, South African Bantu, Kenyan Luhya and Somali groups) and Asian ancestry (randomly selected from Chinese, Philippine Igorot, Bornean Ma'anyan and Malay groups). Masking the haplotypes inferred to derive from Africa, we performed an Ancestry-Specific PCA (Patterson et al., 2006) and TreeMix analysis (Pickrell and Pritchard, 2012). Both show that the Asian genomic components of Malagasy cluster tightly with Southeast Borneo groups (Banjar, South Kalimantan Dayak, Ngaju and Ma'anyan) (1,664 SNPs; Figure 1A and Annex G: Figure S5). This connection is supported by the highest f3-statistics and the lowest F_{ST} genetic distances also being observed between Asian ancestry of the Malagasy and these same Southeast Borneo groups ($f3 > 0.12$; $F_{ST} < 0.02$; Figure 1B; Annex G: Table S3 and Table S4).

To explore this connection in more detail, we turned to the high density SNP dataset (551 individuals from 24 populations genotyped for 374,189 SNPs; Supplementary table S1, Supplementary Material online). This allows more statistically powerful analyses based on haplotypes. We confirmed the earlier result that Malagasy have the highest values of cumulative shared Identity-By-Descent fragments (IBD) with Southeast Borneo populations (Figure 1C; Supplementary figure S6, Supplementary Material online). To expand on this, however, we inferred the population sources of the Malagasy, their relative ratios and the dates of potential admixture events with GLOBETROTTER (Hellenthal et al., 2014), defining each population in our dataset as a donor/surrogate group and the Malagasy as the recipient, using the haplotype 'painting' data obtained with Chromopainter (Lawson et al., 2012). The best fit outcome for the Malagasy was obtained under a model of a single admixture event between two sources: the Banjar representing 37% of modern Malagasy and the South African Bantu population representing the other 63% (r^2=0.99, $P<0.01$; Figure 2 and Annex

G: Table S5). The admixture event was dated to 675 years BP (95% CI: 625-725 years BP, Annex G: Table S5), which is similar to the dates of admixture estimated by ALDER (550-750 years BP) using Banjar population in combination with the South African Bantu (Annex G: Table S6) (Loh et al., 2013). When each Malagasy ethnic group is analysed separately, similar parental populations, admixture proportions and dates are obtained with the noticeable older estimated dates towards the east coast of Madagascar (Annex G: Table S5).

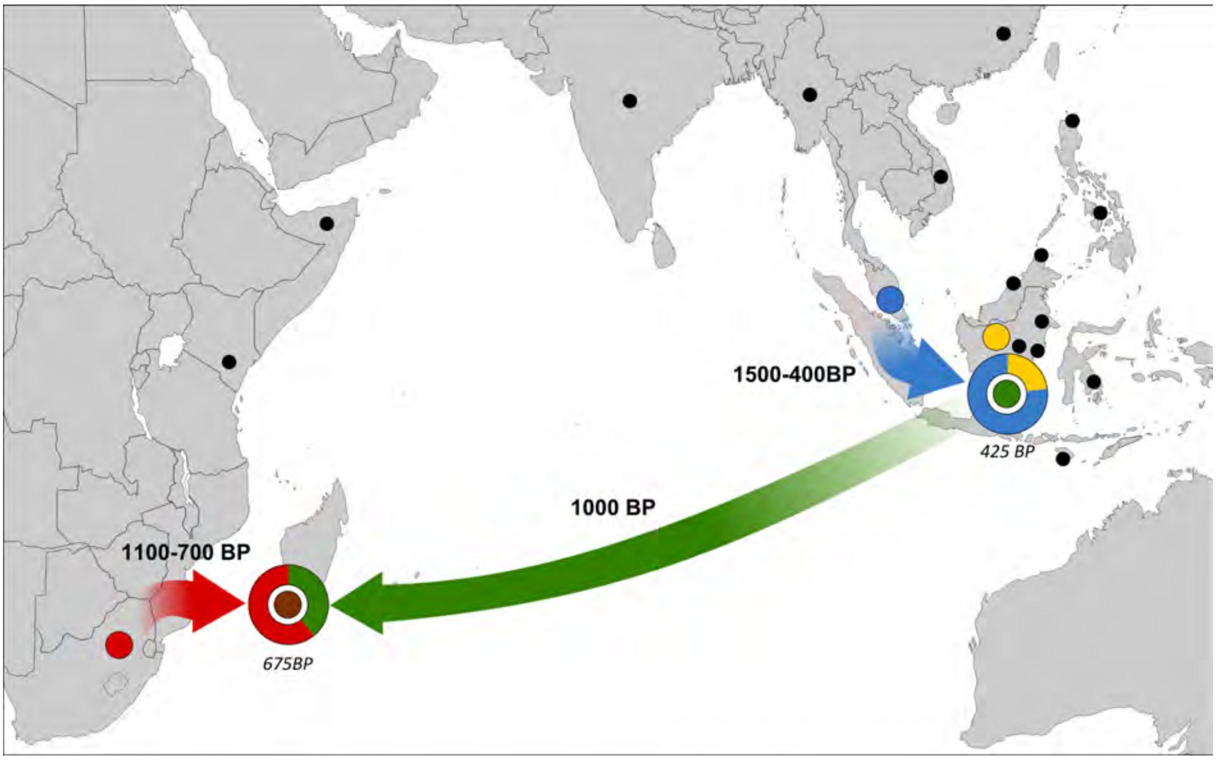


Figure 2. Scenario for the Asian genetic ancestry in Malagasy based on the best fit models inferred by GLOBETROTTER. The brown circle represents the Malagasy (bottom left), while the green circle represents the Banjar (right). Red semicircles show the African ancestry (South African Bantu), while the other semicircles represent Asian ancestry from Malay (blue), Banjar (green) and Ma'anyan (yellow). Black dots highlight other populations included in the high density genomic dataset. The arrows show migration events, with indicative routes, inferred in our analyses with dates of admixture in italic estimated by GLOBETROTTER. Dates in bold correspond to dates of migration estimated from archaeological and historical data (Beaujard, 2012b).

Crucially, these dates of genetic admixture, in agreement with a previous study (Pierron et al., 2014), reflect the midpoint or end of noticeable admixture between groups of Asian and African ancestry in Madagascar, rather than the start of this contact. Therefore they could correspond to the end of the period of the main Austronesian presence in Madagascar that started around the first millennium CE (Adelaar, 1995, 2017; Cox et al., 2012; Dahl, 1951, 1991; Dewar and Wright, 1993). On the other hand, around 1,100-700 years BP, climatic changes in the South of Africa forced Bantu populations to move to more hospitable places (Huffman, 2000). This South Bantu migration has previously been suggested as an explanation for the higher density of populations observed in the South of Madagascar (Beaujard, 2012b). As all of our sampled groups live in the South of Madagascar, and considering that the estimated dates of admixture are more recent on the west coast (Annex G: Table S5 and Table S6), it is tempting to interpret our admixture date as marking the last significant Bantu migration to Madagascar, perhaps initiated by climatic changes in Africa.

The Proto-Malagasy people were a Malay-Ma'anyan admixed group.

These analyses clearly identify an Austronesian-speaking population, the Banjar in the southeast region of Borneo, as the closest Asian sources for modern Malagasy. Linguistic reconstruction of the Proto-Malagasy language indicated that it appears to be derived mainly from the Southeast Barito language spoken today by the Ma'anyan (Dahl, 1951), a Southeast Borneo group. However, we have previously shown (Kusuma et al., 2016), and reconfirm here, that in genetic analyses, the Ma'anyan are only distantly related to the Malagasy in terms of genetics. In contrast, the Banjar currently speak a Malay language. Interestingly the linguistic studies indicated a noticeable proportion of Malay words in present-day Malagasy languages (Adelaar, 1989, 2009). We estimated the best fit scenario for the admixture process by modeling the current Banjar diversity with GLOBETROTTER (Hellenthal et al., 2014). The genetic diversity of the Banjar best fits a model of a unique admixture event (r^2=0.62; $P < 0.01$) between two major ancestries that can be traced back to Malay (77%) and Ma'anyan (23%) (Figure 2 and Annex G: Table S5). We estimate the date of the last noticeable admixture approximately 425 years BP (95% CI: 275-500 years BP). Since the Banjar originated from a Ma'anyan-Malay admixture, at a time preceding the supposed date of migration to Madagascar (i.e. 1,000 years BP), the ancestors of the Banjar were presumably speaking a language close to that reconstructed for Proto-Malagasy (Adelaar, 2017). Although we cannot fully exclude that the Malay-Ma'anyan admixture occurred in Malagasy, prior to the Bantu gene flow, the observed haplotypic structure is so similar to the ones observed in the Banjar that it is more parsimonious to interpret this admixture to have happened first in Borneo. This analysis reconciles both the linguistic and genetic data, strengthening our scenario placing the Banjar as the main Asian parental populations of the Malagasy. Their current genetic diversity appears to be the reflection of the historical relationship between Madagascar, Southeast Borneo and the Malay. The maritime routes linking Madagascar to Borneo were particularly exploited during the rapid expansion of trading networks led by the

Hindu Malay Kingdoms, such as Srivijaya (6th-13th centuries) (Beaujard, 2012b; Ras, 1968). Established on the islands of Sumatra and Java, the Malay traded with far-distant regions, notably across East Asia and reaching as far as East Africa (Beaujard, 2012b). Their influence increased across all the Southeast Asian islands, notably in Borneo where they established several trading posts, such as one in the city of Banjarmasin in S#outheast Borneo (Beaujard, 2012b; Ras, 1968). As related in the only Banjarese historical records available, the Hikayat Banjar ('Tale of Banjar') (Ras, 1968), the main city of the Banjar population was a major trading post in the former Malay Empires. This probably favoured interactions with inland groups in Borneo, such as the Ma'anyan, but also with other populations such as the Bajo sea nomads (Annex G: Figure S6 and Table S5) (Adelaar, 2009; Beaujard, 2012b; Kusuma et al., 2015). The Malay domination of trading networks collapsed during the 15th-16th centuries, with the emergence of several sultanates and the arrival of Europeans, which could correspond to the end of noticeable Malay gene flow into the Banjar population, as indicated by our estimated date of admixture between Malay and Ma'anyan around 425 years BP.

Our study provides strong support for a new scenario for the Austronesian settlement of Madagascar, reconciling cultural, linguistic and genetic data, in which the Banjar population played key roles in establishing the Asian founder population of the Malagasy. The Malay trading networks during the first millennium CE triggered one of the earliest proto-globalization processes, bringing Southeast Asian populations to East Africa (Beaujard, 2012b). The Banjar, currently living in Southeast Borneo, show the highest affinity to the main genetic ancestry component in Malagasy and can therefore be suggested to be likely to have been the ethnic group that accompanied the Malay in their maritime voyages to Madagascar. This population with composite ethnic ancestry emerged from the long-standing presence of Malay in Borneo, creating an admixed community with local Austronesian-speaking groups, such as the Ma'anyan. Before the probable date of migration (around 1,000 years BP), the ancestors of the current Banjar would have contained both Malay and Ma'anyan genetic diversity, and probably linguistic inheritances from both (Adelaar, 2017). Although the exact maritime route(s) of migration from Borneo to Madagascar are still an open question, our study identifies the flow of Malay-Ma'anyan genomic ancestries as carried by Banjar ancestors as the source to Malagasy.

3.2.3.3. Methods

All experimental and analytical procedures are described in the Supplementary Method file (Supplementary Material online). This study was approved by the Research Ethics Commission of the Eijkman Institute for Molecular Biology (Jakarta, Indonesia).

3.2.3.4. Miscellaneous

Acknowledgements

We thank the Eijkman Institute for Molecular Biology (Jakarta, Indonesia), especially Alida Harahap and Deni Feriandika, and the Department of Health, Banjarmasin city in the province of South Kalimantan, for accommodating the sampling program. We wish to acknowledge support from the GenoToul bioinformatics facility of the Genopole Toulouse Midi Pyrénées, France. We thank Harilanto Razafindrazaka for her useful comments on the manuscript. This research was supported by French ANR grant number ANR-14-CE31-0013-01 (grant OceoAdapto to F-X.R), the French Ministry of Foreign and European Affairs (French Archaeological Mission in Borneo (MAFBO) to F-X.R), and the French Embassy in Indonesia through its Cultural and Cooperation Services (Institut Français en Indonésie). Genetic data has been deposited at the European Genome-phenome Archive (EGA, https://ega-archive.org), jointly managed by the EBI and CRG, under accession number EGAS00001001841.

3.2.3.5. References

Adelaar, K.A., 2017. Who were the first Malagasy, and what did they speak?, in: Acri, A., Landmann, A. (Eds.), Spirits and Ships: Cultural Transfer in Early Monsoon Asia. Institute of Southeast Asian Studies - Yusof Ishak Institute, Singapore, pp. 441–469.

Adelaar, K.A., 2009. Loanwords in Malagasy, in: Haspelmath, M., Tadmor, U. (Eds.), Loanwords in the world's languages: A cmparative handbook. De Gruyter Mouton, Berlin, Germany, pp. 717–746.

Adelaar, K.A., 1995. Borneo as a cross-roads for comparative Austronesian linguistics, in: The Austronesians: Historical and comparative perspectives. ANU E Press, Canberra, Australia, pp. 81–102.

Adelaar, K.A., 1989. Malay influence on Malagasy: linguistic and culture-historical implications. Ocean. Linguist. 28, 1–46. doi:10.2307/3622973

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664. doi:10.1101/gr.094052.109

Beaujard, P., 2012b. Les mondes de l'ocean indien. Vol. 2 : L'océan Indien, au cœur des globalisations de l'Ancien Monde (7e-15e siècles). Armand Collin, Paris, France.

Beaujard, P., 2012a. Les mondes de l'océan Indien. Vol. 1 : De la formation de l'État au premier système-monde afro-eurasien (4e millénaire av. J.-C.-6e siècle apr. J.-C.). Armand Collin, Paris, France.

Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., Reynolds, A., Ostrer, H., Mezey, J.G., Bustamante, C.D., 2012. PCAdmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. Hum. Biol. 84, 343–364. doi:10.3378/027.084.0401

Capredon, M., Brucato, N., Tonasso, L., Choesmel-Cadamuro, V., Ricaut, F.-X., Razafindrazaka, H., Rakotondrabe, A.B., Ratolojanahary, M.A., Randriamarolaza, L.-P., Champion, B., Dugoujon, J.-M.,

2013. Tracing arab-islamic inheritance in Madagascar: Study of the Y-chromosome and mitochondrial DNA in the Antemoro. PLoS ONE 8, e80932. doi:10.1371/journal.pone.0080932

Cox, M.P., Nelson, M.G., Tumonggor, M.K., Ricaut, F.-X., Sudoyo, H., 2012. A small cohort of Island Southeast Asian women founded Madagascar. Proc. R. Soc. B Biol. Sci. 279, 2761–2768. doi:10.1098/rspb.2012.0012

Dahl, O.C., 1991. Migration from Kalimantan to Madagascar. Norwegian University Press : Institute for Comparative Research in Human Culture, Oslo, Norway.

Dahl, O.C., 1951. Malgache et maanjan: une comparaison linguistique. Edege-Intituttet, Oslo, Norway.

Dewar, R.E., Wright, H.T., 1993. The culture history of Madagascar. J. World Prehistory 7, 417–466. doi:10.1007/BF00997802

Fourquet, R., Sarthou, J., Roux, J., Acri, K., 1974. Hémoglobine S et origines du peuplement de Madagascar : Nouvelle hypothèse sur son introduction en Afrique. Arch. L'Institut Pasteur Madag. 43, 185–220.

Fuller, D.Q., Boivin, N., Hoogervorst, T., Allaby, R., 2011. Across the Indian Ocean: the prehistoric movement of plants and animals. Antiquity 85, 544–558. doi:10.1017/S0003598X00067934

Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., Myers, S., 2014. A genetic atlas of human admixture history. Science 343, 747–751. doi:10.1126/science.1243518

Hewitt, R., Krause, A., Goldman, A., Campbell, G., Jenkins, T., 1996. Beta-globin haplotype analysis suggests that a major source of Malagasy ancestry is derived from Bantu-speaking Negroids. Am. J. Hum. Genet. 58, 1303–1308.

Huffman, T., 2000. Mapungubwe and the origins of the Zimbabwe culture, in: Leslie, M., Maggs, T. (Eds.), African naissance: The Limpopo valley 1000 years Ago. South African Archaeological Society, Cape Town, South Africa, pp. 14–29.

Hurles, M.E., Sykes, B.C., Jobling, M.A., Forster, P., 2005. The dual origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from maternal and paternal lineages. Am. J. Hum. Genet. 76, 894–901. doi:10.1086/430051

Kusuma, P., Brucato, N., Cox, M.P., Pierron, D., Razafindrazaka, H., Adelaar, A., Sudoyo, H., Letellier, T., Ricaut, F.-X., 2016. Contrasting linguistic and genetic origins of the Asian surce populations of Malagasy. Sci. Rep. 6, 26066. doi:10.1038/srep26066

Kusuma, P., Cox, M.P., Pierron, D., Razafindrazaka, H., Brucato, N., Tonasso, L., Suryadi, H.L., Letellier, T., Sudoyo, H., Ricaut, F.-X., 2015. Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations. BMC Genomics 16, 191. doi:10.1186/s12864-015-1394-7

Lawler, A., 2014. Sailing Sinbad's seas. Science 344, 1440–1445. doi:10.1126/science.344.6191.1440

Lawson, D.J., Hellenthal, G., Myers, S., Falush, D., 2012. Inference of population structure using dense haplotype data. PLOS Genet. 8, e1002453. doi:10.1371/journal.pgen.1002453

Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., Berger, B., 2013. Inferring admixture histories of human populations using linkage disequilibrium. Genetics 193, 1233–1254. doi:10.1534/genetics.112.147330

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012. Ancient admixture in human history. Genetics 192, 1065–1093. doi:10.1534/genetics.112.145037

Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. PLoS Genet 2, e190. doi:10.1371/journal.pgen.0020190

Pickrell, J.K., Pritchard, J.K., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8, e1002967. doi:10.1371/journal.pgen.1002967

Pierron, D., Razafindrazaka, H., Pagani, L., Ricaut, F.-X., Antao, T., Capredon, M., Sambo, C., Radimilahy, C., Rakotoarisoa, J.-A., Blench, R.M., Letellier, T., Kivisild, T., 2014. Genome-wide evidence of Austronesian–Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. Proc. Natl. Acad. Sci. 111, 936–941. doi:10.1073/pnas.1321860111

Ras, J.J., 1968. Hikajat Banjar: a study in Malay historiography. Martinus Nijhoff, The Hague.

Serva, M., Petroni, F., Volchenkov, D., Wichmann, S., 2012. Malagasy dialects and the peopling of Madagascar. J. R. Soc. Interface 9, 54–67. doi:10.1098/rsif.2011.0228

Soodyall, H., Jenkins, T., Stoneking, M., 1995. "Polynesian" mtDNA in the Malagasy. Nat. Genet. 10, 377–378. doi:10.1038/ng0895-377

Tofanelli, S., Bertoncini, S., Castrì, L., Luiselli, D., Calafell, F., Donati, G., Paoli, G., 2009. On the origins and admixture of Malagasy: New evidence from high-resolution analyses of paternal and maternal lineages. Mol. Biol. Evol. 26, 2109–2124. doi:10.1093/molbev/msp120

IV. Conclusions and perspectives

4.1. Conclusions and contributions

My thesis had two main aims: 1) To determine the origin of human genetic diversity in Indonesia; and 2) To assess how this diversity impacts the gene pool of populations in the Indian Ocean region. To achieve these aims I sampled new Indonesian populations in Borneo, and used a broad range of molecular and bioinformatic approaches for sample and data analyses, designed to answer questions regarding human demographic history and migration. Discussions of the results have been made in an interdisciplinary context, drawing on inferences from other disciplines, such as linguistics, history, archeology, and anthropology. I have published the results of my doctoral study as 1st author in five articles in peer-reviewed scientific journals.

In the first part of this study (uniparental markers analyses, section 3.1), I used uniparental markers to determine the genetic background of the Bornean populations within the complex genetic diversity of Indonesia. Genetically, there is clear segregation between western and eastern Indonesian populations, and the genetic profiles of the new Bornean populations are related to those of western Indonesian populations. They have mitochondrial and Y-chromosome haplogroups from East Asia/mainland Southeast Asia, and a limited number of western Eurasian haplogroups. In relation to this, results of which are published in the second article, I detected the genetic influence of western Eurasians on present-day Indonesian populations. I found that populations from Java and Bali have the most significant influence from western Eurasian mtDNA and Y-chromosome haplogroups, and in contrast populations from Borneo, Maluku, Nias, and Mentawai have the least genetic influence.

Concerning the relationship between Indonesian and Malagasy populations, I performed standard genetic distance and haplotype-sharing analyses to observe which Indonesian populations are the closest genetically to the Malagasy. The study showed that populations from eastern Indonesia exhibit the closest link with the Malagasy, specifically Sulawesi and Lesser Sunda island populations. This is most probably driven by the strong presence of the Polynesian motif mitochondrial (macro)haplogroup (B4a1a1*) in eastern Indonesia and Madagascar. Interestingly, no plausible connection with the Malagasy's linguistic sister population, the Ma'anyan was found. Moreover, according to my preliminary results on whole mitogenome sequences in the populations sampled (see Table 2.1 and 2.2), there are no Malagasy motif haplogroups (B4a1a1b) observed (Annex H). This motif has been tested in research by Cox et al. (2012) on ~3,000 individuals from populations across the Indonesian archipelago, but no positive results have yet emerged. It still remains unclear whether this haplogroup emerged in situ, during migration, or in unsampled Indonesian populations.

In the second part of this study (autosomal DNA analyses, section 3.2), I explored the autosomal genetic diversity, using genome-wide SNPs, of the Bornean populations and their connection with the Malagasy. The Ma'anyan have a unique genetic component that has never been observed at such a high frequency in one population. Explored further, this component is strongly related to the Austronesian genetic profile, which is carried by Aboriginal Taiwanese and the Igorot (Kankanaey) of the Philippines, and not to mainland Southeast Asians who are often affiliated to Austroasiatic speakers. A new ancestral component in the Ma'anyan is an important discovery, and its presence does not appear to be linked to any recent admixture events based on the formal admixture tests conducted. This result may give important information on the scenario of Austronesian dispersal throughout ISEA, such as the routes, timing of dispersals, and also the nature of possible admixture events. I also performed analyses on the genetic history of the Bajo sea nomads; their way of life has been instrument in shaping Indonesian genetic diversity and they were potentially involved in the settlement to Madagascar. Dispersed along coastal areas on different islands in Indonesia, the Bajo still speak a single language; the scientific debate on the origin of this language is in progress. In this study, I proposed a scenario for the origin and dispersal of the Bajo, including sex bias admixture during migrations. Their sea-orientated way of life, their prime role in the maritime trading networks and their location at the maritime frontier, placed the Bajo in contact with various populations (including South Asians and Papuans), whose genomes left traces in their communities. By studying the Bajo, we will be able to investigate biological/genetic interactions in such maritime interaction network, which has been the main way for populations to interact in ISEA since the mid-Holocene.

Concerning the origins of the Malagasy, the results of genome-wide SNP analyses confirmed that the Ma'anyan are not the primary ancestor of the Malagasy. After subsequent analyses it was determined that the most plausible Asian ancestor of the Malagasy is the Banjar people, a population who lived in Banjarmasin, southeast Borneo, an admixed population between the Ma'anyans and Malay immigrants. I tested data from the Bajo and the Bornean populations with the Illumina Omni 5 Indonesian data altogether (Annex I), using GLOBETROTTER analysis. The results, however, did not change, and the Banjar remained the most plausible

Asian ancestral population to the Malagasy. This conclusion is supported by historical and linguistic evidence; Indonesians were carried on the Srivijaya ship by the Malay Kingdom during their trading activities to eastern Africa (1st millennium CE), during a period of significant proto-globalization along the Indian Ocean rim.

The key factor that links all of these results together is the presence of the Malay Kingdom in southern Borneo; this Kingdom played a very important role in triggering population admixture and demographic fluctuations in the region, and in Indonesia in general. A Srivijaya Kingdom outpost was established in southern Borneo because of its strategic location. The long coastline might have extended northward as far as 200 kilometers, and further inland than at present, and this outpost was possibly located near the Tanjung region, the auto-identified original homeland of the Ma'anyans. It forced the Ma'anyan people to move further towards the forests in the current East Barito district, where they are still present today. As the coastline moved southward, the trading post was also moved south and it later formed the city of Banjarmasin, which is the main city, commercial state and activity centre of the trading network of this region. During their establishment, the Malay Kingdom introduced Malay cultures and words to the indigenous populations in the region. Moreover, gene flow has been detected from populations with the Malay genetic background to indigenous populations living in the Malay Kingdom, mainly in southern and southeastern Borneo, but not in central or eastern Borneo. This event might have prompted the migration of proto-Malagasy groups from southern Borneo into Madagascar in a Malay-led oceanic voyage, although the cause of their migration remains elusive (potentially related to trading or slavery). This event might also have triggered the outbound dispersal of proto-Bajo groups from southern Borneo, which initiated the life history of the Bajo sea nomads in the Indonesian archipelago. On the other hand, the reign of the Hindu/Malay Kingdom(s) in the archipelago accommodated trading activities by merchants from South Asia and the Middle East. As a consequence, not only goods and commodities were exchanged, but also genes. Gene flow from India and the Middle East to several Indonesian populations has probably been constant since their first contact and potentially intensified during the period of the Hindu/Malay Kingdoms in Indonesia (6th to 16th century CE). Even minor Indian genetic traces that have been found in Madagascar are thought to have been brought from Indonesia, and not directly from India. This indicates that the influence of Indianised Kingdoms in Indonesia was not restricted simply to cultural assimilation, trading and economic power, but was also associated to some degree of gene flow into Indonesia and then to Madagascar.

In the context of the results generated, summary conclusions can be made for each of the four questions of my doctoral study (see p13):

1. Borneo populations enrich the overall diversity of Indonesian populations, and a previously undetermined, unique Austronesian component was identified in Ma'anyan speaking populations.
2. Western Eurasian populations have left genetic traces in some Indonesian populations, mainly in Java, Bali and Sumatra, however, the proportion of the autosomal component is low. This influence was probably brought into Indonesia during the reign of the Malay Kingdoms, which started in the 6th century CE.
3. Despite a close linguistic affinity with the Malagasy, the Ma'anyan are not their direct ancestors. The most probable ancestor of the Malagasy is the Banjar, an admixed population composed of Ma'anyan and Malay individuals, whose genetic influence was present during the trading activities of the Malay Kingdom on the eastern coast of Africa from the mid 7th century CE.
4. Bajo sea nomads have a unique genetic background, which can be explained by their nomadic lifestyle. They are, however, not linked directly to the migration process from southern Borneo to Madagascar, although their dispersal started at the same time as people from southern Borneo departed to Madagascar.

In conclusion, this thesis provides valuable insights into the connection between Indonesian and Malagasy populations, helping to address important questions related to the history and ancestry of these groups.

4.2. Limitations

In this study, the precision of dating generated by software is not absolute, as it is approached by generation interval in time of year. I calculated the date estimation in years by multiplying the date estimate in generations with the average number of years per generation. The date of admixture cannot be equated to the date of migration, but to the date of the last admixture event. The date of migration could be addressed using historical and archaeological evidence, but this evidence is limited, hence should be interpreted with caution.

4.3. Perspectives

Sample coverage

A key area for future work should be to identify the exact (or the closest) population as the source of admixture, as this can provide important insights on the evolutionary history of populations in general. To date, most studies only cover the major global populations in Europe and Africa, and the inclusion of native indigenous regional populations is lacking. Studies would benefit greatly from improving their global coverage of samples, particularly in Indonesia and in other populations from ISEA. Concerning ISEA, a comprehensive coverage of samples in this area would help to identify the source of local Asian ancestry in each of the unsampled populations. Borneo is understudied in Indonesian population genetic research. Almost all major local populations in Sumatra, Java, Bali, Sulawesi, Lesser Sunda Islands and Maluku have been sampled, however, Borneo has not been sampled thoroughly. In collaboration with international research institutions, I am co-leading a genomic research project in Borneo, together with Prof. Herawati Sudoyo from the Eijkman Institute for Molecular Biology, Jakarta, funded by the Indonesian Science Fund (DIPI). Starting this year, I am planning other sampling activities to cover all major ethnic and linguistic groups in Borneo. This data will generate important information on the history of human settlement in Borneo and the population structure of the indigenous Dayak people in Kalimantan: how their genetic diversity reflects that of present-day Indonesians, to reveal how people have adapted to different environments and evolved different lifestyles, continuing the works that have been done during my doctoral study (Publication list, Annex J)

Concerning the history of human settlement in Madagascar, the ancestor of the Malagasy has been pinpointed in this study. However, I only used three southern Malagasy populations as representatives of the Malagasy population, and therefore, more individuals in other Malagasy populations should be included in future studies to confirm this finding. In addition, the exact maritime routes taken during the migration events remain unknown. The inclusion of sample populations across the Indian Ocean rim, may provide information on a possible indirect route from Indonesia to eastern Africa. In this context, the inclusion of various populations on the eastern coast of Africa, such as the Swahili and Comoros, can give further insight on the Indonesian influence in the western corner of the Indian Ocean, and also on the African ancestor of the Malagasy, so concluding the scenario of human settlement in Madagascar. My current involvement in the international research projects led by Dr. François-Xavier Ricaut at the University of Toulouse (genomic and archaeological projects in Papua New Guinea and genomic project on the Swahili corridor in East Africa), will allow us to screen populations in East Africa and in the regions of eastern Indonesia and Melanesia where the Polynesian motif possibly arose. Therefore, the aim is to determine whether the Malagasy's genetic motif is detected outside Madagascar, to investigate the nature of its apparition and diffusion.

Sequence data

Technological advances have made whole genome sequencing from considerable number of individuals cost-effective, providing favorable circumstances to analyse data without ascertainment-bias by sampling all the variation in the genome. Inferences of demographic parameters can be estimated without modeling or correcting any ascertainment biases. Rare alleles will be observed, thus provide information about recent population structuring. In haplotype level, shared segments that contain rare alleles can give insight on ancestry signatures in admixed populations, and also to know the source of admixture. Full and more dense data on haplotype length distribution can give thorough information while modeling the demographic events such as population admixtures and its divergences, as well as its timing. Not only covering demographic questions, whole genome sequence data will also provide robust information on genetic adaptation to selective pressures and environmental changes. The international collaborative genomic projects in which I am involved, will, for the first time, explore these areas in indigenous and native populations.

Mitogenome sequences from various populations in Indonesia and eastern Africa (~800 mitogenome sequences), that were generated during the final year of my doctoral study, will be analysed further to model the settlement of populations in Indonesia. For example, the demographic history of the Ma'anyan, who has a unique genetic background, but doesn't exhibit any admixture events, can be modeled using mitogenome sequence data in the context of the maternal lineage. In addition, mitogenome data on Malagasy populations, once available in the scientific community, will be useful to model the divergence, split, and migrations from respective Asian and African ancestral populations, to complement the results from the genome-wide SNP data presented in this dissertation.

Study of human disease and migration

The migration of large numbers of people creates opportunities for the spread and establishment of common or novel infectious diseases. There are significant inter-regional differences in geographical and climatic conditions, and indigenous people have adapted to the specific conditions and endemic diseases in their respective homelands. The role of environmental change is important in the dissemination of infectious diseases, and when people migrate they also bring pathogens from their specific regions. Studying various diseases as population history and migration markers can be a complement to general population genetics / molecular anthropology studies, to gain a holistic view on where populations originated, where people migrate, and how people have adapted to their old and new environments. Traces of infectious disease parasites and their vectors can bring an understanding, not only of adaptation mechanisms against diseases, but also on the routes taken by humans during their global migrations.

Studies of adaptation to malaria have shown that populations have unique regional-level genetic profiles to counter malaria infection, such as sickle cell anemia and Duffy negativity in western Africa, Hemoglobin E (HbE) and Southeast Asian Ovalocytosis (SAO) in Southeast Asia, and alpha and beta Thalassaemia and G6PD deficiency which have different genotypes in different populations across the world. Research by Rabe et al. (2002) on several villages in the highlands of central Madagascar have shown a small percentage of SAO presence. Further surveys of SAO presence throughout Madagascar, and determining the presence of the SLC4A1 gene haplotype around the SAO-deletion region, can give additional supporting evidence of an Indonesian influence in Madagascar. Studies detecting SAO presence in Indonesian and eastern African populations is currently in progress. In addition, other unique Indonesian genotypes for molecular adaptation against malaria, as mentioned, will also be investigated in the future.

I have started a collaboration with the Pasteur Institute in Paris, to study the distribution of HTLV-1 and HHV-8 viruses in Indonesia. These viruses have wide distributions in Oceania and East Asia, but no studies have been conducted in Indonesia in the context of it being a focal crossroad in human migration from East Asia into Oceania. Traces of these viruses must be present in Indonesia. This study will bring new understanding on HTLV-1 and HHV-8 genotype diversity, and will also complement the scenario of Austronesian eastward dispersal into the Pacific.

V. Bibliography

Adelaar, K.A., 2017. Who were the first Malagasy, and what did they speak?, in: Acri, A., Landmann, A. (Eds.), Spirits and Ships: Cultural Transfer in Early Monsoon Asia. Institute of Southeast Asian Studies - Yusof Ishak Institute, Singapore, pp. 441–469.

Adelaar, K.A., 2009. Loanwords in Malagasy, in: Haspelmath, M., Tadmor, U. (Eds.), Loanwords in the World's Languages: A Comparative Handbook. De Gruyter Mouton, Berlin, Germany, pp. 717–746.

Adelaar, K.A., 2006. The Indonesian migrations to Madagascar: making sense of the multidisciplinary evidence, in: Austronesian Diaspora and the Ethnogenesis of People in Indonesian Archipelago: Proceedings of the International Symposium. Presented at the Austronesian diaspora and the ethnogenesis of people in Indonesian archipelago, LIPI Press, Jakarta, Indonesia, pp. 205–232.

Adelaar, K.A., 2005. The Austronesian languages of Asia and Madagascar: a historical perspective, in: Adelaar, K.A., Himmelman, N.P. (Eds.), The Austronesian Languages of Asia and Madagascar. Routledge, New York, USA, pp. 1–42.

Adelaar, K.A., 1995a. Borneo as a cross-roads for comparative Austronesian linguistics, in: The Austronesians: Historical and Comparative Perspectives. ANU E Press, Canberra, Australia, pp. 81–102.

Adelaar, K.A., 1995b. L'importance du samihim (Bornéo du Sud) pour l'étymologie malgache, in: Champion, B. (Ed.), L'étrange intime. Mélanges Offerts à Paul Ottino. Université de la Réunion, France.

Adelaar, K.A., 1989. Malay influence on Malagasy: linguistic and culture-historical implications. Ocean. Linguist. 28, 1–46. doi:10.2307/3622973

Aghakhanian, F., Yunus, Y., Naidu, R., Jinam, T., Manica, A., Hoh, B.P., Phipps, M.E., 2015. Unravelling the genetic history of Negritos and indigenous populations of Southeast Asia. Genome Biol. Evol. 7, 1206–1215. doi:10.1093/gbe/evv065

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664. doi:10.1101/gr.094052.109

Allen, J., O'Connell, J., 2014. Both half right: Updating the evidence for dating first human arrivals in Sahul. Aust. Archaeol. 79, 86–108. doi:10.1080/03122417.2014.11682025

Allen, S.J., O'Donnell, A., Alexander, N.D., Mgone, C.S., Peto, T.E., Clegg, J.B., Alpers, M.P., Weatherall, D.J., 1999. Prevention of cerebral malaria in children in Papua New Guinea by southeast Asian ovalocytosis band 3. Am. J. Trop. Med. Hyg. 60, 1056–1060.

Anderson, A., 2005. Crossing the Luzon Strait: archaeological chronology in the Batanes Islands, Philippines and the regional sequence of Neolithic dispersal. J. Austronesian Stud. 1, 25-45

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R., Young, I.G., 1981. Sequence and organization of the human mitochondrial genome. Nature 290, 457–465. doi:10.1038/290457a0

Aubert, M., Brumm, A., Ramli, M., Sutikna, T., Saptomo, E.W., Hakim, B., Morwood, M.J., van den Bergh, G.D., Kinsley, L., Dosseto, A., 2014. Pleistocene cave art from Sulawesi, Indonesia. Nature 514, 223–227. doi:10.1038/nature13422

Barker, G. (Ed.), 2013. Rainforest Foraging and Farming in Island Southeast Asia: the Archaeology of the Niah Caves, Sarawak. McDonald Institute for Archaeological Research, Cambridge, UK.

Barker, G., Barton, H., Bird, M., Daly, P., Datan, I., Dykes, A., Farr, L., Gilbertson, D., Harrisson, B., Hunt, C., Higham, T., Kealhofer, L., Krigbaum, J., Lewis, H., McLaren, S., Paz, V., Pike, A., Piper, P., Pyatt, B., Rabett, R., Reynolds, T., Rose, J., Rushworth, G., Stephens, M., Stringer, C., Thompson, J., Turney, C., 2007. The "human revolution" in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo). J. Hum. Evol. 52, 243–261. doi:10.1016/j.jhevol.2006.08.011

Barton, H., Barker, G., Gilbertson, D., Hunt, C., Kealhofer, L., Lewis, H., Paz, V., Piper, P., Rabett, R., Reynolds, T., Szabo, K., 2013. Late Pleistocene foragers, c. 35,000 - 11,500 years ago, in: Barker, G. (Ed.), Rainforest Foraging and Farming in Island Southeast Asia. McDonald Institute for Archaeological Research, Cambridge, UK.

Battaglia, V., Grugni, V., Perego, U.A., Angerhofer, N., Gomez-Palmieri, J.E., Woodward, S.R., Achilli, A., Myres, N., Torroni, A., Semino, O., 2013. The first peopling of South America: New evidence from Y-chromosome haplogroup Q. PLOS ONE 8, e71390. doi:10.1371/journal.pone.0071390

Beaujard, P., 2012a. Les Mondes de l'Océan Indien. Vol. 1 : De la Formation de l'État au Premier Système-Monde Afro-Erasien (4e millénaire av. J.-C.-6e siècle apr. J.-C.). Armand Collin, Paris, France.

Beaujard, P., 2012b. Les Mondes de l'Ocean Indien. Vol. 2 : L'Océan Indien, au Cœur des Globalisations de l'Ancien Monde (7e-15e siècles). Armand Collin, Paris, France.

Beaujard, P., Norel, P., 2014. L'océan Indien avant le XVIe siècle : une charnière économique de l'ancien monde, in: L'Asie, La Mer, Le Monde. Au Temps Des Compagnies Des Indes. Presses Universitaires de Rennes, Rennes, France, pp. 21–41.

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian Computation in population genetics. Genetics 162, 2025–2035.

Bellwood, P., 2013. First Migrants: Ancient Migration in Global Perspective. Wiley-Blackwell, West Sussex, UK.

Bellwood, P., 2007. Prehistory of the Indo-Malaysian Archipelago. ANU E Press, Canberra.

Bellwood, P., Fox, J.J., Tryon, D., 1995. The Austronesians in history: Common origins and diverse transformations, in: The Austronesians: Historical and Comparative Perspectives. ANU E Press, Canberra, Australia, pp. 1–16.

Betsem, E., Cassar, O., Afonso, P.V., Fontanet, A., Froment, A., Gessain, A., 2014. Epidemiology and genetic variability of HHV-8/KSHV in Pygmy and Bantu populations in Cameroon. PLoS Negl. Trop. Dis. 8, e2851. doi:10.1371/journal.pntd.0002851

Bird, M., Taylor, D., Hunt, C., 2005. Palaeoenvironments of insular Southeast Asia during the Last Glacial Period: A savanna corridor in Sundaland? Quat. Sci. Rev. 24, 2228–2242.

Blench, R., 2014. Using diverse sources of evidence for reconstructing the past history of musical exchanges in the Indian Ocean. Afr. Archaeol. Rev. 31, 675–703. doi:10.1007/s10437-014-9178-z

Blench, R., 2011. Was there an Austroasiatic presence in Island Southeast Asia prior to the Austronesian expansion? Bull. Indo-Pac. Prehistory Assoc. 30, 133–144.

Blench, R., 2007. New palaeozoogeographical evidence for the settlement of Madagascar. Azania Archaeol. Res. Afr. 42, 69–82. doi:10.1080/00672700709480451

Blench, R., 1996. The ethnographic evidence for long-distance contacts between Oceania and East Africa, in: Reade, J. (Ed.), The Indian Ocean in Antiquity. Kegan Paul/British Museum, London/ New York, pp. 417–438.

Blust, R., 2013. The Austronesian languages, Revised edition. Asia Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University, Canberra, Australia.

Blust, R., 2007. The linguistic position of Sama-Bajaw. Stud. Philipp. Lang. Cult. 15, 73–114.

Blust, R., 1978. Eastern Malayo-Polynesian: a subgrouping argument, in: Wurm, S.A., Carrington, L. (Eds.), Fascicle 1, Pacific Linguistics, Series C. Presented at the Second international conference on Austronesian linguistics, Department of Linguistics, Research School of Pacific Studies, The Australian National University, Canberra, Australia, pp. 181–234.

Boivin, N., Crowther, A., Helm, R., Fuller, D.Q., 2013. East Africa and Madagascar in the Indian Ocean world. J. World Prehistory 26, 213–281.

Brandão, A., Eng, K.K., Rito, T., Cavadas, B., Bulbeck, D., Gandini, F., Pala, M., Mormina, M., Hudson, B., White, J., Ko, T.-M., Saidin, M., Zafarina, Z., Oppenheimer, S., Richards, M.B., Pereira, L., Soares, P., 2016. Quantifying the legacy of the Chinese Neolithic on the maternal genetic heritage of Taiwan and Island Southeast Asia. Hum. Genet. 135, 363–376. doi:10.1007/s00439-016-1640-3

Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., Reynolds, A., Ostrer, H., Mezey, J.G., Bustamante, C.D., 2012. PCAdmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. Hum. Biol. 84, 343–364. doi:10.3378/027.084.0401

Broushaki, F., Thomas, M.G., Link, V., López, S., Dorp, L. van, Kirsanow, K., Hofmanová, Z., Diekmann, Y., Cassidy, L.M., Díez-del-Molino, D., Kousathanas, A., Sell, C., Robson, H.K., Martiniano, R., Blöcher, J., Scheu, A., Kreutzer, S., Bollongino, R., Bobo, D., Davoudi, H., Munoz, O., Currat, M., Abdi, K., Biglari, F., Craig, O.E., Bradley, D.G., Shennan, S., Veeramah, K.R., Mashkour, M., Wegmann, D., Hellenthal, G., Burger, J., 2016. Early Neolithic genomes from the eastern Fertile Crescent. Science 353, 499–503. doi:10.1126/science.aaf7943

Bulbeck, D., 2013. Craniodental affinities of Southeast Asia's "Negritos" and the concordance with their genetic affinities. Hum. Biol. 85.

Bulbeck, D., 2008. An integrated perspective on the Austronesian diaspora: The switch from cereal agriculture to maritime foraging in the colonisation of Island Southeast Asia. Aust. Archaeol. 67, 31–52.

Burney, D.A., Burney, L.P., Godfrey, L.R., Jungers, W.L., Goodman, S.M., Wright, H.T., Jull, A.J.T., 2004. A chronology for late prehistoric Madagascar. J. Hum. Evol. 47, 25–63. doi:10.1016/j.jhevol.2004.05.005

Busby, G.B., Band, G., Le, Q.S., Jallow, M., Bougama, E., Mangano, V.D., Amenga-Etego, L.N., Enimil, A., Apinjoh, T., Ndila, C.M., Manjurano, A., Nyirongo, V., Doumba, O., Rockett, K.A., Kwiatkowski, D.P., Spencer, C.C., Network, M.G.E., 2016. Admixture into and within sub-Saharan Africa. eLife 5, e15266. doi:10.7554/eLife.15266

Cann, R.L., Stoneking, M., Wilson, A.C., 1987. Mitochondrial DNA and human evolution. Nature 325, 31–36. doi:10.1038/325031a0

Capredon, M., 2011. Histoire biologique d'une population du sud-est malgache : les Antemoro (Ph.D. Thesis). Université de La Réunion, La Réunion.

Cassar, O., Afonso, P.V., Bassot, S., Plancoulaine, S., Duprez, R., Capuano, C., Abel, M., Martin, P.M.V., Gessain, A., 2007. Novel Human Herpesvirus 8 Subtype D Strains in Vanuatu, Melanesia. Emerg. Infect. Dis. 13, 1745–1748. doi:10.3201/eid1311.070636

Cassar, O., Charavay, F., Bassot, S., Plancoulaine, S., Grangeon, J.-P., Laumond-Barny, S., Martin, P.M.V., Chanteau, S., Gessain, A., 2012. Divergent KSHV/HHV-8 subtype D strains in New Caledonia and Solomon Islands, Melanesia. J. Clin. Virol. 53, 214–218. doi:10.1016/j.jcv.2011.12.016

Cassar, O., Charavay, F., Touzain, F., Jeannin, P., Grangeon, J.-P., Laumond, S., Chungue, E., Martin, P.M.V., Gessain, A., 2017. A novel human T-lymphotropic virus type 1c molecular variant in an indigenous individual from New Caledonia, Melanesia. PLoS Negl. Trop. Dis. 11, e0005278. doi:10.1371/journal.pntd.0005278

Cassar, O., Einsiedel, L., Afonso, P.V., Gessain, A., 2013. Human T-Cell lymphotropic virus type 1 subtype C molecular variants among indigenous Australians: New insights into the molecular epidemiology of HTLV-1 in Australo-Melanesia. PLoS Negl. Trop. Dis. 7, e2418. doi:10.1371/journal.pntd.0002418

Charlesworth, B., 1996. The evolution of chromosomal sex determination and dosage compensation. Curr. Biol. 6, 149–162. doi:10.1016/S0960-9822(02)00448-7

Charlesworth, B., Charlesworth, D., 2000. The degeneration of Y chromosomes. Philos. Trans. R. Soc. B Biol. Sci. 355, 1563–1572. doi:10.1098/rstb.2000.0717

Clayton, D., 2003. Mitochondrial DNA Replication: What We Know. IUBMB Life 55, 213–217. doi:10.1080/1521654031000134824

International HapMap Consortium, 2010. Integrating common and rare genetic variation in diverse human populations. Nature 467, 52–58. doi:10.1038/nature09298

Cox, M.P., Hudjashov, G., Sim, A., Savina, O., Karafet, T.M., Sudoyo, H., Lansing, J.S., 2016. Small traditional human communities sustain genomic diversity over microgeographic scales despite linguistic isolation. Mol. Biol. Evol. 33, 2273–2284. doi:10.1093/molbev/msw099

Cox, M.P., Karafet, T.M., Lansing, J.S., Sudoyo, H., Hammer, M.F., 2010. Autosomal and X-linked single nucleotide polymorphisms reveal a steep Asian–Melanesian ancestry cline in eastern Indonesia and a sex bias in admixture rates. Proc. R. Soc. B Biol. Sci. 277, 1589–1596. doi:10.1098/rspb.2009.2041

Cox, M.P., Nelson, M.G., Tumonggor, M.K., Ricaut, F.-X., Sudoyo, H., 2012. A small cohort of Island Southeast Asian women founded Madagascar. Proc. R. Soc. B Biol. Sci. 279, 2761–2768. doi:10.1098/rspb.2012.0012

Crawford, G.W., Shen, C., 1998. The origins of rice agriculture: recent progress in East Asia. Antiquity 72, 858–866. doi:10.1017/S0003598X00087494

Crowther, A., Lucas, L., Helm, R., Horton, M., Shipton, C., Wright, H.T., Walshaw, S., Pawlowicz, M., Radimilahy, C., Douka, K., Picornell-Gelabert, L., Fuller, D.Q., Boivin, N.L., 2016. Ancient crops provide first archaeological signature of the westward Austronesian expansion. Proc. Natl. Acad. Sci. 113, 6635–6640. doi:10.1073/pnas.1522714113

Cruciani, F., Trombetta, B., Massaia, A., Destro-Bisol, G., Sellitto, D., Scozzari, R., 2011. A revised root for the human Y chromosomal phylogenetic tree: The origin of patrilineal diversity in Africa. Am. J. Hum. Genet. 88, 814–818. doi:10.1016/j.ajhg.2011.05.002

Csilléry, K., François, O., Blum, M.G.B., 2012. abc: an R package for approximate Bayesian computation (ABC). Methods Ecol. Evol. 3, 475–479. doi:10.1111/j.2041-210X.2011.00179.x

Curnoe, D., Datan, I., Taçon, P.S.C., Leh Moi Ung, C., Sauffi, M.S., 2016. Deep skull from Niah Cave and the Pleistocene peopling of Southeast Asia. Front. Ecol. Evol. 4. doi:10.3389/fevo.2016.00075

Dahl, O.C., 1951. Malgache et Maanjan: une comparaison linguistique. Edege-Intituttet, Oslo, Norway.

Deng, L., Hoh, B.-P., Lu, D., Saw, W.-Y., Twee-Hee Ong, R., Kasturiratne, A., Janaka de Silva, H., Zilfalil, B.A., Kato, N., Wickremasinghe, A.R., Teo, Y.-Y., Xu, S., 2015. Dissecting the genetic structure and admixture of four geographical Malay populations. Sci. Rep. 5, 14375. doi:10.1038/srep14375

Deng, Z., Qin, L., Gao, Y., Weisskopf, A.R., Zhang, C., Fuller, D.Q., 2015. From early domesticated rice of the middle Yangtze Basin to millet, rice and wheat agriculture: Archaeobotanical macro-remains from Baligang, Nanyang Basin, Central China (6700–500 BC). PLOS ONE 10, e0139885. doi:10.1371/journal.pone.0139885

Détroit, F., Dizon, E., Falguères, C., Hameau, S., Ronquillo, W., Sémah, F., 2004. Upper pleistocene Homo sapiens from the Tabon cave (Palawan, The Philippines): Description and dating of new discoveries. Comptes Rendus Palevol 3, 705–712. doi:10.1016/j.crpv.2004.06.004

Dewar, R.E., Wright, H.T., 1993. The culture history of Madagascar. J. World Prehistory 7, 417–466. doi:10.1007/BF00997802

Diamond, J., Bellwood, P., 2003. Farmers and their languages: The first expansions. Science 300, 597–603. doi:10.1126/science.1078208

Duggan, A.T., Evans, B., Friedlaender, F.R., Friedlaender, J.S., Koki, G., Merriwether, D.A., Kayser, M., Stoneking, M., 2014. Maternal history of Oceania from complete mtDNA genomes: Contrasting ancient diversity with recent homogenization due to the Austronesian expansion. Am. J. Hum. Genet. 94, 721–733. doi:10.1016/j.ajhg.2014.03.014

Duggan, A.T., Stoneking, M., 2014. Recent developments in the genetic history of East Asia and Oceania. Curr. Opin. Genet. Dev., Genetics of human evolution 29, 9–14. doi:10.1016/j.gde.2014.06.010

Ferrand, G., 1908. L'Origine Africaine des Malgaches. Impr. nationale, Paris.

Field, J.S., Petraglia, M.D., Lahr, M.M., 2007. The southern dispersal hypothesis and the South Asian archaeological record: Examination of dispersal routes through GIS analysis. J. Anthropol. Archaeol. 26, 88–108. doi:10.1016/j.jaa.2006.06.001

Fitzpatrick, S.M., Callaghan, R., 2008. Seafaring simulations and the origin of prehistoric settlers to Madagascar, in: Clark, G., Leach, F., O'Connor, S. (Eds.), Islands of Inquiry (Terra Australis 29). The Australian National University, Canberra, Australia.

Fuller, D.Q., 2011. Pathways to Asian civilizations: Tracing the origins and spread of rice and rice cultures. Rice 4, 78–92. doi:10.1007/s12284-011-9078-7

Gilbert, S.F., Singer, S., 2006. Developmental Biology, Eighth Edition, 8th edition. ed. Sinauer Associates Inc., Sunderland, Mass.

Gommery, D., Ramanivosoa, B., Faure, M., Guérin, C., Kerloc'h, P., Sénégas, F., Randrianantenaina, H., 2011. Les plus anciennes traces d'activités anthropiques de Madagascar sur des ossements d'hippopotames subfossiles d'Anjohibe (Province de Mahajanga). Comptes Rendus Palevol 10, 271–278. doi:10.1016/j.crpv.2011.01.006

Gonda, J., 1975. The Indian Religions in Pre-Islamic Indonesia and Their Survival in Bali, in: Kähler, H. (Ed.), Handbook of Oriental Studies. E.J. Brill, Leiden, Netherlands.

Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. Nat. Rev. Genet. 17, 333–351. doi:10.1038/nrg.2016.49

Gray, R.D., Jordan, F.M., 2000. Language trees support the express-train sequence of Austronesian expansion. Nature 405, 1052–1055. doi:10.1038/35016575

Greenhill, S.J., 2015. TransNewGuinea.org: An online database of New Guinea languages. PLOS ONE 10, e0141563. doi:10.1371/journal.pone.0141563

Grenet, M., Sarel, J., Fauzy, R., Oktaviana, A.A., Sugiyanto, B., Chazine, J.-M., Ricaut, F.-X., 2016. New insights on the late Pleistocene–Holocene lithic industry in East Kalimantan (Borneo): The contribution of three rock shelter sites in the karstic area of the Mangkalihat peninsula. Quat. Int., Southeast Asia: human evolution, dispersals and adaptation 416, 126–150. doi:10.1016/j.quaint.2015.11.044

Hallast, P., Batini, C., Zadik, D., Delser, P.M., Wetton, J.H., Arroyo-Pardo, E., Cavalleri, G.L., Knijff, P.D., Bisol, G.D., Dupuy, B.M., Eriksen, H.A., Jorde, L.B., King, T.E., Larmuseau, M.H., Munain, A.L.D., López-Parra, A.M., Loutradis, A., Milasin, J., Novelletto, A., Pamjav, H., Sajantila, A., Schempp, W., Sears, M., Tolun, A., Tyler-Smith, C., Geystelen, A.V., Watkins, S., Winney, B., Jobling, M.A., Ramachandran, S., 2014. The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. Mol. Biol. Evol. 32, 661–673. doi:10.1093/molbev/msu327

Hammer, M.F., Karafet, T., Rasanayagam, A., Wood, E.T., Altheide, T.K., Jenkins, T., Griffiths, R.C., Templeton, A.R., Zegura, S.L., 1998. Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol. Biol. Evol. 15, 427–441. doi:10.1093/oxfordjournals.molbev.a025939

Harrison, T., 1957. The Great Cave of Niah: A preliminary report on Borneo prehistory. Man 57, 161–166.

Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., Myers, S., 2014. A genetic atlas of human admixture history. Science 343, 747–751. doi:10.1126/science.1243518

Henn, B.M., Gignoux, C.R., Jobin, M., Granka, J.M., Macpherson, J.M., Kidd, J.M., Rodríguez-Botigué, L., Ramachandran, S., Hon, L., Brisbin, A., Lin, A.A., Underhill, P.A., Comas, D., Kidd, K.K., Norman, P.J., Parham, P., Bustamante, C.D., Mountain, J.L., Feldman, M.W., 2011. Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc. Natl. Acad. Sci. 108, 5154–5162. doi:10.1073/pnas.1017511108

Higham, C., 2014. Early mainland Southeast Asia: From first humans to Angkor. River Books, Bangkok, Thailand.

Higham, C., 2013. Hunter-gatherers in Southeast Asia: From prehistory to the present. Hum. Biol. 85, 21–44. doi:10.3378/027.085.0302

Higham, C.F.W., Douka, K., Higham, T.F.G., 2015. A new chronology for the Bronze Age of Northeastern Thailand and its implications for Southeast Asian prehistory. PLOS ONE 10, e0137542. doi:10.1371/journal.pone.0137542

Hill, C., Soares, P., Mormina, M., Macaulay, V., Clarke, D., Blumbach, P.B., Vizuete-Forster, M., Forster, P., Bulbeck, D., Oppenheimer, S., Richards, M., 2007. A mitochondrial stratigraphy for island Southeast Asia. Am. J. Hum. Genet. 80, 29–43. doi:10.1086/510412

Hudson, A.B., 1972. Padju Epat: The Maʼanyan of Indonesian Borneo. Irvington Publishers, New York, USA.

HUGO Pan-Asian SNP Consortium, Abdulla, M.A., Ahmed, I., Assawamakin, A., Bhak, J., Brahmachari, S.K., Calacal, G.C., Chaurasia, A., Chen, C.-H., Chen, J., Chen, Y.-T., Chu, J., Cutiongco-de la Paz, E.M.C., De Ungria, M.C.A., Delfin, F.C., Edo, J., Fuchareon, S., Ghang, H., Gojobori, T., Han, J., Ho, S.-F., Hoh, B.P., Huang, W., Inoko, H., Jha, P., Jinam, T.A., Jin, L., Jung, J., Kangwanpong, D., Kampuansai, J., Kennedy, G.C., Khurana, P., Kim, H.-L., Kim, K., Kim, S., Kim, W.-Y., Kimm, K., Kimura, R., Koike, T., Kulawonganunchai, S., Kumar, V., Lai, P.S., Lee, J.-Y., Lee, S., Liu, E.T., Majumder, P.P., Mandapati, K.K., Marzuki, S., Mitchell, W., Mukerji, M., Naritomi, K., Ngamphiw, C., Niikawa, N., Nishida, N., Oh, B., Oh, S., Ohashi, J., Oka, A., Ong, R., Padilla, C.D., Palittapongarnpim, P., Perdigon, H.B., Phipps, M.E., Png, E., Sakaki, Y., Salvador, J.M., Sandraling, Y., Scaria, V., Seielstad, M., Sidek, M.R., Sinha, A., Srikummool, M., Sudoyo, H., Sugano, S., Suryadi, H., Suzuki, Y., Tabbada, K.A., Tan, A., Tokunaga, K., Tongsima, S., Villamor, L.P., Wang, E., Wang, Y., Wang, H., Wu, J.-Y., Xiao, H., Xu, S., Yang, J.O., Shugart, Y.Y., Yoo, H.-S., Yuan, W., Zhao, G., Zilfalil, B.A., Indian Genome Variation Consortium, 2009. Mapping human genetic diversity in Asia. Science 326, 1541–1545. doi:10.1126/science.1177074

Hurles, M.E., Sykes, B.C., Jobling, M.A., Forster, P., 2005. The dual origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from maternal and paternal lineages. Am. J. Hum. Genet. 76, 894–901. doi:10.1086/430051

Jinam, T.A., Hong, L.-C., Phipps, M.E., Stoneking, M., Ameen, M., Edo, J., Consortium, H.P.-A.S., Saitou, N., 2012. Evolutionary history of continental Southeast Asians: “Early Train” hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. Mol. Biol. Evol. 29, 3513–3527. doi:10.1093/molbev/mss169

Jobling, M.A., Tyler-Smith, C., 2003. The human Y chromosome: an evolutionary marker comes of age. Nat. Rev. Genet. 4, 598–612. doi:10.1038/nrg1124

Jobling, M.A., Tyler-Smith, C., 1995. Fathers and sons: the Y chromosome and human evolution. Trends Genet. 11, 449–456. doi:10.1016/S0168-9525(00)89144-1

Kanchan, R.K., 1990. Hindu Kingdoms of South-East Asia, Indian Historical Researches ; Cosmo Publications, New Delhi, India.

Kao, J.-H., Chen, P.-J., Lai, M.-Y., Chen, D.-S., 2004. Hepatitis B virus genotypes and spontaneous hepatitis B e antigen seroconversion in Taiwanese hepatitis B carriers. J. Med. Virol. 72, 363–369. doi:10.1002/jmv.10534

Karafet, T.M., Hallmark, B., Cox, M.P., Sudoyo, H., Downey, S., Lansing, J.S., Hammer, M.F., 2010. Major east–west division underlies Y chromosome stratification across Indonesia. Mol. Biol. Evol. 27, 1833–1844. doi:10.1093/molbev/msq063

Karmin, M., Saag, L., Vicente, M., Sayres, M.A.W., Järve, M., Talas, U.G., Rootsi, S., Ilumäe, A.-M., Mägi, R., Mitt, M., Pagani, L., Puurand, T., Faltyskova, Z., Clemente, F., Cardona, A., Metspalu, E., Sahakyan, H., Yunusbayev, B., Hudjashov, G., DeGiorgio, M., Loogväli, E.-L., Eichstaedt, C., Eelmets, M., Chaubey, G., Tambets, K., Litvinov, S., Mormina, M., Xue, Y., Ayub, Q., Zoraqi, G., Korneliussen, T.S., Akhatova, F., Lachance, J., Tishkoff, S., Momynaliev, K., Ricaut, F.-X., Kusuma, P., Razafindrazaka, H., Pierron, D., Cox, M.P., Sultana, G.N.N., Willerslev, R., Muller, C., Westaway, M., Lambert, D., Skaro, V., Kovačević, L., Turdikulova, S., Dalimova, D., Khusainova, R., Trofimova, N., Akhmetova, V., Khidiyatova, I., Lichman, D.V., Isakova, J., Pocheshkhova, E., Sabitov, Z., Barashkov, N.A., Nymadawa, P., Mihailov, E., Seng, J.W.T., Evseeva, I., Migliano, A.B., Abdullah, S., Andriadze, G., Primorac, D., Atramentova, L., Utevska, O., Yepiskoposyan, L., Marjanović, D., Kushniarevich, A., Behar, D.M., Gilissen, C., Vissers, L., Veltman, J.A., Balanovska, E., Derenko, M., Malyarchuk, B., Metspalu, A., Fedorova, S., Eriksson, A., Manica, A., Mendez, F.L., Karafet, T.M., Veeramah, K.R., Bradman, N., Hammer, M.F., Osipova, L.P., Balanovsky, O., Khusnutdinova, E.K., Johnsen, K., Remm, M., Thomas, M.G., Tyler-Smith, C., Underhill, P.A., Willerslev, E., Nielsen, R., Metspalu, M., Villems, R., Kivisild, T., 2015. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. Genome Res. doi:10.1101/gr.186684.114

Kayser, M., Brauer, S., Weiss, G., Schiefenhövel, W., Underhill, P., Shen, P., Oefner, P., Tommaseo-Ponzetta, M., Stoneking, M., 2003. Reduced Y-chromosome, but not Mitochondrial DNA, diversity in human populations from West New Guinea. Am. J. Hum. Genet. 72, 281–302. doi:10.1086/346065

Kayser, M., Lao, O., Saar, K., Brauer, S., Wang, X., Nürnberg, P., Trent, R.J., Stoneking, M., 2008. Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. Am. J. Hum. Genet. 82, 194–198. doi:10.1016/j.ajhg.2007.09.010

Kivisild, T., 2017. The study of human Y chromosome variation through ancient DNA. Hum. Genet. 136, 529–546. doi:10.1007/s00439-017-1773-z

Kivisild, T., 2015. Maternal ancestry and population history from whole mitochondrial genomes. Investig. Genet. 6, 3. doi:10.1186/s13323-015-0022-2

Kraytsberg, Y., Schwartz, M., Brown, T.A., Ebralidse, K., Kunz, W.S., Clayton, D.A., Vissing, J., Khrapko, K., 2004. Recombination of human mitochondrial DNA. Science 304, 981–981. doi:10.1126/science.1096342

Krigbaum, J., Manser, J., 2005. The West Mouth burial series from Niah Cave: past and present, in: Majid, Z. (Ed.), The Perak Man and Other Prehistoric Skeletons of Malaysia. Universiti Sains Malaysia, Penang, Malaysia, pp. 175–206.

Lambek, M., 1998. People of the sea: Identity and descent among the Vezo of Madagascar. Am. Anthropol. 100, 203–204. doi:10.1525/aa.1998.100.1.203.2

Lawson, D.J., Hellenthal, G., Myers, S., Falush, D., 2012. Inference of population structure using dense haplotype data. PLOS Genet. 8, e1002453. doi:10.1371/journal.pgen.1002453

Lewis, M.P., Simons, G.F., Fennig, C.D. (Eds.), 2017. Ethnologue: Languages of the world, Nineteenth edition. ed. SIL International, Texas.

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., Myers, R.M., 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319, 1100–1104. doi:10.1126/science.1153717

Lipson, M., Loh, P.-R., Patterson, N., Moorjani, P., Ko, Y.-C., Stoneking, M., Berger, B., Reich, D., 2014. Reconstructing Austronesian population history in Island Southeast Asia. Nat. Commun. 5. doi:10.1038/ncomms5689

Liu, S.C., Jarolim, P., Rubin, H.L., Palek, J., Amato, D., Hassan, K., Zaik, M., Sapak, P., 1994. The homozygous state for the band 3 protein mutation in Southeast Asian Ovalocytosis may be lethal. Blood 84, 3590–3591.

Lusida, M.I., Nugrahaputra, V.E., Soetjipto, Handajani, R., Nagano-Fujii, M., Sasayama, M., Utsumi, T., Hotta, H., 2008. Novel subgenotypes of Hepatitis B Virus genotypes C and D in Papua, Indonesia. J. Clin. Microbiol. 46, 2160–2166. doi:10.1128/JCM.01681-07

Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., Taha, A., Shaari, N.K., Raja, J.M., Ismail, P., Zainuddin, Z., Goodwin, W., Bulbeck, D., Bandelt, H.-J., Oppenheimer, S., Torroni, A., Richards, M., 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science 308, 1034–1036. doi:10.1126/science.1109792

MacPhee, R.D.E., Burney, D.A., 1991. Dating of modified femora of extinct dwarf Hippopotamus from Southern Madagascar: Implications for constraining human colonization and vertebrate extinction events. J. Archaeol. Sci. 18, 695–706. doi:10.1016/0305-4403(91)90030-S

Mahdi, W., 2016. Origins of Southeast Asian shipping and maritime communication across the Indian Ocean, in: Campbell, G. (Ed.), Early Exchange between Africa and the Wider Indian Ocean World, Palgrave Series in Indian Ocean World Studies. Springer International Publishing, pp. 25–49. doi:10.1007/978-3-319-33822-4_2

Malaspinas, A.-S., Westaway, M.C., Muller, C., Sousa, V.C., Lao, O., Alves, I., Bergström, A., Athanasiadis, G., Cheng, J.Y., Crawford, J.E., Heupink, T.H., Macholdt, E., Peischl, S., Rasmussen, S., Schiffels, S., Subramanian, S., Wright, J.L., Albrechtsen, A., Barbieri, C., Dupanloup, I., Eriksson, A., Margaryan, A., Moltke, I., Pugach, I., Korneliussen, T.S., Levkivskyi, I.P., Moreno-Mayar, J.V., Ni, S., Racimo, F., Sikora, M., Xue, Y., Aghakhanian, F.A., Brucato, N., Brunak, S., Campos, P.F., Clark, W., Ellingvåg, S., Fourmile, G., Gerbault, P., Injie, D., Koki, G., Leavesley, M., Logan, B., Lynch, A., Matisoo-Smith, E.A., McAllister, P.J., Mentzer, A.J., Metspalu, M., Migliano, A.B., Murgha, L., Phipps, M.E., Pomat, W., Reynolds, D., Ricaut, F.-X., Siba, P., Thomas, M.G., Wales, T., Wall, C.M., Oppenheimer, S.J., Tyler-Smith, C., Durbin, R., Dortch, J., Manica, A., Schierup, M.H., Foley, R.A., Lahr, M.M., Bowern, C., Wall, J.D., Mailund, T., Stoneking, M., Nielsen, R., Sandhu, M.S., Excoffier, L., Lambert, D.M., Willerslev, E., 2016. A genomic history of Aboriginal Australia. Nature 538, 207–214. doi:10.1038/nature18299

Maricic, T., Whitten, M., Pääbo, S., 2010. Multiplexed DNA Sequence capture of mitochondrial genomes using PCR products. PLOS ONE 5, e14004. doi:10.1371/journal.pone.0014004

McFarland, R., Taylor, R.W., Turnbull, D.M., 2010. A neurological perspective on mitochondrial disease. Lancet Neurol. 9, 829–840. doi:10.1016/S1474-4422(10)70116-2

McLeod, N., 1977. Musical instruments and history in Madagascar, in: Essays for a Humanist. An Offering to Klaus Wachsmann. The Town House Press, New York, USA.

Mellars, P., Gori, K.C., Carr, M., Soares, P.A., Richards, M.B., 2013. Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. Proc. Natl. Acad. Sci. 110, 10699–10704. doi:10.1073/pnas.1306043110

Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. 2010, pdb.prot5448. doi:10.1101/pdb.prot5448

Migliano, A.B., Romero, I.G., Metspalu, M., Leavesley, M., Pagani, L., Antao, T., Huang, D.-W., Sherman, B.T., Siddle, K., Scholes, C., Hudjashov, G., Kaitokai, E., Babalu, A., Belatti, M., Cagan, A., Hopkinshaw, B., Shaw, C., Nelis, M., Metspalu, E., Mägi, R., Lempicki, R.A., Villems, R., Lahr, M.M., Kivisild, T., 2013. Evolution of the Pygmy phenotype: Evidence of positive selection from genome-wide scans in African, Asian, and Melanesian pygmies. Hum. Biol. 85, 251–284. doi:10.13110/humanbiology.85.1-3.0251

Mijares, A.S., Détroit, F., Piper, P., Grün, R., Bellwood, P., Aubert, M., Champion, G., Cuevas, N., De Leon, A., Dizon, E., 2010. New evidence for a 67,000-year-old human presence at Callao Cave, Luzon, Philippines. J. Hum. Evol. 59, 123–132. doi:10.1016/j.jhevol.2010.04.008

Mohandas, N., Lie-Injo, L.E., Friedman, M., Mak, J.W., 1984. Rigid membranes of Malayan ovalocytes: a likely genetic barrier against malaria. Blood 63, 1385–1392.

Moodley, Y., Linz, B., Yamaoka, Y., Windsor, H.M., Breurec, S., Wu, J.-Y., Maady, A., Bernhöft, S., Thiberge, J.-M., Phuanukoonnon, S., Jobb, G., Siba, P., Graham, D.Y., Marshall, B.J., Achtman, M., 2009. The Peopling of the Pacific from a bacterial perspective. Science 323, 527–530. doi:10.1126/science.1166083

Moorjani, P., Thangaraj, K., Patterson, N., Lipson, M., Loh, P.-R., Govindaraj, P., Berger, B., Reich, D., Singh, L., 2013. Genetic evidence for recent population mixture in India. Am. J. Hum. Genet. 93, 422–438. doi:10.1016/j.ajhg.2013.07.006

Mörseburg, A., Pagani, L., Ricaut, F.-X., Yngvadottir, B., Harney, E., Castillo, C., Hoogervorst, T., Antao, T., Kusuma, P., Brucato, N., Cardona, A., Pierron, D., Letellier, T., Wee, J., Abdullah, S., Metspalu, M., Kivisild, T., 2016. Multi-layered population structure in Island Southeast Asians. Eur. J. Hum. Genet. 24, 1605–1611. doi:10.1038/ejhg.2016.60

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R., Auton, A., Indap, A., King, K.S., Bergmann, S., Nelson, M.R., Stephens, M., Bustamante, C.D., 2008. Genes mirror geography within Europe. Nature 456, 98–101. doi:10.1038/nature07331

Nuraini, C., 2016. The intangible legacy of the Indonesian Bajo. Wacana 17, 1–18. doi:10.17510/wacana.v17i1.427

O'donnell, Allen, Mgone, Martinson, Clegg, Weatherall, 1998. Red cell morphology and malaria anaemia in children with Southeast-Asian ovalocytosis band 3 in Papua New Guinea. Br. J. Haematol. 101, 407–412. doi:10.1046/j.1365-2141.1998.00742.x

Oppenheimer, S., Richards, M., 2001. Fast trains, slow boats, and the ancestry of the Polynesian islanders. Sci. Prog. 84, 157–181.

Pagani, L., Schiffels, S., Gurdasani, D., Danecek, P., Scally, A., Chen, Y., Xue, Y., Haber, M., Ekong, R., Oljira, T., Mekonnen, E., Luiselli, D., Bradman, N., Bekele, E., Zalloua, P., Durbin, R., Kivisild, T., Tyler-Smith, C., 2015. Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. Am. J. Hum. Genet. 96, 986–991. doi:10.1016/j.ajhg.2015.04.019

Pakendorf, B., Stoneking, M., 2005. Mitochondrial DNA and human evolution. Annu. Rev. Genomics Hum. Genet. 6, 165–183. doi:10.1146/annurev.genom.6.080604.162249

Paquette, A.M., Harahap, A., Laosombat, V., Patnode, J.M., Satyagraha, A., Sudoyo, H., Thompson, M.K., Yusoff, N.M., Wilder, J.A., 2015. The evolutionary origins of Southeast Asian Ovalocytosis. Infect. Genet. Evol. 34, 153–159. doi:10.1016/j.meegid.2015.06.002

Pasaniuc, B., Sankararaman, S., Torgerson, D.G., Gignoux, C., Zaitlen, N., Eng, C., Rodriguez-Cintron, W., Chapela, R., Ford, J.G., Avila, P.C., Rodriguez-Santana, J., Chen, G.K., Le Marchand, L., Henderson, B., Reich, D., Haiman, C.A., Gonzàlez Burchard, E., Halperin, E., 2013. Analysis of Latino populations from GALA and MEC studies reveals genomic loci with biased local ancestry estimation. Bioinformatics 29, 1407–1415. doi:10.1093/bioinformatics/btt166

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012. Ancient admixture in human history. Genetics 192, 1065–1093. doi:10.1534/genetics.112.145037

Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. PLoS Genet 2, e190. doi:10.1371/journal.pgen.0020190

Pawley, A., Ross, M., 1995. The prehistory of the oceanic languages: a current view, in: Fox, J.J., Tryon, D., Bellwood, P. (Eds.), The Austronesians: Historical and Comparative Perspective. ANU E Press, Canberra, Australia, pp. 43–80.

Paz, V., 2002. Island Southeast Asia: spread or friction zone?, in: Bellwood, P., Renfrew, C. (Eds.), Examining the Farming/Language Dispersal Hypothesis. McDonald Institute for Archaeological Research, Cambridge, UK, pp. 275–285.

Pierron, D., Razafindrazaka, H., Pagani, L., Ricaut, F.-X., Antao, T., Capredon, M., Sambo, C., Radimilahy, C., Rakotoarisoa, J.-A., Blench, R.M., Letellier, T., Kivisild, T., 2014. Genome-wide evidence of Austronesian–Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. Proc. Natl. Acad. Sci. 111, 936–941. doi:10.1073/pnas.1321860111

Pierron, D., Rocher, C., Amati-Bonneau, P., Reynier, P., Martin-Négrier, M.-L., Allouche, S., Batandier, C., de Camaret, B.M., Godinot, C., Rotig, A., Feldmann, D., Bellanne-Chantelot, C., Arveiler, B., Pennarun, E., Rossignol, R., Crouzet, M., Murail, P., Thoraval, D., Letellier, T., 2008. New evidence of a mitochondrial genetic background paradox: Impact of the J haplogroup on the A3243G mutation. BMC Med. Genet. 9, 41. doi:10.1186/1471-2350-9-41

Plutniak, S., Oktaviana, A.A., Sugiyanto, B., Chazine, J.M., Ricaut, F.X., 2014. New ceramic data from East Kalimantan: Pottery chronology and the cord-marked and red-slipped sherds of Liang Abu's layer 2. J. Pac. Archaeol. 5, 90–99.

Poznik, G.D., Henn, B.M., Yee, M.-C., Sliwerska, E., Euskirchen, G.M., Lin, A.A., Snyder, M., Quintana-Murci, L., Kidd, J.M., Underhill, P.A., Bustamante, C.D., 2013. Sequencing Y Chromosomes resolves discrepancy in time to common ancestor of males versus females. Science 341, 562–565. doi:10.1126/science.1237619

Price, A.L., Tandon, A., Patterson, N., Barnes, K.C., Rafaels, N., Ruczinski, I., Beaty, T.H., Mathias, R., Reich, D., Myers, S., 2009. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. PLOS Genet. 5, e1000519. doi:10.1371/journal.pgen.1000519

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Rabe, T., Jambou, R., Rabarijaona, L., Raharimalala, L., Rason, M.A., Ariey, F., Dhermy, D., 2002. South-East Asian ovalocytosis among the population of the Highlands of Madagascar: a vestigé of the island's settlement. Trans. R. Soc. Trop. Med. Hyg. 96, 143–144.

Ras, J.J., 1968. Hikajat Banjar: a study in Malay historiography. Martinus Nijhoff, The Hague.

Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., Albrechtsen, A., Skotte, L., Lindgreen, S., Metspalu, M., Jombart, T., Kivisild, T., Zhai, W., Eriksson, A., Manica, A., Orlando, L., Vega, F.M.D.L., Tridico, S., Metspalu, E., Nielsen, K., Ávila-Arcos, M.C., Moreno-Mayar, J.V., Muller, C., Dortch, J., Gilbert, M.T.P., Lund, O., Wesolowska, A., Karmin, M., Weinert, L.A., Wang, B., Li, J., Tai, S., Xiao, F., Hanihara, T., Driem, G. van, Jha, A.R., Ricaut, F.-X., Knijff, P. de, Migliano, A.B., Romero, I.G., Kristiansen, K., Lambert, D.M., Brunak, S., Forster, P., Brinkmann, B., Nehlich, O., Bunce, M., Richards, M., Gupta, R., Bustamante, C.D., Krogh, A., Foley, R.A., Lahr, M.M., Balloux, F., Sicheritz-Pontén, T., Villems, R., Nielsen, R., Wang, J., Willerslev, E., 2011. An Aboriginal Australian genome reveals separate human dispersals into Asia. Science 334, 94–98. doi:10.1126/science.1211177

Rasoloson, J., Rubino, C., 2005. Malagasy, in: Adelaar, K.A., Himmelman, N.P. (Eds.), The Austronesian Languages of Asia and Madagascar. Routledge, New York, USA, pp. 456–488.

Razafindrazaka, H., 2010. Le peuplement humain de Madagascar : Anthropologie génétique de trois groupes traditionnels (Ph.D. Thesis). Université Paul Sabatier - Tolouse III, Toulouse, France.

Razafindrazaka, H., Ricaut, F.-X., Cox, M.P., Mormina, M., Dugoujon, J.-M., Randriamarolaza, L.P., Guitard, E., Tonasso, L., Ludes, B., Crubezy, E., 2010. Complete mitochondrial DNA sequences provide new insights into the Polynesian motif and the peopling of Madagascar. Eur. J. Hum. Genet. 18, 575–581. doi:10.1038/ejhg.2009.222

Reich, D., Thangaraj, K., Patterson, N., Price, A.L., Singh, L., 2009. Reconstructing Indian population history. Nature 461, 489–494. doi:10.1038/nature08365

Robert Blust, 2010. The Greater North Borneo hypothesis. Ocean. Linguist. 49, 44–118. doi:10.1353/ol.0.0060

Rosenberg, N.A., Li, L.M., Ward, R., Pritchard, J.K., 2003. Informativeness of genetic markers for inference of ancestry. Am. J. Hum. Genet. 73, 1402–1422. doi:10.1086/380416

Sankararaman, S., Sridhar, S., Kimmel, G., Halperin, E., 2008. Estimating local ancestry in admixed populations. Am. J. Hum. Genet. 82, 290–303. doi:10.1016/j.ajhg.2007.09.022

Scozzari, R., Massaia, A., Trombetta, B., Bellusci, G., Myres, N.M., Novelletto, A., Cruciani, F., 2014. An unbiased resource of novel SNP markers provides a new chronology for the human Y chromosome and reveals a deep phylogenetic structure in Africa. Genome Res. gr.160788.113. doi:10.1101/gr.160788.113

Sellato, B., 1994. Nomads of the Borneo Rainforest: The Economics, Politics, and Ideology of Settling Down. University of Hawaii Press, Honolulu, USA.

Serva, M., Petroni, F., Volchenkov, D., Wichmann, S., 2012. Malagasy dialects and the peopling of Madagascar. J. R. Soc. Interface 9, 54–67. doi:10.1098/rsif.2011.0228

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S.-F., Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T., Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfing, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S.-P., Waterston, R.H., Wilson, R.K., Rozen, S., Page, D.C., 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423, 825–837. doi:10.1038/nature01722

Skoglund, P., Mallick, S., Bortolini, M.C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M.L., Salzano, F.M., Patterson, N., Reich, D., 2015. Genetic evidence for two founding populations of the Americas. Nature 525, 104–108. doi:10.1038/nature14895

Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., Clark, G.R., Reepmeyer, C., Petchey, F., Fernandes, D., Fu, Q., Harney, E., Lipson, M., Mallick, S., Novak, M., Rohland, N., Stewardson, K., Abdullah, S., Cox, M.P., Friedlaender, F.R., Friedlaender, J.S., Kivisild, T., Koki, G., Kusuma, P., Merriwether, D.A., Ricaut, F.-X., Wee, J.T.S., Patterson, N., Krause, J., Pinhasi, R., Reich, D., 2016. Genomic insights into the peopling of the Southwest Pacific. Nature 538, 510–513. doi:10.1038/nature19844

Soares, P., Rito, T., Trejaut, J., Mormina, M., Hill, C., Tinkler-Hundal, E., Braid, M., Clarke, D.J., Loo, J.-H., Thomson, N., Denham, T., Donohue, M., Macaulay, V., Lin, M., Oppenheimer, S., Richards, M.B., 2011. Ancient voyaging and Polynesian origins. Am. J. Hum. Genet. 88, 239–247. doi:10.1016/j.ajhg.2011.01.009

Soares, P.A., Trejaut, J.A., Rito, T., Cavadas, B., Hill, C., Eng, K.K., Mormina, M., Brandão, A., Fraser, R.M., Wang, T.-Y., Loo, J.-H., Snell, C., Ko, T.-M., Amorim, A., Pala, M., Macaulay, V., Bulbeck, D., Wilson, J.F., Gusmão, L., Pereira, L., Oppenheimer, S., Lin, M., Richards, M.B., 2016. Resolving the ancestry of Austronesian-speaking populations. Hum. Genet. 135, 309–326. doi:10.1007/s00439-015-1620-z

Soodyall, H., Jenkins, T., Stoneking, M., 1995. “Polynesian” mtDNA in the Malagasy. Nat. Genet. 10, 377–378. doi:10.1038/ng0895-377

Spriggs, M., 2007. The Neolithic and Austronesian expansion within Island Southeast Asia and into the Pacific, in: Chiu, S., Sand, C. (Eds.), From Southeast Asia to the Pacific: Archaeological Perspectives on the Austronesian Expansion and the Lapita Cultural Complex. Academia Sinica, Taiwan, pp. 104–125.

Spriggs, M., 1995. The lapita culture and Austronesian prehistory in Oceania, in: Fox, J.J., Tryon, D., Bellwood, P. (Eds.), The Austronesians: Historical and Comparative Perspectives. ANU E Press, Canberra, Australia.

Steinemann, M., Steinemann, S., 1998. Enigma of Y chromosome degeneration: Neo-Y and Neo-X chromosomes of Drosophila miranda a model for sex chromosome evolution, in: Woodruff, R.C., Jr, J.N.T. (Eds.), Mutation and Evolution, Contemporary Issues in Genetics and Evolution. Springer Netherlands, pp. 409–420. doi:10.1007/978-94-011-5210-5_33

Stoneking, M., 2008. Human origins. EMBO Rep. 9, S46–S50. doi:10.1038/embor.2008.64

Sudoyo, H., Suryadi, H., Sitorus, N., Soegondo, S., Pranoto, A., Marzuki, S., 2003. Mitochondrial genome and susceptibility to diabetes mellitus, in: Marzuki, S., Verhoef, J., Snippe, H. (Eds.), Tropical Diseases, Advances in Experimental Medicine and Biology. Springer US, pp. 19–36. doi:10.1007/978-1-4615-0059-9_2

The 1000 Genomes Project Consortium, 2015. A global reference for human genetic variation. Nature 526, 68–74. doi:10.1038/nature15393

Thedja, M.D., Muljono, D.H., Nurainy, N., Sukowati, C.H.C., Verhoef, J., Marzuki, S., 2011. Ethnogeographical structure of hepatitis B virus genotype distribution in Indonesia and discovery of a new subgenotype, B9. Arch. Virol. 156, 855–868. doi:10.1007/s00705-011-0926-y

Tobler, R., Rohrlach, A., Soubrier, J., Bover, P., Llamas, B., Tuke, J., Bean, N., Abdullah-Highfold, A., Agius, S., O'Donoghue, A., O'Loughlin, I., Sutton, P., Zilio, F., Walshe, K., Williams, A.N., Turney, C.S.M., Williams, M., Richards, S.M., Mitchell, R.J., Kowal, E., Stephen, J.R., Williams, L., Haak, W., Cooper, A., 2017. Aboriginal mitogenomes reveal 50,000 years of regionalism in Australia. Nature 544, 180–184. doi:10.1038/nature21416

Tofanelli, S., Bertoncini, S., Castrì, L., Luiselli, D., Calafell, F., Donati, G., Paoli, G., 2009. On the origins and admixture of Malagasy: New evidence from high-resolution analyses of paternal and maternal lineages. Mol. Biol. Evol. 26, 2109–2124. doi:10.1093/molbev/msp120

Trombetta, B., D'Atanasio, E., Massaia, A., Myres, N.M., Scozzari, R., Cruciani, F., Novelletto, A., 2015. Regional differences in the accumulation of SNPs on the male-specific portion of the human Y chromosome replicate autosomal patterns: Implications for genetic dating. PLOS ONE 10, e0134646. doi:10.1371/journal.pone.0134646

Tryon, D., 1995. Proto-Austronesian and the major Austronesian subgroups, in: Fox, J.J., Tryon, D., Bellwood, P. (Eds.), The Austronesians: Historical and Comparative Perspectives. ANU E Press, Canberra, Australia, pp. 19–41.

Tsukahara, T., Hombhanje, F.W., Lum, J.K., Hwaihwanje, I., Masta, A., Kaneko, A., Kobayakawa, T., 2006. Austronesian origin of the 27-bp deletion of the erythrocyte band 3 gene in East Sepik, Papua New Guinea inferred from mtDNA analysis. J. Hum. Genet. 51, 244–248. doi:10.1007/s10038-005-0352-7

Tumonggor, M.K., Karafet, T.M., Hallmark, B., Lansing, J.S., Sudoyo, H., Hammer, M.F., Cox, M.P., 2013. The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific. J. Hum. Genet. 58, 165–173. doi:10.1038/jhg.2012.154

Underhill, P.A., Kivisild, T., 2007. Use of Y cromosome and mitochondrial DNA population structure in tracing human migrations. Annu. Rev. Genet. 41, 539–564. doi:10.1146/annurev.genet.41.110306.130407

Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Mirazón Lahr, M., Foley, R.A., Oefner, P.J., Cavalli-Sforza, L.L., 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann. Hum. Genet. 65, 43–62. doi:10.1046/j.1469-1809.2001.6510043.x

Van Bemmelen, R.W., 1949. The Geology of Indonesia. Government Printing Office; Martinus Nijhoff, The Hague.

van Oven, M., Kayser, M., 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. 30, E386–E394. doi:10.1002/humu.20921

van Oven, M., Van Geystelen, A., Kayser, M., Decorte, R., Larmuseau, M.H., 2014. Seeing the wood for the trees: A minimal reference phylogeny for the human Y chromosome. Hum. Mutat. 35, 187–191. doi:10.1002/humu.22468

Voris, H.K., 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. J. Biogeogr. 27, 1153–1167. doi:10.1046/j.1365-2699.2000.00489.x

Wallace, D.C., 1994. Mitochondrial DNA sequence variation in human evolution and disease. Proc. Natl. Acad. Sci. 91, 8739–8746.

Wilder, J.A., Cox, M.P., Paquette, A.M., Alford, R., Satyagraha, A.W., Harahap, A., Sudoyo, H., 2011. Genetic continuity across a deeply divergent linguistic contact zone in North Maluku, Indonesia. BMC Genet. 12, 100. doi:10.1186/1471-2156-12-100

Wilder, J.A., Stone, J.A., Preston, E.G., Finn, L.E., Ratcliffe, H.L., Sudoyo, H., 2009. Molecular population genetics of SLC4A1 and Southeast Asian Ovalocytosis. J. Hum. Genet. 54, 182–187. doi:10.1038/jhg.2009.12

Wollstein, A., Lao, O., Becker, C., Brauer, S., Trent, R.J., Nürnberg, P., Stoneking, M., Kayser, M., 2010. Demographic history of Oceania inferred from genome-wide data. Curr. Biol. 20, 1983–1992. doi:10.1016/j.cub.2010.10.040

Wurster, C.M., Rifai, H., Haig, J., Titin, J., Jacobsen, G., Bird, M., 2017. Stable isotope composition of cave guano from eastern Borneo reveals tropical environments over the past 15,000 cal yr BP. Palaeogeogr. Palaeoclimatol. Palaeoecol. 473, 73–81. doi:10.1016/j.palaeo.2017.02.029

Xu, S., Pugach, I., Stoneking, M., Kayser, M., Jin, L., 2012. Genetic dating indicates that the Asian–Papuan admixture through Eastern Indonesia corresponds to the Austronesian expansion. Proc. Natl. Acad. Sci. 109, 4574–4579. doi:10.1073/pnas.1118892109

Zhang, Z.-H., Wu, C.-C., Chen, X.-W., Li, X., Li, J., Lu, M.-J., 2016. Genetic variation of hepatitis B virus and its significance for pathogenesis. World J. Gastroenterol. 22, 126–144. doi:10.3748/wjg.v22.i1.126

VI. Annexes

Annex A. Documentations on sampling activities

Figure A1. A phlebotomist (blue blouse) from local health center helped in verifying the questionaire of a volunteer (yellow dress) before conducting the blood and saliva sampling during the Bajo sampling campaign in Kotabaru, South Kalimantan.

Figure A2. Prof. Herawati Sudoyo (red shirt) from Eijkman Institute, Jakarta gave a presentation about general overview of the study to all volunteers before conducting the sampling.

Figure A3. The author, Pradiptajati Kusuma (left, black shirt), gave interview to volunteers and fill the questionnaire before taking the samples. This sampling was conducted in Buluh Kuning village, South Kalimantan, in the house of the chief of Samihim ethnic group in the region.

Figure A4. One of the sampling activity conducted in a villager's house terrace in Buluh Kuning village, South Kalimantan.

Annex B. List of analyses performed on each population brought in this study.

No	**Ethnic group**	**mtDNA HVR**	**Whole mtDNA**	**Y-chromosome SNPs**	**Array Omni Express (700K SNPs)**	**Array Omni 2.5 (2.5M SNPs)**	**Array Omni 5 (4.5M SNPs)**
1	Ma’anyan	v	*	v	v		
2	SK Dayak	v	*	v	v		
3	Ngaju	v	*	v	v		
4	Banjar		*	*	v		
5	Samihim		*	*			v
6	Bajo Kotabaru		*	*			v
7	Bajo Derawan		*	*			v
8	Lebbo’	v	*	v	v		
9	Bajo Kendari	v	*	v	v		
10	Bugis		*	*			v
11	Mandar		*	*		v	
12	North Maluku	v	*	v			v

* Data have been generated but not analysed in this thesis due to different data production time-frame.

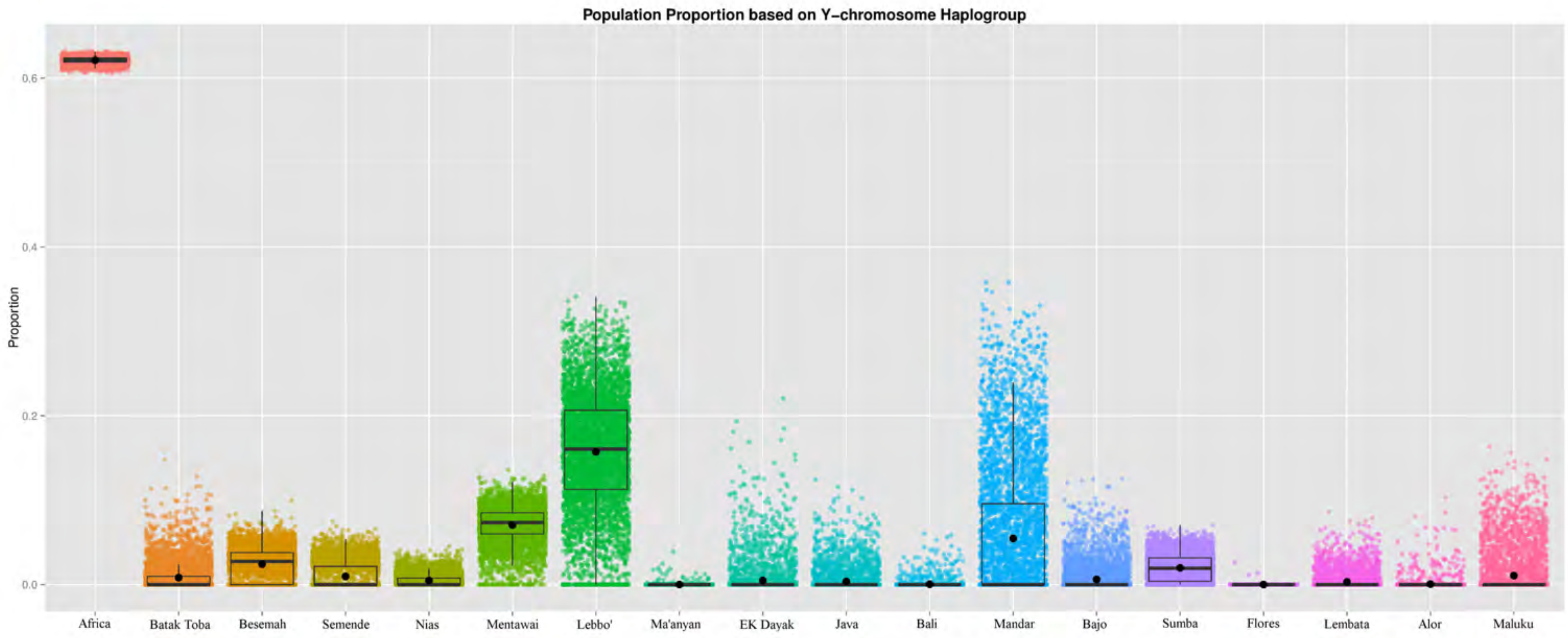
Annex C. Supplementary informations from article entitled “Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations”

(Published in BMC Genomics, 16:191, March 17th 2015. doi: 10.1186/s12864-015-1394-7)

Pradiptajati Kusuma[1,4], Murray P. Cox[2], Denis Pierron[1], Harilanto Razafindrazaka[1], Nicolas Brucato[3], Laure Tonasso[1], Helena Loa Suryadi[4], Thierry Letellier[1], Herawati Sudoyo[4,5], François-Xavier Ricaut[1§]

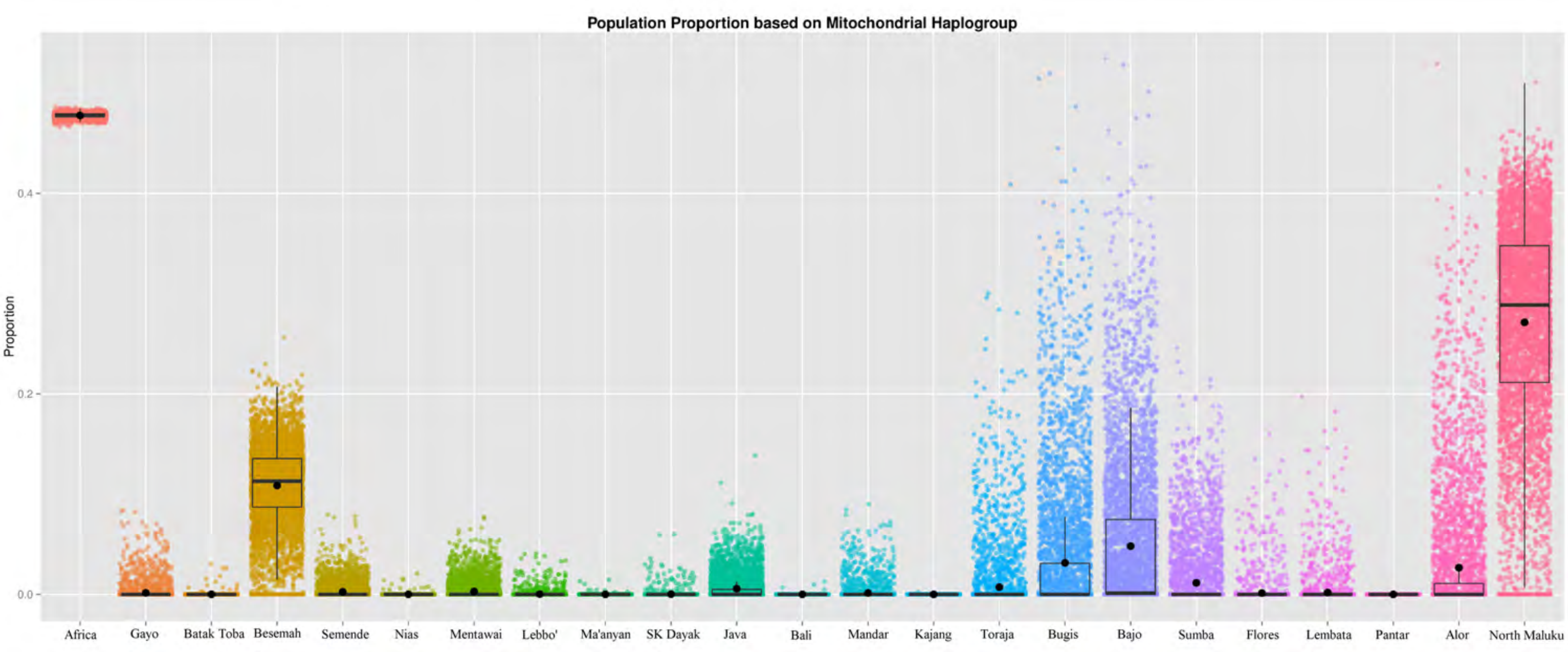
[1] Laboratoire d’Anthropologie Moléculaire et Imagérie de Synthèse UMR-5288, Université de Toulouse, Toulouse, France
[2] Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
[3] Center for Linguistics, University of Leiden, Leiden, Netherlands
[4] Eijkman Institute for Molecular Biology, Jakarta, Indonesia
[5] Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia

Figure S1. Population proportion based on Y-chromosome haplogroups estimated with the lsei algorithm.



Note: Black dots represent the mean value

Figure S2. Population proportion based on mitochondrial DNA haplogroups estimated with the lsei algorithm.



Note: Black dots represent the mean value

Table S1. List of populations used in this study.

Group	Islands	Code	Population	N mtDNA	N Y-chr	Ref.
Malagasy***	Madagascar		Antemoro	135	129	[50]
			Merina	47	34	[19**, 51]
			Mikea	127	59	[51]
			Vezo	98	48	[51]
			Antandroy	59	46	[19]**
			Antanosy	53	47	[19]**
			Antaisaka	10	8	[19]**
Western Indonesia	Sumatera	1	Gayo	62	*NA*	Eijkman Institute archived samples
		2	Batak Toba	42	37	[32, 33]
		3	Besemah	36	38	[40]
		4	Semende	36	37	[40]
		5	Nias	59	60	[32, 33]
		6	Mentawai	128	74	[32, 33]
	Borneo	7	**Lebbo'**	**19**	**15**	**This study**
		8	**Ma'anyan**	**159**	**90**	**This study**
		9	SK Dayak*	64	*NA*	Eijkman Institute archived samples
		10	EK Dayak*	*NA*	85	[32]
	Java	11	Java	51	61	[32, 33]
	Bali	12	Bali	487	634	[32, 33]
Eastern Indonesia	Sulawesi	13	Mandar	54	54	[32, 33]
		14	Kajang	46	*NA*	[33]
		15	Toraja	50	*NA*	[33]
		16	Bugis	50	*NA*	[33]
		17	**Bajo**	**27**	**27**	**This study**
	Lesser Sunda	18	Sumba	634	349	[32, 33]
		19	Flores	469	388	[32, 33]
		20	Lembata	92	89	[32, 33]
		21	Pantar	29	*NA*	[33]
		22	Alor	23	27	[32, 33]
	Maluku	23	North Maluku	224	*NA*	Eijkman Institute archived samples
		24	Maluku (Hiri and Ternate)	*NA*	30	[33]

Note : NA = Not Available for analysis

* *Samples of SK Dayak were collected from various Dayak ethnics in South Kalimantan, while the EK Dayak from various Dayak ethnics in*

** *Populations not used for shared mitochondrial haplotype analysis due to short fragment sequence available*

*** *These populations have been pooled as single group for all analysis based on their genetic homogeinity (see main text)*

Table S2. List of Y chromosome SNP markers screened.

Index	RefSNP ID	SNPs	Haplogroup
1	-	P114	A0a1a
2	-	M31	A1a
3	rs3905	M14	A1b1a1
4	rs2032603	M190	A1b1b2
5	rs34078768	M51	A1b1b2a
6	rs3904	M13	A1b1b2b
7	rs2534636	SRY10831.1	BT
8	-	P97	BT
9	rs2032651	M91	BT
10	rs2032623	M60	B
11	-	M150	B2a
12	-	M112	B2b
13	rs2032594	M169	B2b1a1a
14	-	M129	B2b1a1b
15	rs2032663	M211	B2b1a1c
16	-	P70	B2b1a2
17	-	P6	B2b1b
18	rs2032653	M203	DE
19	rs3848982	M145	DE
20	rs60115999	P29	E
21	rs9306841	M96	E
22	-	M33	E1a
23	rs9785756	P2	E1b1
24	rs768983	V38	E1b1a
25	rs9785941	M2	E1b1a1
26	rs2032590	M191	E1b1a1a1c1a
27	rs16980586	U174	E1b1a1a1c1a1
28	-	P115	E1b1a1a1c1a1b
29	-	P116	E1b1a1a1c1a1c
30	rs16980588	U175	E1b1a1a1d
31	rs16980502	U209	E1b1a1a1d1
32	rs16980406	U290	E1b1a1a1d1a
33	rs16980589	U181	E1b1a1a1d1a1
34	-	M154	E1b1a1a1d1c
35	rs2032654	M215	E1b1b
36	-	M35	E1b1b1
37	-	M78	E1b1b1a1
38	rs2032640	M81	E1b1b1b1a
39	-	M123	E1b1b1b2a
40	-	M34	E1b1b1b2a1
41	rs2032639	M75	E2
42	-	M41	E2a
43	rs2032620	M54	E2b
44	rs2032666	M216	C
45	rs35284970	M130	C
46	-	M38	C2
47	rs2032668	M217	C3
48	rs2032652	M89	F
49	rs2032665	M213	F
50	rs2032636	M201	G
51	rs4116820	P287	G2
52	-	P15	G2a
53	rs2032673	M69	H1
54	-	M52	H1a
55	rs2032675	M82	H1a1
56	rs13447352	M304	J
57	rs17315835	P209	J
58	rs9341313	M267	J1
59	rs34043621	P58	J1a2b
60	rs2032604	M172	J2
61	-	M410	J2a
62	rs2032628	M67	J2a1b
63	rs3903	M12	J2b
64	rs8179022	M241	J2b2
65	rs2033003	M526	K(xLT)
66	rs17250121	P128	K
67	rs3900	M9	K
68	-	P256	M
69	rs2032681	M186	M1
70	rs2032631	M45	P
71	rs2032658	M207	R
72	rs2032624	M173	R1
73	-	M434	R1a (private)
74	rs3908	M17	R1a1a
75	-	M458	R1a1a1b1a1
76	rs9786194	M415	R1b1
77	rs9786153	M269	R1b1a2
78	-	M124	R2a
79	rs16981290	P186	O
80	-	M119	O1a
81	rs13447354	P203	O1a1 / I1
82	rs2032632	M50	O1a2
83	rs2032650	M95	O2a1
84	rs2032645	M88	O2a1a
85	-	M122	O3
86	rs2267801	P201	O3a2
87	rs3898	M7	O3a2b
88	-	M134	O3a2c1
89	rs3911	M20	L
90	rs9341318	M295	L1
91	-	M76	L1a
92	-	M357	L1c
93	rs20320	M184	T
94	rs9341308	M272	T
95	rs2032672	M70	T1a
96	rs2215828	L131	T1a2

Table S3. Y chromosome haplogroup frequencies in Indonesia and Malagasy populations.

			Western Indonesia										Eastern Indonesia						
		Malagasy	**Batak Toba**	**Besemah**	**Semende**	**Nias**	**Mentawai**	**Lebbo'**	**Ma'anyan**	**EK Dayak**	**Java**	**Bali**	**Mandar**	**Bajo**	**Sumba**	**Flores**	**Lembata**	**Alor**	**Maluku**
C	**C-RPS4Y***	0.0054	--	--	--	--	0.1351	0.1333	0.5222	0.2235	0.0328	0.0142	0.1111	0.0370	0.0029	0.2912	0.2247	0.0370	--
C1c	**C-M38***	--	--	--	--	--	--	--	--	--	--	--	0.1111	0.2222	0.5731	0.1546	0.2135	0.4444	0.3333
C2	**C-M217***	--	--	--	--	--	--	--	--	--	--	0.0016	--	--	--	--	--	--	--
F	**F-P14***	--	--	--	--	--	--	--	--	0.0235	0.0164	--	0.0556	--	--	0.0155	0.0449	--	--
K	**K-M9***	--	--	--	--	--	--	0.0667	--	--	--	--	--	--	--	--	--	--	--
KxLT	**K-M526***	--	0.1351	--	--	--	--	0.1333	0.0222	0.0588	0.0328	--	0.0741	0.2222	0.0401	0.0619	0.0337	0.3333	0.1667
K3	**K-P261**	--	--	--	--	--	--	--	--	--	--	0.0095	--	--	--	--	--	--	--
M	**M-P256**	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0258	--	--	--
M1a	**M-P34**	--	--	--	--	--	--	--	--	--	--	0.0063	0.0370	--	0.0487	0.1856	0.1236	0.1481	0.1667
O	**O-M175***	--	--	0.0811	0.0811	--	--	--	--	--	--	--	--	--	--	--	--	--	--
O1a	**O-M119***	0.0458	--	--	--	--	0.5676	--	0.0222	0.0118	0.0328	0.0899	0.1296	0.0370	0.0573	0.0026	0.0225	--	--
O1a1	**O-P203**	--	0.0270	--	--	0.8667	0.2838	--	0.1000	0.0235	0.0984	0.1151	--	--	0.0086	0.0928	0.0112	--	0.0333
O1a2	**O-M110**	0.1105	0.1081	0.0811	0.0541	0.1333	0.0135	0.2667	0.0222	0.0588	0.0164	0.0079	0.0741	--	0.0630	--	--	--	0.0333
O2	**O-P31***	--	--	0.7027	0.7297	--	--	--	--	--	--	--	--	--	--	--	--	--	--
O2a1	**O-M95***	0.0243	0.1351	--	--	--	--	0.3333	0.0778	0.2118	0.4918	0.5694	0.1296	--	0.0029	0.0464	--	--	--
O2a1a	**O-M88**	0.0377	--	--	--	--	--	--	0.0111	--	--	--	--	--	--	--	--	--	--
O3	**O-M122***	--	--	0.1351351351	0.1351	--	--	--	0.0111	--	--	0.0126	--	0.0370	0.0201	--	--	--	--
O3a2	**O-P201***	--	0.5676	--	--	--	--	0.0667	0.1778	0.1529	0.0164	0.0521	0.1667	0.0741	0.0086	0.0464	--	--	0.1000
O3a2b	**O-M7**	--	--	--	--	--	--	--	--	0.2000	0.1148	0.0079	--	0.0370	--	--	--	--	--
M1a	**M-186**	--	--	--	--	--	--	--	--	--	--	--	--	0.0741	--	--	--	--	--
P	**P-M45***	--	--	--	--	--	--	--	0.0222	--	--	--	--	--	--	--	--	--	--
S	**S-M230***	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0086	0.0052	0.0225	--	--
S1	**S-M254**	--	--	--	--	--	--	--	--	--	--	0.0032	0.0556	--	0.1662	0.0722	0.3034	0.0370	0.1667
Q1a2	**Q-M346**	--	--	--	--	--	--	--	--	--	--	0.0032	--	--	--	--	--	--	--
R	**M207**	--	--	--	--	--	--	--	--	--	--	--	--	0.1481	--	--	--	--	--
R1a	**R-M17**	0.0081	--	--	--	--	--	--	0.0111	0.0118	0.0984	0.0142	0.0185	--	--	--	--	--	--
R1b1	**R-M415**	0.0162	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
R1b1a2	**R-M269**	--	--	--	--	--	--	--	--	--	--	0.0063	--	--	--	--	--	--	--
R2a	**R-M124**	--	0.0270	--	--	--	--	--	--	0.0118	--	0.0174	0.0185	--	--	--	--	--	--
H1	**H-M69***	--	--	--	--	--	--	--	--	--	--	0.0268	--	--	--	--	--	--	--
H1a	**H-M52**	--	--	--	--	--	--	--	--	0.0118	--	0.0016	--	--	--	--	--	--	--
H1b1	**H-Apt**	--	--	--	--	--	--	--	--	--	--	0.0016	--	--	--	--	--	--	--
J	**J-M304**	--	--	--	--	--	--	--	--	--	--	0.0032	--	--	--	--	--	--	--
J1	**J-M267**	0.0458	--	--	--	--	--	--	--	--	--	0.0095	--	--	--	--	--	--	--
J2	**J-M172***	0.0108	--	--	--	--	--	--	--	--	0.0492	--	0.0185	--	--	--	--	--	--
J2b	**J-M12**	0.0108	--	--	--	--	--	--	--	--	--	0.0063	--	--	--	--	--	--	--
T	**T-M272***	0.0701	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
T1a	**T-M70**	--	--	--	--	--	--	--	--	--	--	--	--	0.0741	--	--	--	--	--
L	**L-M20***	0.0027	--	--	--	--	--	--	--	--	--	0.0205	--	--	--	--	--	--	--
L1a	**L-M76**	--	--	--	--	--	--	--	--	--	--	--	--	0.0370	--	--	--	--	--
D	**D-M174**	0.0027	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
African Hg	**E***	0.6092	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Table S4. Mitochondrial haplogroup frequencies in Indonesian and Malagasy populations.

		Western Indonesia											Eastern Indonesia										
	Malagasy	**Gayo**	**Batak Toba**	**Besemah**	**Semende**	**Nias**	**Mentawai**	**Lebbo'**	**Ma'anyan**	**SK Dayak**	**Java**	**Bali**	**Mandar**	**Kajang**	**Toraja**	**Bugis**	**Bajo**	**Sumba**	**Flores**	**Lembata**	**Pantar**	**Alor**	**North Maluku**
B4	--	--	--	--	--	--	--	--	--	--	--	0.0021	--	--	--	--	--	--	--	--	--	--	0.0045
B4a	0.0151	--	0.0952	0.0278	0.0833	0.0339	--	0.2105	0.0943	0.0462	--	0.0103	--	0.0652	0.1200	0.0600	0.0741	0.0252	0.0277	0.0761	--	0.0870	0.1161
B4a1a1	0.2098	--	--	--	--	--	--	--	--	--	--	0.0041	--	--	0.0200	0.0200	0.0370	0.0205	0.0171	0.0109	--	0.0435	0.0848
B4a1a3	--	--	--	--	--	--	--	--	--	--	--	0.0062	--	--	0.0400	--	--	0.0095	0.0085	0.0435	--	0.0435	0.0045
B4a1c3	--	--	--	--	0.0278	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
B4a2a	--	--	--	--	0.0556	--	--	--	0.0440	--	--	--	--	--	--	--	--	--	--	--	--	--	--
B4a3	--	--	--	--	--	--	--	--	--	--	--	0.0062	--	--	--	--	--	--	--	--	--	--	--
B4a4	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0741	--	--	--	--	--	--
B4b1	--	--	--	--	--	--	--	--	0.0818	--	--	--	--	0.1739	--	--	--	0.0584	0.0384	0.0326	--	0.0435	0.0179
B4c1b	--	0.0484	0.0714	0.0278	0.0556	0.1186	--	--	0.0252	0.0462	0.0392	0.1027	0.0741	0.1739	0.0600	0.0800	0.0741	0.0410	0.0043	--	0.1724	--	0.0313
B4c2	--	0.0161	--	0.0278	--	--	--	--	0.1887	0.0923	0.0392	0.0678	--	0.0217	--	0.0200	--	0.0016	0.0149	--	--	--	--
B5	--	--	--	--	--	0.0169	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
B5a	--	--	--	0.0278	--	--	--	0.1579	0.0377	0.0769	0.0588	0.1745	--	--	--	--	0.0370	0.0158	0.0320	0.0109	--	--	0.0045
B5a1d	--	--	--	0.0556	0.0278	--	--	--	--	--	0.0196	--	--	--	--	0.0200	--	0.0095	0.0085	--	--	--	--
B5b	--	0.0323	--	--	--	0.0508	0.2344	--	--	0.0308	--	0.0349	0.0370	--	--	--	--	0.0615	0.0256	--	--	0.0435	0.0580
B5b2	--	0.0323	--	--	--	0.0508	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0109	--	--	--
B6	--	--	--	--	--	--	--	--	--	--	0.0980	0.0021	--	--	--	--	--	--	--	--	--	--	--
C1b	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0021	--	--	--	--
D1	--	--	--	--	--	--	--	--	--	--	0.0196	--	--	--	--	--	--	--	--	--	--	--	--
D2	--	0.0161	0.0238	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0047	--	--	--	--	--
D4a	--	--	--	--	--	--	--	--	--	--	--	0.0041	--	--	--	0.0200	--	--	--	--	--	--	--
D4b	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0217	--	--	--	0.0189	--	--	--	--	--
D4s	--	0.0161	0.0238	0.0278	--	--	--	--	0.0629	0.0769	0.0392	--	0.0185	--	--	0.0200	--	0.0095	0.0192	--	--	--	0.0045
D4i	--	--	0.0238	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
D4j	--	--	--	--	--	--	--	--	--	0.0154	--	--	--	--	--	--	--	0.0095	--	--	--	--	--
D5b1c1	--	0.0161	0.0238	--	--	0.0339	0.0078	--	--	0.0154	0.0196	--	--	--	0.0400	0.0200	--	0.0505	0.0384	0.0435	--	--	0.0089
D6	--	--	--	--	--	--	--	--	--	--	--	0.0082	--	--	--	--	--	--	0.0299	--	0.0345	--	0.0134
E	--	--	--	--	--	--	--	--	--	--	0.0196	0.0123	--	--	0.0400	0.0600	--	0.0189	0.0021	--	--	--	0.0714
E1a	0.0491	--	--	0.1667	0.0278	--	--	0.2105	--	--	--	0.0226	0.2593	0.1087	0.2000	0.1200	0.0741	0.0284	0.0235	0.0870	0.0345	0.0870	0.0580
E1b	--	--	--	--	--	--	0.0313	--	--	0.0154	0.0196	0.0062	0.1296	0.0217	0.0400	0.0800	--	0.0489	0.0128	0.0109	0.0345	0.0435	0.0357
E2	--	0.0161	--	0.0278	--	--	--	--	--	0.0462	0.0392	0.0021	0.0370	--	--	0.0200	--	0.0237	0.0426	--	--	--	0.0536
E2a3	--	--	--	--	--	--	--	--	--	--	--	0.0062	0.0556	0.0217	0.0400	0.0400	--	0.0095	--	--	--	--	--
F1a	--	0.0806	0.0476	0.0278	0.0833	--	--	--	0.0377	0.0154	0.0196	0.0164	0.0556	--	0.0400	0.0200	0.0370	0.0331	0.0021	0.0109	--	--	0.0045
F1a1a	--	0.0323	0.0238	--	0.0278	0.0169	--	--	0.0377	0.0923	0.0784	0.0082	--	--	--	0.0200	--	0.0126	0.0256	0.0109	--	--	--
F1a1c1	--	0.0161	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
F1a1d	--	--	--	--	--	--	--	--	--	--	--	0.0082	--	--	--	--	--	--	--	--	--	--	--
F1a2	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0079	--	--	--	--	--
F1a3	--	--	--	--	--	--	--	--	--	0.0308	--	0.0226	0.0185	0.0217	0.0400	0.0200	0.0741	0.0521	0.0704	0.0217	0.0345	--	0.0223
F1a4	--	--	--	0.0278	--	--	--	--	0.0252	0.0308	--	0.0144	0.0185	--	0.0600	0.0200	--	0.0473	0.0426	0.1413	--	0.1304	0.0223
F1c	--	--	--	--	--	--	--	--	--	--	--	0.0021	--	--	--	--	--	--	--	--	--	--	--
F1e3	--	--	--	--	0.0278	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
F3a	--	--	--	--	--	--	--	--	--	--	0.1373	--	--	--	--	--	--	--	--	--	--	--	--
F3b1a	0.0548	0.0161	--	--	0.0278	--	--	--	0.0189	0.0308	--	--	--	--	--	--	--	0.0095	0.0021	--	--	--	--
F4b	--	0.0161	--	--	--	--	--	--	--	--	--	--	0.0370	--	--	--	--	--	--	--	--	--	--
G	--	--	--	--	--	--	--	--	--	--	--	0.0021	--	--	--	0.0200	--	0.0252	--	--	--	--	--
H*	--	0.0645	0.0238	--	--	--	--	--	--	--	--	0.0123	--	--	0.0400	--	--	--	--	0.0109	--	--	0.0179
M*	--	0.0645	0.0714	0.0278	--	--	0.0078	--	--	--	0.0196	0.0472	0.0185	--	--	0.0200	--	0.0174	0.0576	0.0109	--	--	0.0045
M12	--	0.0161	0.0238	--	--	--	--	--	0.0377	0.0154	0.0392	0.0021	--	--	--	--	--	--	0.0277	--	--	--	--
M17	--	--	--	--	--	--	--	--	--	--	--	0.0534	--	--	--	--	--	0.0047	0.0064	--	0.0345	0.0435	--
M18	--	0.0161	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
M19	--	--	--	--	--	--	--	--	--	--	--	0.0411	--	--	--	--	--	--	--	--	--	--	--
M2	--	--	--	--	--	--	--	--	0.0063	--	--	--	--	--	--	--	--	--	--	--	--	--	--
M20	--	0.0645	0.0714	--	--	0.0508	0.1250	0.1579	0.0440	0.0308	0.0196	0.0041	--	--	--	0.0200	--	--	0.0021	--	--	--	0.0089
M21	--	--	--	--	--	--	--	--	--	--	--	0.0226	--	--	--	0.0200	--	--	--	0.0109	--	0.0435	0.0045
M22	--	0.0161	--	--	--	--	--	--	--	0.0154	--	--	--	--	--	--	--	--	--	--	--	--	--
M23	0.0681	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Table S4 (continued). Mitochondrial haplogroup frequencies in Indonesian and Malagasy populations.

M26	--	0.0323	--	0.0278	0.1944	--	--	--	--	--	--	--	--	--	--	--	--	0.0063	--	--	--	--	--
M29	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0047	--	--	--	--	--
M32c	0.0435	--	--	--	--	--	--	--	--	--	0.0196	--	--	--	--	--	--	--	--	--	--	--	--
M33	--	0.0161	--	--	--	--	--	--	--	--	0.0588	--	--	--	--	--	--	--	--	--	--	--	--
M35a	--	0.0161	--	--	--	--	--	--	0.0189	--	--	--	--	--	--	--	--	--	--	--	--	--	--
M40	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0200	--	--	0.0021	--	0.0690	--	--
M41a	--	0.0161	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
M45	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0200	--	--	--	--	--	--	--	--
M47	--	--	0.0952	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0032	0.0021	--	--	--	--
M5	--	--	--	--	--	--	--	--	0.0063	--	--	0.0205	--	--	--	--	--	--	0.0171	--	0.0345	--	--
M50	--	0.0161	0.0476	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
M51a1a	--	--	0.0238	--	0.2500	--	--	--	--	0.0154	--	0.0103	--	--	0.0200	--	--	0.0032	0.0043	--	--	--	--
M59	--	--	--	--	--	--	--	--	--	--	--	0.0041	--	--	--	--	--	--	--	--	--	--	--
M68	--	--	--	--	--	--	--	--	--	--	--	0.0041	--	--	--	--	--	--	--	--	--	--	--
M71a2	--	--	--	--	--	--	--	0.1579	--	--	--	0.0021	--	0.0217	--	--	--	--	0.0256	--	--	--	0.0045
M72	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0047	--	--	0.0345	0.0435	--
M73'79	--	--	--	--	--	--	--	--	--	0.0154	--	--	--	--	--	--	--	--	0.0384	--	--	--	--
M73	--	--	--	--	--	--	--	--	0.0566	--	--	0.0082	0.0741	0.3261	--	--	0.0370	--	0.0107	--	0.0345	--	0.0268
M74b1	--	--	--	0.0278	--	0.0169	--	--	0.1069	0.0308	--	--	--	--	--	--	--	0.0221	--	0.0109	--	--	--
M77	--	--	--	--	--	--	--	--	--	--	--	0.0287	--	--	--	--	--	--	--	--	--	--	--
M79	--	--	--	--	--	--	--	--	--	--	--	0.0021	--	--	--	--	--	--	--	--	--	--	--
M7b1a1+(16192)	--	0.0484	--	--	--	--	--	--	--	0.0154	0.0392	0.0144	--	--	0.0200	--	--	0.0032	0.0043	0.0109	--	--	0.0089
M7b1a1i	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.1481	--	--	--	--	--	0.0089
M7b1a1a	--	--	--	--	--	--	--	--	--	--	0.0196	0.0082	--	--	--	--	--	0.0126	0.0064	--	--	--	--
M7b1a2	--	--	--	--	--	--	--	--	0.0126	0.0308	--	--	0.0370	--	--	0.0400	0.0741	0.0521	0.0149	--	--	--	0.0179
M7c1	--	--	--	--	--	--	--	--	--	--	--	0.0021	--	--	--	--	--	0.0016	--	--	--	--	--
M7c1a4a	0.0794	0.0484	0.0714	0.3056	0.0556	0.1356	0.2500	--	0.0126	0.0154	0.0784	0.0287	0.0926	--	0.0200	0.1000	0.1111	0.1136	0.0490	0.0870	0.0690	0.0435	0.0313
M7c1c3a	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0400	--	--	--	--	--	--	--	--
M81	--	--	--	--	--	--	--	--	--	--	--	0.0021	--	--	--	--	--	--	--	--	--	--	--
N	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0016	--	0.0217	0.0345	--	0.0045
N10	--	--	--	--	--	--	--	--	--	0.0154	--	0.0021	--	--	--	--	--	--	--	--	--	--	0.0045
N21	--	--	--	--	0.0278	--	--	--	--	--	0.0196	0.0185	--	--	--	--	--	--	--	--	--	0.0435	--
N22	--	--	0.0238	0.0278	--	--	--	--	0.0189	0.0308	--	0.0062	--	--	--	--	--	0.0079	--	--	--	--	--
N5	--	--	--	--	--	--	--	--	--	--	--	0.0021	--	--	--	--	--	0.0016	--	--	--	--	--
N8	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0085	--	--	--	--
N9a6a	--	0.0161	--	0.0556	0.0278	--	--	--	0.0252	0.0462	--	0.0123	0.0185	--	0.0200	0.0200	--	0.0032	0.0128	--	--	--	--
N9b	--	--	--	--	--	--	--	--	--	--	--	0.0062	--	--	--	--	--	--	0.0064	--	--	--	--
P1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0200	--	0.0063	0.0171	0.0326	--	--	0.0625
P1d1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0095	0.0213	0.0978	0.0345	0.0435	0.0179
P4	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0320	--	--	--	0.0045
Q	--	--	--	--	--	--	--	--	--	--	--	0.0062	--	--	--	--	--	--	--	--	--	--	--
Q1	0.0019	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0400	0.0200	740740740740	0.0284	0.0576	0.0761	0.3103	0.2174	0.1071
Q2	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0043	0.0326	--	--	0.0179
Q3	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0045
R14	--	--	--	--	--	--	--	--	--	--	--	0.0021	--	--	--	0.0200	--	--	0.0021	0.0761	--	--	--
R21	--	0.0161	--	--	--	--	0.0078	--	--	0.0154	--	--	--	--	--	--	--	--	--	--	--	--	--
R22	--	--	--	--	--	0.0169	0.0391	--	--	--	0.0392	0.0308	--	--	--	--	370370370370	0.0079	0.0256	--	--	--	--
R23	--	--	--	--	--	--	--	--	--	--	--	0.0123	--	--	--	--	--	0.0016	--	--	--	--	--
R24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0045
R9b1a1a	--	0.0323	--	--	--	--	0.0859	0.1053	--	--	--	--	--	0.0217	--	--	--	0.0016	0.0320	0.0109	--	--	--
R9c1a	--	--	0.0238	--	--	--	--	--	--	0.0308	--	0.0041	--	--	--	--	--	0.0284	0.0171	--	0.0345	--	0.0045
U2	--	0.0323	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
U7	--	0.0323	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
V1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0200	--	--	--	--	--	--	0.0045
X	--	0.0161	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0.0370	--	--	--	--	--	--
Y2	--	0.0645	0.1905	0.0833	--	0.4576	0.2109	--	--	0.0154	--	0.0329	0.0185	--	0.0400	--	--	--	0.0043	--	--	--	0.0134
African Hg (L*)	0.4783	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Table S5. Y chromosome haplogroup sharing frequencies in Indonesian and Malagasy populations.

			Shared haplogroups													
Population		**n ind**	**C**		**O1a**		**O1a2**		**O2a1**		**O2a1a**		**J1 (WE)**		**J2 (WE)**	
			n	**freq**	**n**	**freq**	**n**	**freq**	**n**	**freq**	**n**	**freq**	**n**	**freq**	**n**	**freq**
Malagasy		371	2	0.00539	17	0.04582	41	0.11051	9	0.02426	14	0.03774	17	0.04582	4	0.01078
Western Indonesia	Batak Toba	37	--	--	--	--	4	0.10811	5	0.13514	--	--	--	--	--	--
	Besemah	38	--	--	--	--	3	0.08108	--	--	--	--	--	--	--	--
	Semende	37	--	--	--	--	2	0.05405	--	--	--	--	--	--	--	--
	Nias	60	--	--	--	--	8	0.13333	--	--	--	--	--	--	--	--
	Mentawai	74	10	0.13514	42	0.56757	1	0.01351	--	--	--	--	--	--	--	--
	Lebbo'	15	2	0.13333	--	--	4	0.26667	5	0.33333	--	--	--	--	--	--
	Ma'anyan	90	47	0.52222	2	0.02222	2	0.02222	8	0.08889	1	0.01111	--	--	--	--
	EK Dayak	85	19	0.22353	1	0.01176	5	0.05882	18	0.21176	--	--	--	--	--	--
	Java	61	2	0.03279	2	0.03279	1	0.01639	30	0.49180	--	--	--	--	3	0.04918
	Bali	634	9	0.01420	57	0.08991	5	0.00789	361	0.56940	--	--	6	0.00946	--	--
Eastern Indonesia	Mandar	54	6	0.11111	7	0.12963	4	0.07407	7	0.12963	--	--	--	--	1	0.01852
	Bajo	27	1	0.03704	1	0.03704	--	--	--	--	--	--	--	--	--	--
	Sumba	349	1	0.00287	20	0.05731	22	0.06304	1	0.00287	--	--	--	--	--	--
	Flores	388	113	0.29124	1	0.00258	--	--	18	0.04639	--	--	--	--	--	--
	Lembata	89	20	0.22472	2	0.02247	--	--	--	--	--	--	--	--	--	--
	Alor	27	1	0.03704	--	--	--	--	--	--	--	--	--	--	--	--
	Maluku	30	--	--	--	--	1	0.03333	--	--	--	--	--	--	--	--

			Shared Haplogroups								Not shared haplogroups					
Population		**n ind**	**J2b (WE)**		**T* (WE)**		**L* (WE)**		**R1a (WE)**		**Other Asian Hg**		**Other WE Hg**		**African Hg**	
			n	**freq**	**n**	**freq**	**n**	**freq**	**n**	**freq**	**n**	**freq**	**n**	**freq**	**n**	**freq**
Malagasy		371	4	0.01078	26	0.07008	1	0.00270	3	0.00809	--	--	7	0.01887	226	0.60916
Western Indonesia	Batak Toba	37	--	--	--	--	--	--	--	--	27	0.72973	1	0.02703	--	--
	Besemah	38	--	--	--	--	--	--	--	--	35	0.91892	--	--	--	--
	Semende	37	--	--	--	--	--	--	--	--	35	0.94595	--	--	--	--
	Nias	60	--	--	--	--	--	--	--	--	52	0.86667	--	--	--	--
	Mentawai	74	--	--	--	--	--	--	--	--	21	0.28378	--	--	--	--
	Lebbo'	15	--	--	--	--	--	--	--	--	4	0.26667	--	--	--	--
	Ma'anyan	90	--	--	--	--	--	--	1	0.01111	30	0.33333	--	--	--	--
	EK Dayak	85	--	--	--	--	--	--	1	0.01176	39	0.45882	2	0.02353	--	--
	Java	61	--	--	--	--	--	--	6	0.09836	17	0.27869	--	--	--	--
	Bali	634	4	0.00631	--	--	13	0.02050	9	0.01420	132	0.20820	38	0.05994	--	--
Eastern Indonesia	Mandar	54	--	--	--	--	--	--	1	0.01852	27	0.50000	1	0.01852	--	--
	Bajo	27	--	--	2	0.07407	1	0.03704	--	--	22	0.81481	--	--	--	--
	Sumba	349	--	--	--	--	--	--	--	--	305	0.87393	--	--	--	--
	Flores	388	--	--	--	--	--	--	--	--	256	0.65979	--	--	--	--
	Lembata	89	--	--	--	--	--	--	--	--	67	0.75281	--	--	--	--
	Alor	27	--	--	--	--	--	--	--	--	26	0.96296	--	--	--	--
	Maluku	30	--	--	--	--	--	--	--	--	29	0.96667	--	--	--	--

Table S6. Y chromosome F_{ST} distances between Malagasy and Indonesian populations.

Population		Fst	p-value
Mandar	EI	0.25891	<0.01
Flores	EI	0.27588	<0.01
Bajo	EI	0.28268	<0.01
EK Dayak	WI	0.28773	<0.01
Lebbo'	WI	0.29014	<0.01
Lembata	EI	0.31388	<0.01
Maluku	EI	0.31488	<0.01
Java	WI	0.33563	<0.01
Bali	WI	0.35691	<0.01
Ma'anyan	WI	0.35744	<0.01
Alor	EI	0.36493	<0.01
Batak Toba	WI	0.37053	<0.01
Sumba	EI	0.37383	<0.01
Mentawai	WI	0.38326	<0.01
Besemah	WI	0.42927	<0.01
Semende	WI	0.43841	<0.01
Nias	WI	0.50283	<0.01

Note.

U test p value = 0.06; EI=Eastern Indonesia; WI= Western Indonesia

Table S7. Mitochondrial haplogroup sharing frequencies between Malagasy and Indonesian populations.

Population		n ind	Shared haplogroups									
			B4a1a		B4a1a1		E1a1a		F3b		M7c1a4a	
			n	freq	n	freq	n	freq	n	freq	n	freq
Malagasy		529	8	0.01512	111	0.20983	26	0.04915	29	0.05482	42	0.07940
Western Indonesia	Gayo	62	--	--	--	--	--	--	1	0.01613	3	0.04839
	Batak Toba	42	4	0.09524	--	--	--	--	--	--	4	0.09524
	Besemah	36	2	0.05556	--	--	6	0.16667	--	--	11	0.30556
	Semende	36	3	0.08333	--	--	1	0.02778	1	0.02778	2	0.05556
	Nias	59	2	0.03390	--	--	--	--	--	--	8	0.13559
	Mentawai	128	--	--	--	--	--	--	--	--	32	0.25000
	Lebbo'	19	3	0.15789	--	--	3	0.15789	--	--	--	--
	Ma'anyan	159	14	0.08805	--	--	--	--	3	0.01887	2	0.01258
	SK Dayak	64	3	0.04688	--	--	--	--	2	0.03125	1	0.01563
	Java	51	--	--	--	--	--	--	--	--	3	0.05882
	Bali	487	5	0.01027	2	0.00411	11	0.02259	--	--	14	0.02875
Eastern Indonesia	Mandar	54	--	--	--	--	14	0.25926	--	--	5	0.09259
	Kajang	46	3	0.06522	--	--	5	0.10870	--	--	--	--
	Toraja	50	5	0.10000	1	0.02000	10	0.20000	--	--	1	0.02000
	Bugis	50	3	0.06000	1	0.02000	6	0.12000	--	--	5	0.10000
	Bajo	27	2	0.07407	1	0.03704	2	0.07407	--	--	3	0.11111
	Sumba	634	16	0.02524	12	0.01893	18	0.02839	6	0.00946	71	0.11199
	Flores	469	13	0.02772	--	--	11	0.02345	1	0.00213	23	0.04904
	Lembata	92	7	0.07609	1	0.01087	8	0.08696	--	--	8	0.08696
	Pantar	29	--	--	--	--	1	0.03448	--	--	2	0.06897
	Alor	23	2	0.08696	1	0.04348	2	0.08696	--	--	1	0.04348
	North Maluku	224	21	0.09375	19	0.08482	11	0.04911	--	--	7	0.03125

Table S7 (continued). Mitochondrial haplogroup sharing frequencies between Malagasy and Indonesian populations.

			Shared haplogroups				**Not shared haplogroups**					
Population		**n ind**	**M32**		**Q1**		**M23 (WE)**		**Other Asian Hg**		**African Hg**	
			n	**freq**	**n**	**freq**	**n**	**freq**	**n**	**freq**	**n**	**freq**
Malagasy		529	23	0.04348	1	0.00189	36	0.06805	--	--	253	0.47826
Western Indonesia	Gayo	62	--	--	--	--	--	--	58	0.93548	--	--
	Batak Toba	42	--	--	--	--	--	--	34	0.80952	--	--
	Besemah	36	--	--	--	--	--	--	17	0.47222	--	--
	Semende	36	--	--	--	--	--	--	29	0.80556	--	--
	Nias	59	--	--	--	--	--	--	49	0.83051	--	--
	Mentawai	128	--	--	--	--	--	--	96	0.75000	--	--
	Lebbo'	19	--	--	--	--	--	--	13	0.68421	--	--
	Ma'anyan	159	--	--	--	--	--	--	140	0.88050	--	--
	SK Dayak	64	--	--	--	--	--	--	58	0.90625	--	--
	Java	51	1	0.01961	--	--	--	--	47	0.92157	--	--
	Bali	487	--	--	3	0.00616	--	--	452	0.92813	--	--
Eastern Indonesia	Mandar	54	--	--	--	--	--	--	35	0.64815	--	--
	Kajang	46	--	--	--	--	--	--	38	0.82609	--	--
	Toraja	50	--	--	2	0.04000	--	--	31	0.62000	--	--
	Bugis	50	--	--	1	0.02000	--	--	34	0.68000	--	--
	Bajo	27	--	--	2	0.07407	--	--	17	0.62963	--	--
	Sumba	634	--	--	18	0.02839	--	--	493	0.77760	--	--
	Flores	469	--	--	25	0.05330	--	--	396	0.84435	--	--
	Lembata	92	--	--	7	0.07609	--	--	61	0.66304	--	--
	Pantar	29	--	--	9	0.31034	--	--	17	0.58621	--	--
	Alor	23	--	--	5	0.21739	--	--	12	0.52174	--	--
	North Maluku	224	--	--	25	0.11161	--	--	141	0.62946	--	--

Table S8. Mitochondrial haplotype sharing between Malagasy and Indonesian populations.

			Shared Haplotypes													
	Population	**n ind**	**B4+16261 (H1)**[1]		**B4+16261 (H2)**[2]		**B4a1a1 (H3)**[3]		**B4a1a1 (H4)**[4]		**M7c1a4a (H5)**[5]		**M7c1a4a (H6)**[6]		**M7c1a4a (H7)**[7]	
			n	**freq**	**n**	**freq**	**n**	**freq**	**n**	**freq**	**n**	**freq**	**n**	**freq**	**n**	**freq**
	Malagasy*	398	1	0.00251	3	0.00754	21	0.05276	50	0.12563	2	0.00503	20	0.05025	1	0.00251
Western Indonesia	Gayo	62	--	--	--	--	--	--	--	--	1	0.01613	1	0.01613	--	--
	Batak Toba	42	--	--	4	0.09524	--	--	--	--	--	--	1	0.02381	--	--
	Besemah	36	1	0.02778	1	0.02778	--	--	--	--	--	--	11	0.30556	--	--
	Semende	36	--	--	2	0.05556	--	--	--	--	--	--	2	0.05556	--	--
	Nias	59	--	--	1	0.01695	--	--	--	--	1	0.01695	5	0.08475	--	--
	Mentawai	128	--	--	--	--	--	--	--	--	--	--	7	0.05469	--	--
	Lebbo'	19	--	--	2	0.10526	--	--	--	--	--	--	--	--	--	--
	Ma'anyan	159	--	--	13	0.08176	--	--	--	--	--	--	2	0.01258	--	--
	SK Dayak	64	--	--	3	0.04688	--	--	--	--	1	0.01563	--	--	--	--
	Java	51	--	--	--	--	--	--	--	--	1	0.01961	1	0.01961	1	0.01961
	Bali	487	--	--	2	0.00411	--	--	--	--	--	--	5	0.01027	--	--
Eastern Indonesia	Mandar	54	--	--	--	--	--	--	--	--	--	--	3	0.05556	--	--
	Kajang	46	3	0.06522	--	--	--	--	--	--	--	--	--	--	--	--
	Toraja	50	--	--	5	0.10000	--	--	--	--	--	--	1	0.02000	--	--
	Bugis	50	2	0.04000	1	0.02000	--	--	1	0.02000	--	--	4	0.08000	--	--
	Bajo	27	--	--	2	0.07407	--	--	--	--	--	--	2	0.07407	--	--
	Sumba	634	3	0.00473	8	0.01262	1	0.00158	--	--	1	0.00158	35	0.05521	--	--
	Flores	469	--	--	9	0.01919	--	--	--	--	--	--	15	0.03198	--	--
	Lembata	92	--	--	7	0.07609	--	--	--	--	--	--	2	0.02174	--	--
	Pantar	29	--	--	--	--	--	--	--	--	--	--	2	0.06897	--	--
	Alor	23	1	0.04348	--	--	--	--	--	--	--	--	--	--	--	--
	North Maluku	224	--	--	14	0.06250	--	--	11	0.04911	1	0.00446	3	0.01339	--	--

Note : **Malagasy individuals from Tofanelli etal, (2009) are not included due to their shorter sequences.*

[1]*B4+16261 (H1)*	*16189C, 16217C, 16261T, 16519C*
[2]*B4+16261 (H2)*	*16182C, 16183C, 16189C, 16217C, 16261T, 16519C*
[3]*B4a1a1 (H3)*	*16189C, 16217C, 16247G, 16261T, 16519C*
[4]*B4a1a1 (H4)*	*16182C, 16183C, 16189C, 16217C*
[5]*M7c1a4a (H5)*	*16223T, 16362C, 16519C*
[6]*M7c1a4a (H6)*	*16223T, 16295T, 16362C, 16519C*
[7]*M7c1a4a (H7)*	*16093C, 16223T, 16295T, 16362C, 16519C*
[8]*M32c (H8)*	*16086C, 16148T, 16223T, 16259T*
[9]*E1a1a (H9)*	*16223T, 16291T, 16362C, 16390A, 16519C*
[10]*Q1 (H10)*	*16129A, 16144C, 16148T, 16192T, 16223T, 16241G, 16265C, 16274A, 16311C, 16343G, 16362C*

Table S8 (continued). Mitochondrial haplotype sharing between Malagasy and Indonesian populations.

			Shared Haplotypes						Not Shared Haplotypes					
Population		n ind	M32c (H8)[8]		E1a1a (H9)[9]		Q1 (H10)[10]		M23 (WE)		Other Asian Hp		African Hp	
			n	freq	n	freq	n	freq	n	freq	n	freq	n	freq
Malagasy*		398	11	0.02764	1	0.00251	1	0.00251	29	0.07286	56	0.14070	202	0.50754
Western Indonesia	Gayo	62	--	--	--	--	--	--	--	--	60	0.96774	--	--
	Batak Toba	42	--	--	--	--	--	--	--	--	37	0.88095	--	--
	Besemah	36	--	--	6	0.16667	--	--	--	--	17	0.47222	--	--
	Semende	36	--	--	1	0.02778	--	--	--	--	31	0.86111	--	--
	Nias	59	--	--	--	--	--	--	--	--	52	0.88136	--	--
	Mentawai	128	--	--	--	--	--	--	--	--	121	0.94531	--	--
	Lebbo'	19	--	--	3	0.15789	--	--	--	--	14	0.73684	--	--
	Ma'anyan	159	--	--	--	--	--	--	--	--	144	0.90566	--	--
	SK Dayak	64	--	--	--	--	--	--	--	--	60	0.93750	--	--
	Java	51	1	0.01961	--	--	--	--	--	--	47	0.92157	--	--
	Bali	487	--	--	11	0.02259	--	--	--	--	469	0.96304	--	--
Eastern Indonesia	Mandar	54	--	--	6	0.11111	--	--	--	--	45	0.83333	--	--
	Kajang	46	--	--	5	0.10870	--	--	--	--	38	0.82609	--	--
	Toraja	50	--	--	8	0.16000	--	--	--	--	36	0.72000	--	--
	Bugis	50	--	--	3	0.06000	--	--	--	--	39	0.78000	--	--
	Bajo	27	--	--	1	0.03704	--	--	--	--	22	0.81481	--	--
	Sumba	634	--	--	9	0.01420	--	--	--	--	577	0.91009	--	--
	Flores	469	--	--	5	0.01066	1	0.00213	--	--	439	0.93603	--	--
	Lembata	92	--	--	6	0.06522	--	--	--	--	77	0.83696	--	--
	Pantar	29	--	--	--	--	--	--	--	--	27	0.93103	--	--
	Alor	23	--	--	1	0.04348	--	--	--	--	21	0.91304	--	--
	North Maluku	224	--	--	5	0.02232	--	--	--	--	190	0.84821	--	--

Note : **Malagasy individuals from Tofanelli etal, (2009) are not included due to their shorter sequences.*

[1] *B4+16261 (H1)*	*16189C, 16217C, 16261T, 16519C*
[2] *B4+16261 (H2)*	*16182C, 16183C, 16189C, 16217C, 16261T, 16519C*
[3] *B4a1a1 (H3)*	*16189C, 16217C, 16247G, 16261T, 16519C*
[4] *B4a1a1 (H4)*	*16182C, 16183C, 16189C, 16217C*
[5] *M7c1a4a (H5)*	*16223T, 16362C, 16519C*
[6] *M7c1a4a (H6)*	*16223T, 16295T, 16362C, 16519C*
[7] *M7c1a4a (H7)*	*16093C, 16223T, 16295T, 16362C, 16519C*
[8] *M32c (H8)*	*16086C, 16148T, 16223T, 16259T*
[9] *E1a1a (H9)*	*16223T, 16291T, 16362C, 16390A, 16519C*
[10] *Q1 (H10)*	*16129A, 16144C, 16148T, 16192T, 16223T, 16241G, 16265C, 16274A, 16311C, 16343G, 16362C*

Table S9. Mitochondrial DNA F_{ST} distances between Malagasy and Indonesian populations.

Population		Fst	p-value
Sumba	EI	0.15078	<0.01
Flores	EI	0.15433	<0.01
North Maluku	EI	0.15873	<0.01
Bugis	EI	0.16618	<0.01
Bali	WI	0.17206	<0.01
Bajo	EI	0.17262	<0.01
Gayo	WI	0.17399	<0.01
SK Dayak	WI	0.17835	<0.01
Lembata	EI	0.17868	<0.01
Toraja	EI	0.18076	<0.01
Java	WI	0.18135	<0.01
Alor	EI	0.18272	<0.01
Batak Toba	WI	0.19116	<0.01
Ma'anyan	WI	0.19201	<0.01
Besemah	WI	0.19431	<0.01
Mandar	EI	0.19549	<0.01
Semende	WI	0.20944	<0.01
Pantar	EI	0.21693	<0.01
Lebbo'	WI	0.21835	<0.01
Mentawai	WI	0.22808	<0.01
Kajang	EI	0.23728	<0.01
Nias	WI	0.26224	<0.01

Note.

U test p value < 0.05; EI=Eastern Indonesia; WI= Western Indonesia

Annex D. Supplementary information from article entitled “West Eurasian genetic influences in the Indonesian archipelago”

(Published in Quaternary International, 416, September 19th 2016. doi:10.1016/j.quaint. 2015.06.048)

Pradiptajati Kusuma[a,b], Murray P. Cox[c], Nicolas Brucato[d], Herawati Sudoyo[b], Thierry Letellier[a], François-Xavier Ricaut[a*]

[a] Laboratoire d’Anthropologie Moléculaire et Imagérie de Synthèse UMR-5288, Université de Toulouse, Toulouse, France
[b] Laboratory of Human Genetic Diversity, Eijkman Institute for Molecular Biology, Jakarta, Indonesia
[c] Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
[d] Center for Linguistics, University of Leiden, Leiden, Netherlands

Table S1. Frequency of Western Eurasian origin Y-haplogroups in Indonesia

		Western Indonesia									
West Eurasian Origin		**Batak Toba 37**	**Besemah 38**	**Semende 37**	**Nias 60**	**Mentawai 74**	**Lebbo' 15**	**Ma'anyan 90**	**EK Dayak 85**	**Java 61**	**Bali 634**
Q1a2	**Q-M346**	--	--	--	--	--	--	--	--	--	0.32
R	**M207**	--	--	--	--	--	--	--	--	--	--
R1a	**R-M17**	--	--	--	--	--	--	1.11	1.18	9.84	1.42
R1b1a2	**R-M269**	--	--	--	--	--	--	--	--	--	0.63
R2a	**R-M124**	2.70	--	--	--	--	--	--	1.18	--	1.74
H1	**H-M69***	--	--	--	--	--	--	--	--	--	2.68
H1a	**H-M52**	--	--	--	--	--	--	--	1.18	--	0.16
H1b1	**H-Apt**	--	--	--	--	--	--	--	--	--	0.16
J	**J-M304**	--	--	--	--	--	--	--	--	--	0.32
J1	**J-M267**	--	--	--	--	--	--	--	--	--	0.95
J2	**J-M172***	--	--	--	--	--	--	--	--	4.92	--
J2b	**J-M12**	--	--	--	--	--	--	--	--	--	0.63
T1a	**T-M70**	--	--	--	--	--	--	--	--	--	--
L	**L-M20***	--	--	--	--	--	--	--	--	--	2.05
L1a	**L-M76**	--	--	--	--	--	--	--	--	--	--
Total Frequency		**2.70**	**0**	**0**	**0**	**0**	**0**	**1.11**	**3.53**	**14.75**	**11.04**
Number of Individuals		(1/37)						(1/90)	(3/85)	(9/61)	(70/634)
Asian Origin (Total Frequency)		97.30	100	100	100	100	100	98.89	96.47	85.25	88.96

Table S1 (continued). Frequency of Western Eurasian origin Y-haplogroups in Indonesia

		Eastern Indonesia							**Indonesia**	**Haplogroups likely origin/high frequency**
		Mandar	**Bajo**	**Sumba**	**Flores**	**Lembata**	**Alor**	**Maluku**		
West Eurasian Origin		**54**	**27**	**349**	**388**	**89**	**27**	**30**	**2095**	
Q1a2	**Q-M346**	--	--	--	--	--	--	--	0.10	South Siberia, South Asia, West Asia
R	**M207**	--	14.81	--	--	--	--	--	0.19	Central Asia, South Asia
R1a	**R-M17**	1.85	--	--	--	--	--	--	0.86	Central Asia, West Asia, West India, Pakistan
R1b1a2	**R-M269**	--	--	--	--	--	--	--	0.19	Europe
R2a	**R-M124**	1.85	--	--	--	--	--	--	0.67	Central Asia, South Asia, West Asia, Sri Lanka
H1	**H-M69***	--	--	--	--	--	--	--	0.81	India, Pakistan
H1a	**H-M52**	--	--	--	--	--	--	--	0.10	India
H1b1	**H-Apt**	--	--	--	--	--	--	--	0.05	India
J	**J-M304**	--	--	--	--	--	--	--	0.10	Southwest Asia, Near east
J1	**J-M267**	--	--	--	--	--	--	--	0.29	Near East, West Asia
J2	**J-M172***	1.85	--	--	--	--	--	--	0.19	Near East, West Asia
J2b	**J-M12**	--	--	--	--	--	--	--	0.19	Europe
T1a	**T-M70**	--	7.41	--	--	--	--	--	0.10	Near East, India
L	**L-M20***	--	--	--	--	--	--	--	0.62	India, Pakistan
L1a	**L-M76**	--	3.70	--	--	--	--	--	0.05	India
Total Frequency		**5.56**	**25.93**	**0**	**0**	**0**	**0**	**0**	**4.49**	
Number of Individuals		(3/54)	(7/27)						(94/2095)	
Asian Origin (Total Frequency)		94.44	74.07	100	100	100	100	100	95.51	

Table S2. Frequency of Western Eurasian origin mitochondrial haplogroups in Indonesian populations.

	Western Indonesia										
	Gayo	**Batak Toba**	**Besemah**	**Semende**	**Nias**	**Mentawai**	**Lebbo'**	**Ma'anyan**	**SK Dayak**	**Java**	**Bali**
West Eurasian Origin	**62**	**42**	**36**	**36**	**59**	**128**	**19**	**159**	**64**	**51**	**487**
M18	1.61	--	--	--	--	--	--	--	--	--	--
M2	--	--	--	--	--	--	--	0.63	--	--	--
M32c	--	--	--	--	--	--	--	--	--	1.96	--
M33	1.61	--	--	--	--	--	--	--	--	5.88	--
M35a	1.61	--	--	--	--	--	--	1.89	--	--	--
M40	--	--	--	--	--	--	--	--	--	--	--
M41a	1.61	--	--	--	--	--	--	--	--	--	--
M45	--	--	--	--	--	--	--	--	--	--	--
M5	--	--	--	--	--	--	--	0.63	--	--	2.05
M81	--	--	--	--	--	--	--	--	--	--	0.21
N5	--	--	--	--	--	--	--	--	--	--	0.21
U2	3.23	--	--	--	--	--	--	--	--	--	--
U7	3.23	--	--	--	--	--	--	--	--	--	--
Total Frequency	**12.90**	**0**	**0**	**0**	**0**	**0**	**0**	**3.14**	**0**	**7.84**	**2.46**
Number of Individuals	(8/62)							(5/159)		(4/51)	(12/487)
Asian Origin Total Frequency	87.10	100	100	100	100	100	100	96.86	100	92.16	97.54

Table S2 (continued). Frequency of Western Eurasian origin mitochondrial haplogroups in Indonesian populations.

	Eastern Indonesia											**Indonesia**	**Haplogroups likely origin/high frequency**
	Mandar	**Kajang**	**Toraja**	**Bugis**	**Bajo**	**Sumba**	**Flores**	**Lembata**	**Pantar**	**Alor**	**North Maluku**		
West Eurasian Origin	**54**	**46**	**50**	**50**	**27**	**634**	**469**	**92**	**29**	**23**	**224**	**2841**	
M18	--	--	--	--	--	--	--	--	--	--	--	0.04	India
M2	--	--	--	--	--	--	--	--	--	--	--	0.04	India
M32c	--	--	--	--	--	--	--	--	--	--	--	0.04	West Eurasian
M33	--	--	--	--	--	--	--	--	--	--	--	0.14	India, Nepal
M35a	--	--	--	--	--	--	--	--	--	--	--	0.14	India
M40	--	--	--	2.00	--	--	0.21	--	6.90	--	--	0.14	India
M41a	--	--	--	--	--	--	--	--	--	--	--	0.04	India
M45	--	--	2.00	--	--	--	--	--	--	--	--	0.04	India, Nepal
M5	--	--	--	--	--	--	1.71	--	3.45	--	--	0.70	India
M81	--	--	--	--	--	--	--	--	--	--	--	0.04	India
N5	--	--	--	--	--	0.16	--	--	--	--	--	0.07	India
U2	--	--	--	--	--	--	--	--	--	--	--	0.07	India, Pakistan
U7	--	--	--	--	--	--	--	--	--	--	--	0.07	Gujarat, Punjab
Total Frequency	**0**	**0**	**2.00**	**2.00**	**0**	**0.16**	**1.92**	**0**	**10.34**	**0**	**0**	**1.55**	
Number of Individuals			(1/50)	(1/50)		(1/634)	(9/469)		(3/29)			(44/2841)	
Asian Origin Total Frequency	100	100	98	96	100	99.84	98.08	100	89.66	100	100	98.45	

Table S3. Western Eurasian halogroup diversity of Y-chromosome and mitochondria.

			Western Eurasian Y-Haplogroups			Western Eurasian Mitchondrial Haplogroups		
No.	**Island**	**Population**	**Sample Size**	**Haplogroup Diversity (± S.D)**	**Number of Haplogroups**	**Sample Size**	**Haplogroup Diversity (± S.D)**	**Number of Haplogroups**
1	Sumatera	Gayo	*NA*	*NA*	*NA*	62	0.2422 ± 0.0722	7
2		Batak Toba	37	0.0541 ± 0.0505	1	42	--	--
3		Besemah	38	--	--	36	--	--
4		Semende	37	--	--	36	--	--
5		Nias	60	--	--	59	--	--
6		Mentawai	74	--	--	128	--	--
7	Borneo	Lebbo'	15	--	--	19	--	--
8		Ma'anyan	90	0.0222 ± 0.0216	1	159	0.0619 ± 0.0264	3
9		SK Dayak	*NA*	*NA*	*NA*	64	--	--
10		EK Dayak	85	0.0697 ± 0.0382	3	*NA*	*NA*	*NA*
11	Java	Java	61	0.2656 ± 0.0702	2	51	0.1498 ± 0.0658	2
12	Bali	Bali	634	0.2071 ± 0.0217	11	487	0.0483 ± 0.0135	3
13	Sulawesi	Mandar	54	0.1090 ± 0.0578	3	54	--	--
14		Kajang	*NA*	*NA*	*NA*	46	--	--
15		Toraja	*NA*	*NA*	*NA*	50	0.0400 ± 0.0380	1
16		Bugis	*NA*	*NA*	*NA*	50	0.0400 ± 0.0380	1
17		Bajo	27	0.4387 ± 0.1078	3	27	--	--
18	Lesser Sunda	Sumba	349	--	--	634	0.0032 ± 0.0031	1
19		Flores	388	--	--	469	0.0378 ± 0.0122	2
20		Lembata	89	--	--	92	--	--
21		Pantar	*NA*	*NA*	*NA*	29	0.1970 ± 0.0952	2
22		Alor	27	--	--	23	--	--
23	Maluku	North Maluku	*NA*	*NA*	*NA*	224	--	--
24		Maluku	30	--	--	*NA*	*NA*	*NA*

Notes:
-- : No Western Eurasian haplogroups found in population
NA : No data available

Annex E. Supplementary information from article entitled “Contrasting linguistic and genetic influences during the Austronesian settlement of Madagascar”

(Published in Scientific Reports, 6:26066, May 18th 2016. doi:10.1038/srep26066)

Pradiptajati Kusuma[1,2], Nicolas Brucato[1], Murray P. Cox[3], Denis Pierron[1], Harilanto Razafindrazaka[1], Alexander Adelaar[4], Herawati Sudoyo[2,5], Thierry Letellier[1], and François-Xavier Ricaut[1,*]

[1] Evolutionary Medicine Group, Laboratoire d’Anthropologie Moléculaire et Imagerie de Synthèse UMR-5288, Université de Toulouse, Toulouse, France

[2] Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular Biology, Jakarta, Indonesia

[3] Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

[4] Asia Institute, University of Melbourne, Melbourne, Australia

[5] Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia

Figure S1. Principal Component Analysis on all East Asian/Mainland Southeast Asian/Island Southeast Asian populations from the low density dataset.

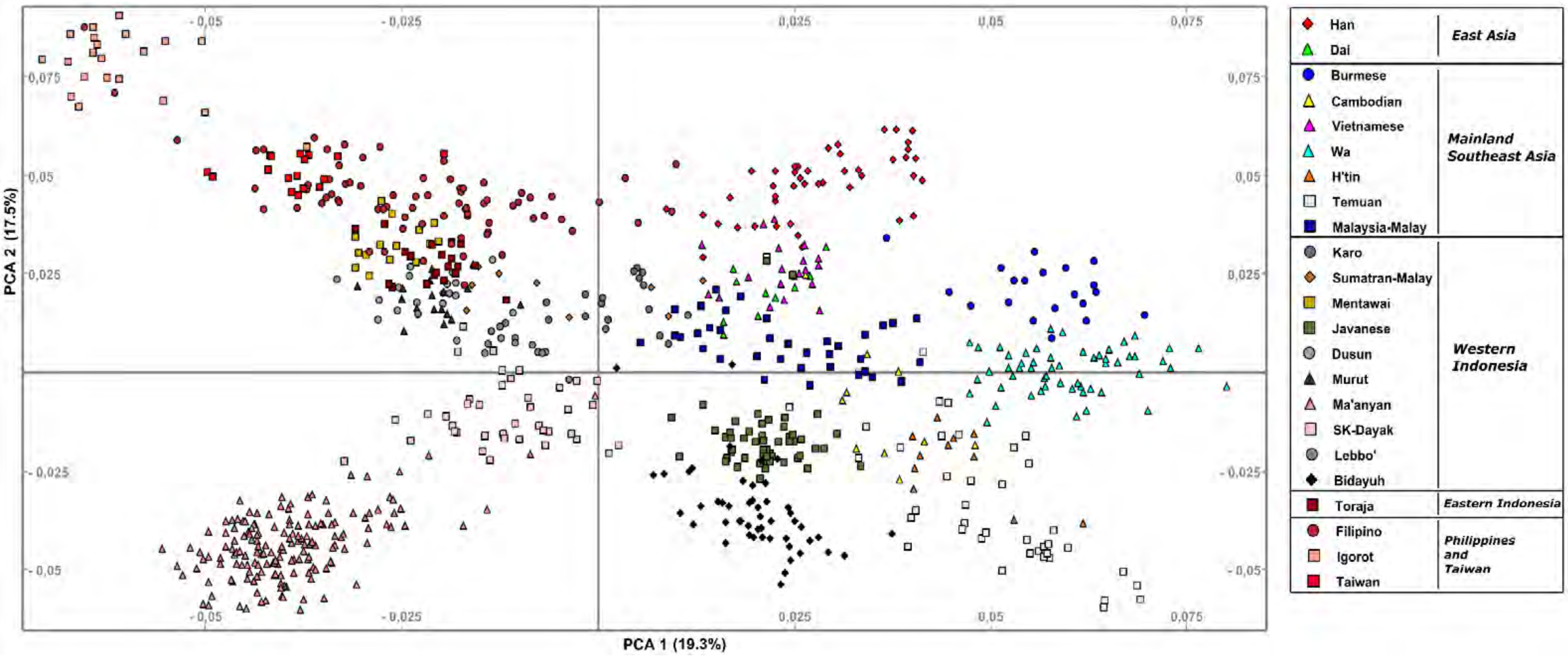


Figure S2. Cross-validation plot generated from the ADMIXTURE analysis.

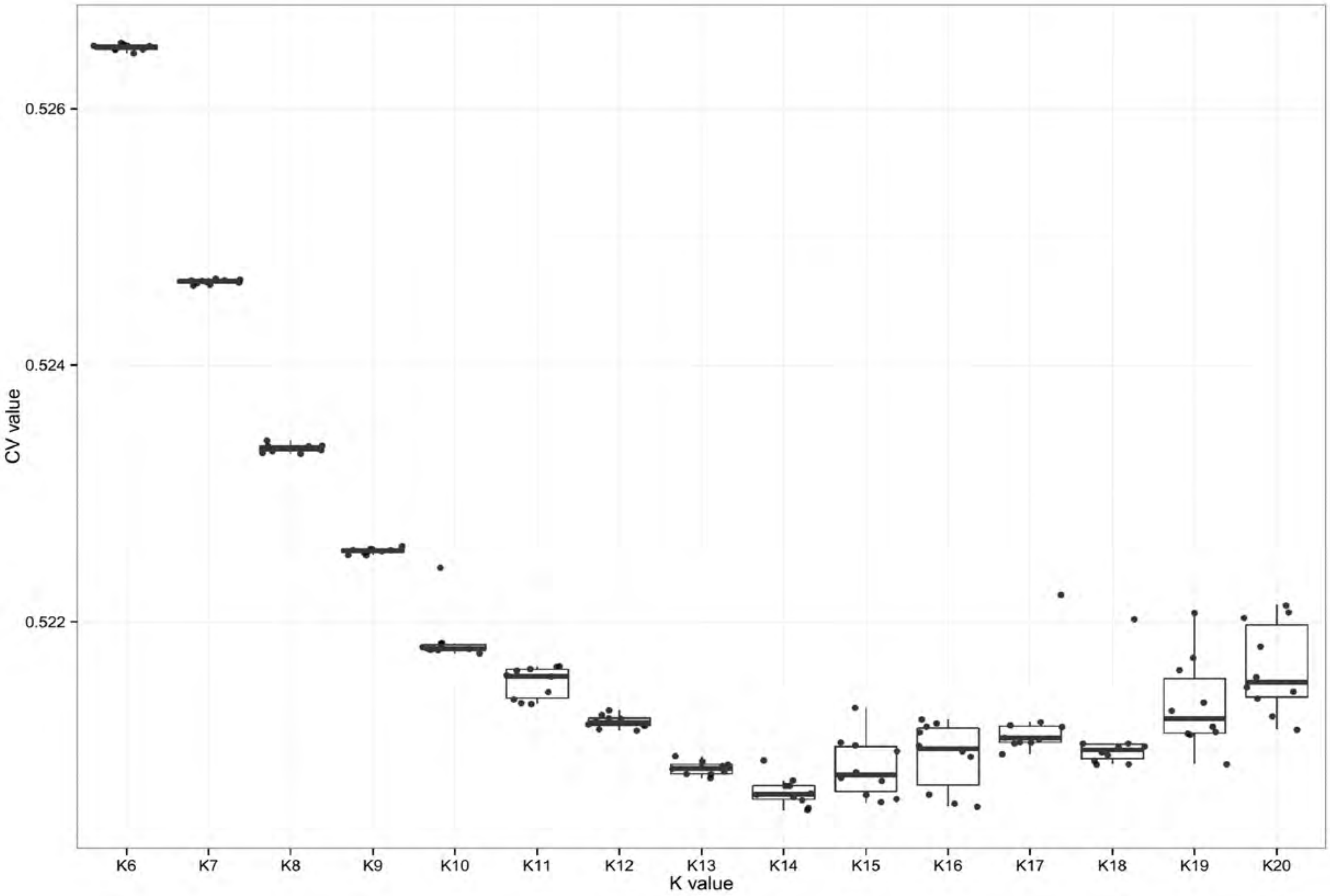


Figure S3. Runs of Homozygosity results showing that the Ma’anyan have comparable homozygosity relative to other Borneo populations.

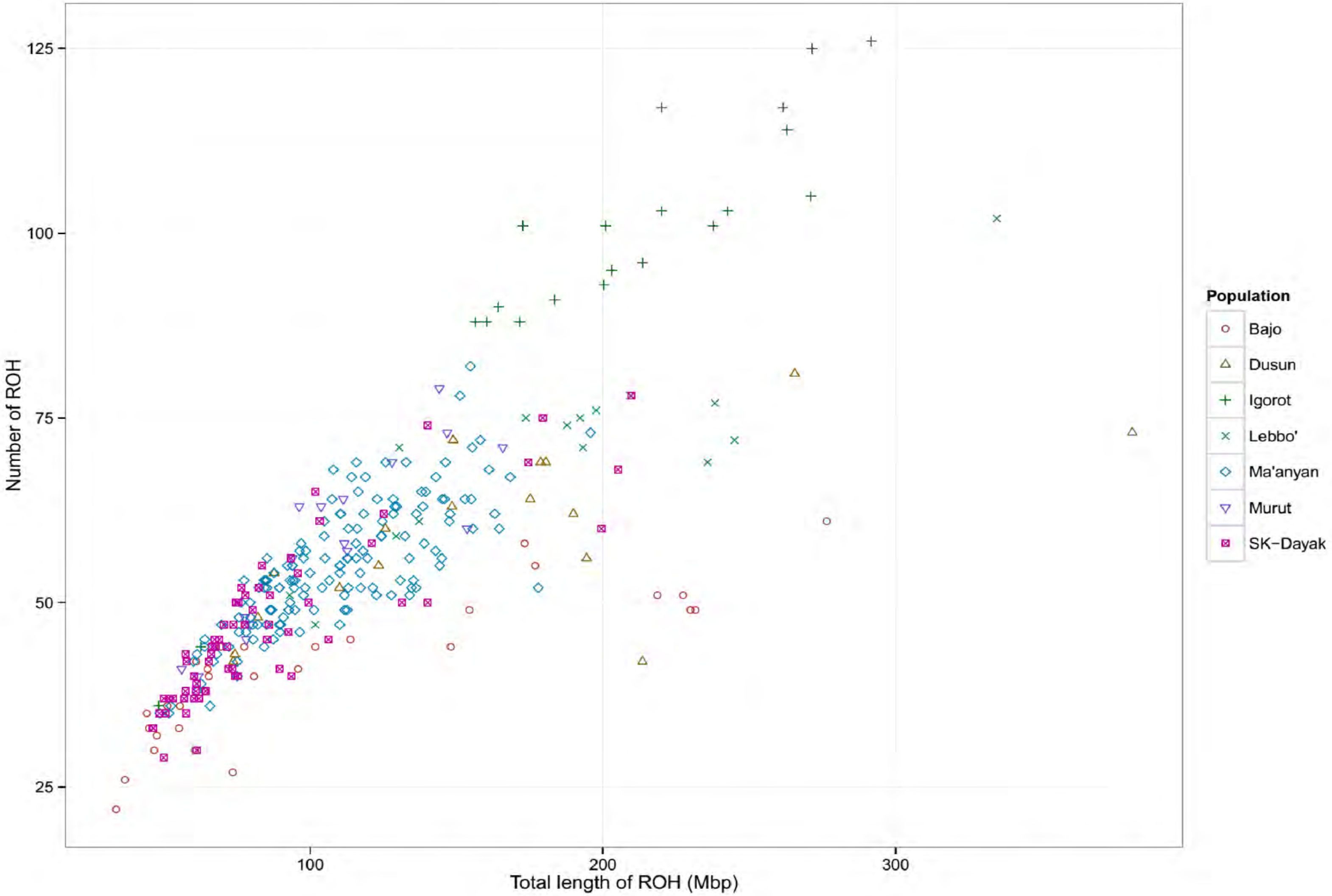


Figure S4. TreeMix analysis on the low density dataset with eight migration nodes showing no gene flow to or from the Ma'anyan.

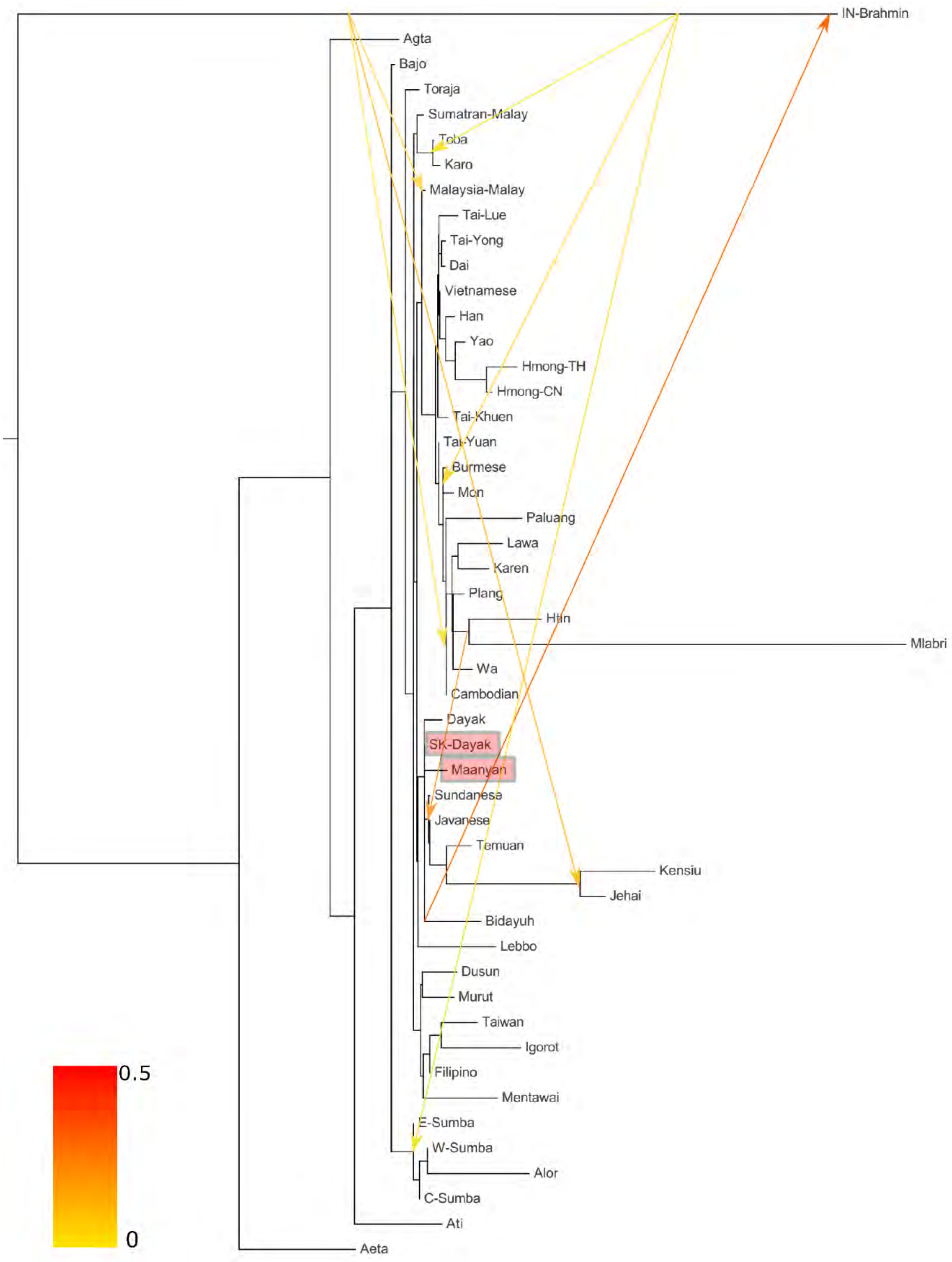


Figure S5. Shared Identity-By-Descent fragments between pairs of individuals in Southeast Asia, using nine filtering thresholds (from 10cM to 90cM). The maps were generated using Global Mapper v.15 software (http://www.bluemarblegeo.com/ products/global-mapper.php). The networks lines were generated using Cytoscape v.3.2.152 software.

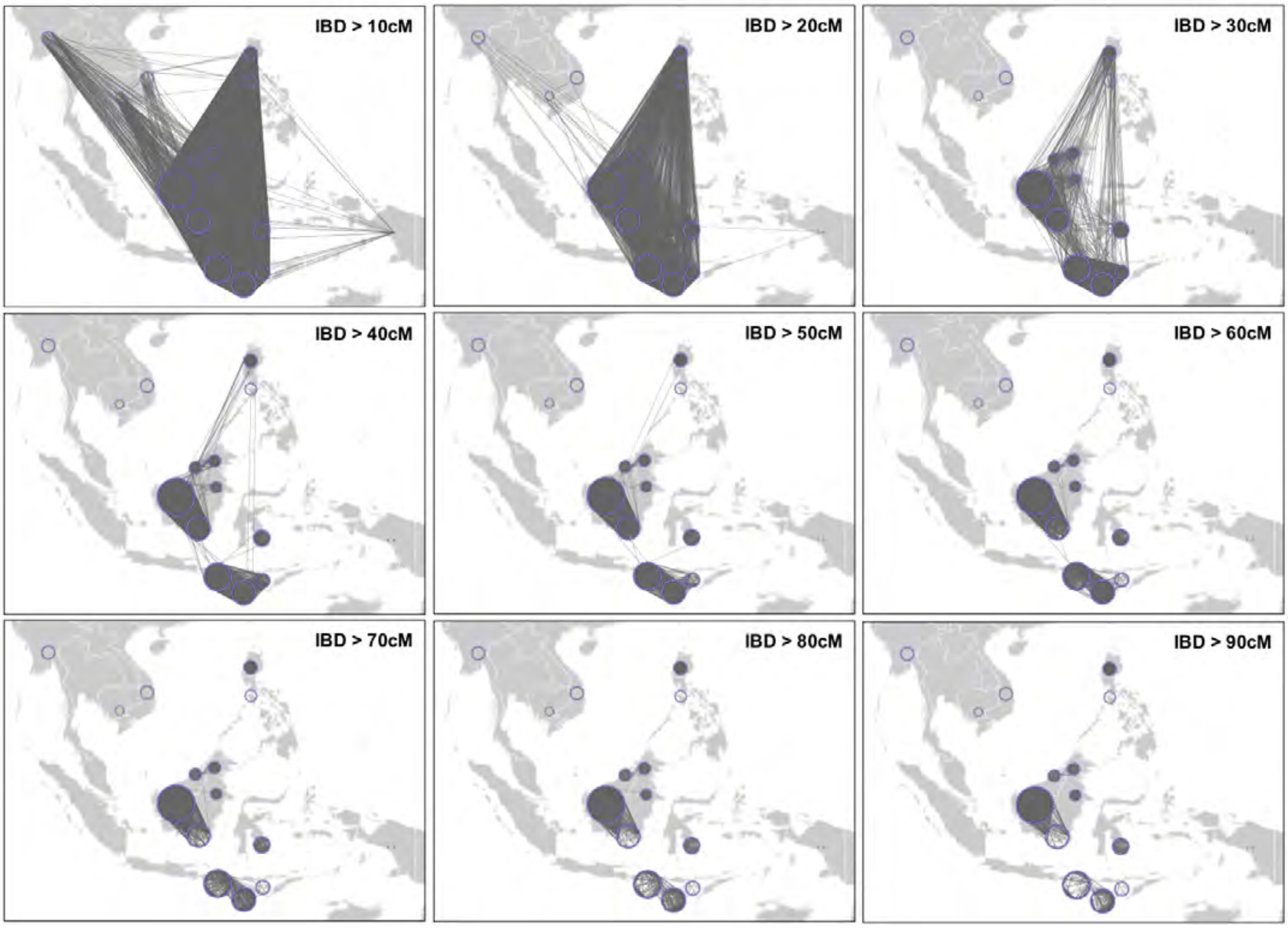


Figure S6. Principal Component Analysis on all populations in the low density dataset. Ma’anyan and South Kalimantan Dayak individuals are clustered together with other Island Southeast Asian individuals.

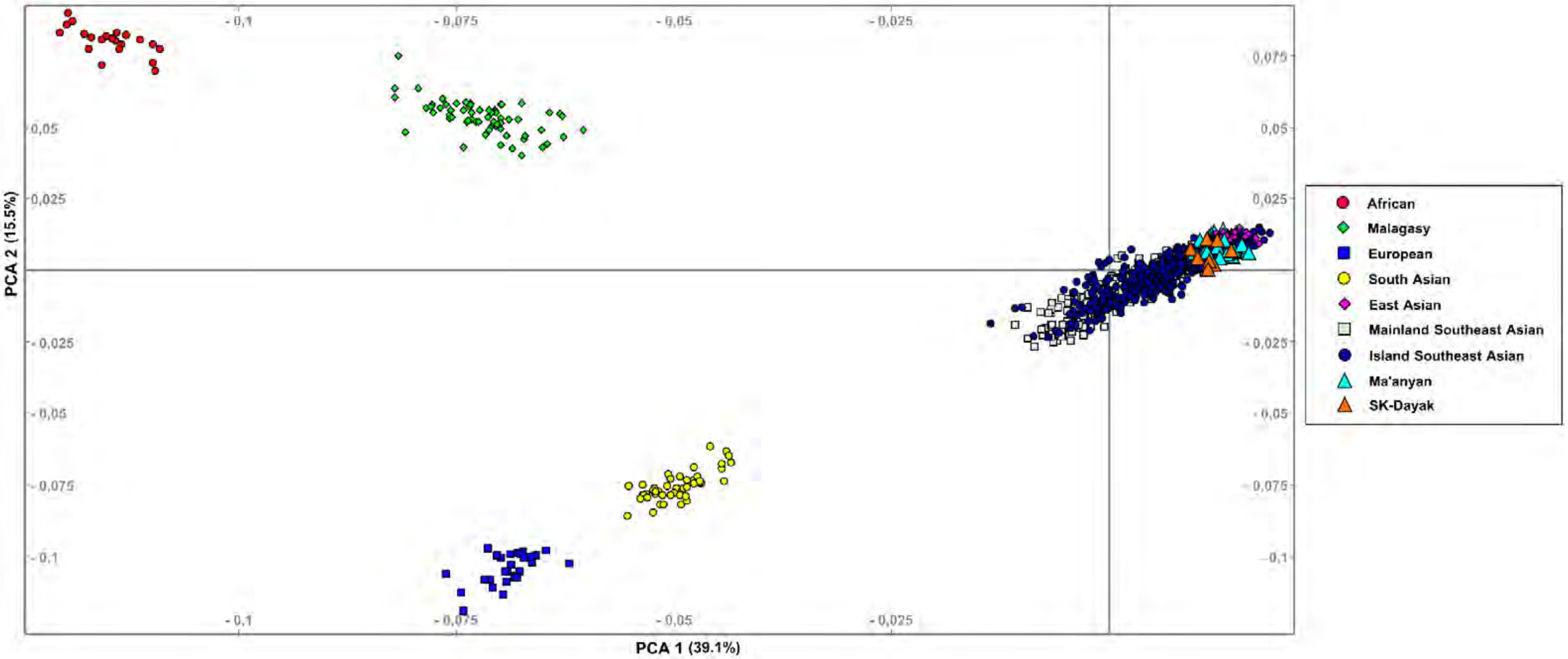


Figure S7. ADMIXTURE plots on the low density dataset from K = 2-14

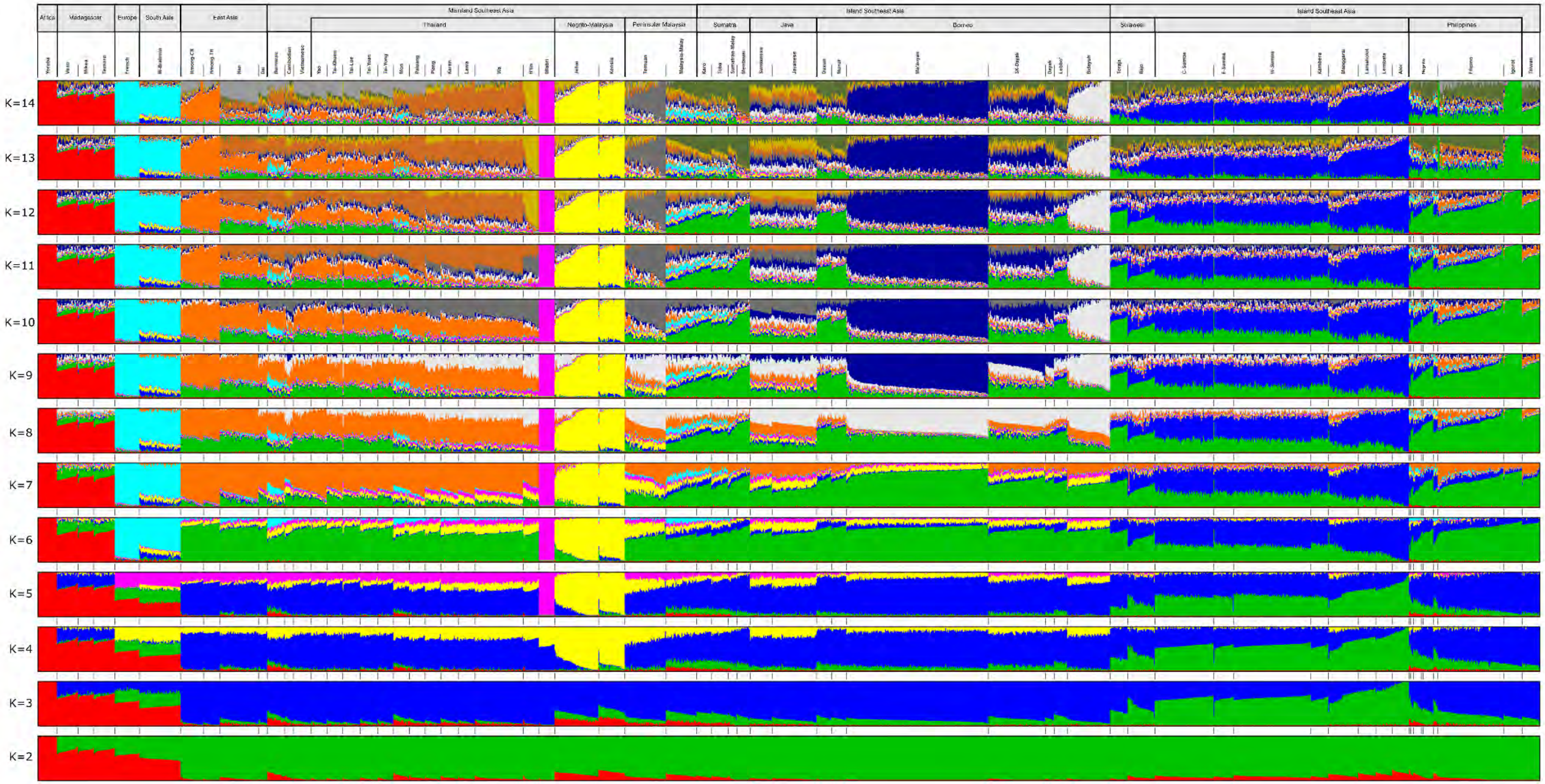


Figure S8. TreeMix analysis on the low density dataset with six migration nodes showing no specific Southeast Asian gene flow into Malagasy populations.

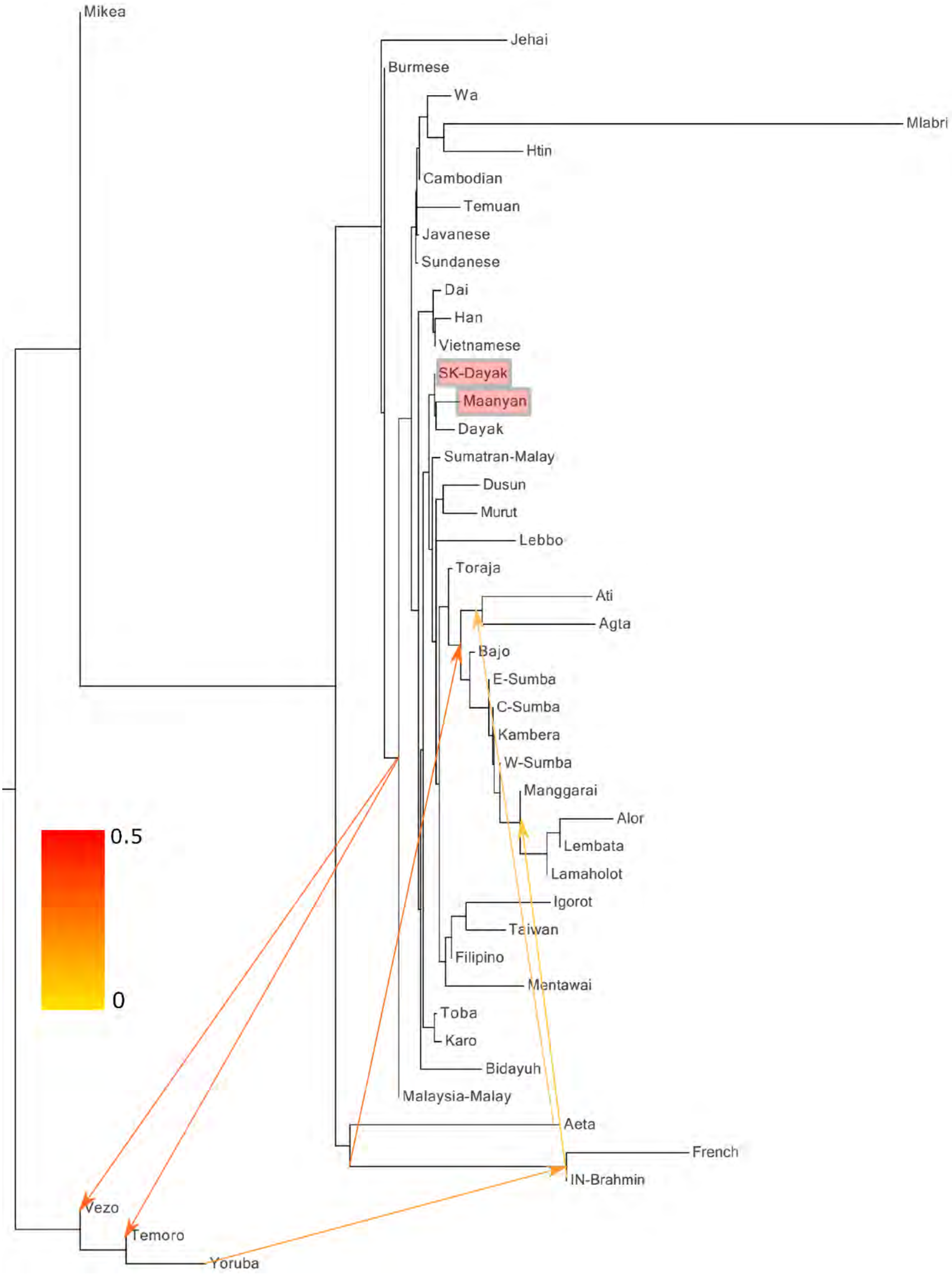


Table S1. Populations used in the study.

Population	No. of Individuals	Grouped as	Location	Region	References	LDD*	HDD*
Yoruba	21	Yoruba	Nigeria	Africa	Ref. 41	v	v
French	28	French	France	Europe	Ref. 41	v	v
Cambodian	10	Cambodian	Cambodia	East Asia	Ref. 41	v	v
Han	44	Han	China	East Asia	Ref. 41	v	v
Dai	10	Dai	China	East Asia	Ref. 41	v	v
Vezo	24	Vezo	Madagascar	Africa	Ref. 9	v	v
Mikea	18	Mikea	Madagascar	Africa	Ref. 9	v	v
Temoro	24	Temoro	Madagascar	Africa	Ref. 9	v	v
Lebbo	15	Lebbo	Borneo	Island Southeast Asia	Ref. 9	v	v
Bajo	31	Bajo	Sulawesi	Island Southeast Asia	Ref. 9	v	v
Brahmin	47	Brahmin	India	South Asia	Mörseburg etal., unpublished data	v	v
Vietnamese	20	Vietnamese	Singapore	Mainland Southeast Asia	Mörseburg etal., unpublished data	v	v
Burmese	20	Burmese	Singapore	Mainland Southeast Asia	Mörseburg etal., unpublished data	v	v
Dusun	17	Dusun	Borneo	Island Southeast Asia	Mörseburg etal., unpublished data	v	v
Murut	17	Murut	Brunei	Island Southeast Asia	Mörseburg etal., unpublished data	v	v
Luzon	12	Filipino	Philippines	Island Southeast Asia	Mörseburg etal., unpublished data	v	v
Visayas	4	Filipino	Philippines	Island Southeast Asia	Mörseburg etal., unpublished data	v	v
Igorot	21	Igorot	Philippines	Island Southeast Asia	Mörseburg etal., unpublished data	v	v
Ami	10	Taiwan-Aborigine	Taiwan	East Asia	Ref. 42	v	-
Atayal	10	Taiwan-Aborigine	Taiwan	East Asia	Ref. 42	v	-
Wa	56	Wa	China	East Asia	Ref. 42	v	-
Hmong	26	Hmong-CN	China	East Asia	Ref. 42	v	-
Hmong	20	Hmong-TH	Thailand	Mainland Southeast Asia	Ref. 42	v	-
Karen	20	Karen	Thailand	Mainland Southeast Asia	Ref. 42	v	-
Lawa	19	Lawa	Thailand	Mainland Southeast Asia	Ref. 42	v	-
Mlabri	18	Mlabri	Thailand	Mainland Southeast Asia	Ref. 42	v	-
Mon	19	Mon	Thailand	Mainland Southeast Asia	Ref. 42	v	-
Paluang	18	Paluang	Thailand	Mainland Southeast Asia	Ref. 42	v	-
Plang	18	Plang	Thailand	Mainland Southeast Asia	Ref. 42	v	-
TaiKhuen	19	TaiKhuen	Thailand	Mainland Southeast Asia	Ref. 42	v	-
TaiLue	20	TaiLue	Thailand	Mainland Southeast Asia	Ref. 42	v	-
H'tin	18	H'tin	Thailand	Mainland Southeast Asia	Ref. 42	v	-
TaiYuan	20	TaiYuan	Thailand	Mainland Southeast Asia	Ref. 42	v	-
TaiYong	18	TaiYong	Thailand	Mainland	Ref. 42	v	-

				Southeast Asia			
Yao	19	Yao	Thailand	Mainland Southeast Asia	Ref. 42	v	-
Jehai	50	Jehai	Malaysia	Mainland Southeast Asia	Ref. 42	v	-
Kensiu	30	Kensiu	Malaysia	Mainland Southeast Asia	Ref. 42	v	-
Malaysian-Malay	38	Malaysian-Malay	Malaysia	Mainland Southeast Asia	Ref. 42	v	-
Temuan	49	Temuan	Malaysia	Mainland Southeast Asia	Ref. 42	v	-
Bidayuh	50	Bidayuh	Malaysia	Island Southeast Asia	Ref. 42	v	-
Alorese	19	Alorese	Indonesia	Island Southeast Asia	Ref. 42	v	-
Dayak	12	Dayak	Indonesia	Island Southeast Asia	Ref. 42	v	-
Javanese-Jakarta	34	Javanese	Indonesia	Island Southeast Asia	Ref. 42	v	-
Javanese-Dieng	19	Javanese	Indonesia	Island Southeast Asia	Ref. 42	v	-
BatakKaro	17	Karo	Indonesia	Island Southeast Asia	Ref. 42	v	-
Lamaholot	20	Lamaholot	Indonesia	Island Southeast Asia	Ref. 42	v	-
Lembata	19	Lembata	Indonesia	Island Southeast Asia	Ref. 42	v	-
Sumatran-Malay	12	Sumatran-Malay	Indonesia	Island Southeast Asia	Ref. 42	v	-
Mentawai	15	Mentawai	Indonesia	Island Southeast Asia	Ref. 42	v	-
Manggarai	36	Manggarai	Indonesia	Island Southeast Asia	Ref. 42	v	-
Kambera	20	Kambera	Indonesia	Island Southeast Asia	Ref. 42	v	-
Sunda	25	Sunda	Indonesia	Island Southeast Asia	Ref. 42	v	-
BatakToba	20	Toba	Indonesia	Island Southeast Asia	Ref. 42	v	-
Toraja	20	Toraja	Indonesia	Island Southeast Asia	Ref. 42	v	-
Ayta	8	Ayta	Philippines	Island Southeast Asia	Ref. 42	v	-
Agta	8	Agta	Philippines	Island Southeast Asia	Ref. 42	v	-
Ati	23	Ati	Philippines	Island Southeast Asia	Ref. 42	v	-
Iraya	9	Iraya	Philippines	Island Southeast Asia	Ref. 42	v	-
Manobo	18	Manobo	Philippines	Island Southeast Asia	Ref. 42	v	-
Mamanwa	19	Mamanwa	Philippines	Island Southeast Asia	Ref. 42	v	-
Ilocano	20	Filipino	Philippines	Island Southeast Asia	Ref. 42	v	-
Visaya,Chabakano	20	Filipino	Philippines	Island Southeast Asia	Ref. 42	v	-
Tagalog	19	Filipino	Philippines	Island Southeast Asia	Ref. 42	v	-
Kodi	30	West Sumba	Indonesia	Island Southeast Asia	Cox, unpublished data	v	v
Lamboya	30	West Sumba	Indonesia	Island Southeast Asia	Cox, unpublished data	v	v
Loli	30	West Sumba	Indonesia	Island Southeast Asia	Cox, unpublished data	v	v
Wanokaka	31	West Sumba	Indonesia	Island Southeast Asia	Cox, unpublished data	v	v

Anakalang	30	Central Sumba	Indonesia	Island Southeast Asia	Cox, unpublished data	v	v
Wunga	30	Central Sumba	Indonesia	Island Southeast Asia	Cox, unpublished data	v	v
Mamboro	30	Central Sumba	Indonesia	Island Southeast Asia	Cox, unpublished data	v	v
Rindi	24	East Sumba	Indonesia	Island Southeast Asia	Cox, unpublished data	v	v
Ma'anyan	162	Ma'anyan	Borneo	Island Southeast Asia	this study	v	v
South Kalimantan Dayak	40	SK-Dayak	Borneo	Island Southeast Asia	this study	v	v

Note: * Data from this publication is used only on low density dataset (9,742 SNPs), not on high density dataset (311,871 SNPs).

Table S2. F_{ST} genetic distances between Ma’anyan and other populations in the low density dataset.

Populations	Ma'anyan
SK-Dayak	0.006
Sundanese	0.012
Javanese	0.012
Dayak	0.012
Cambodian	0.014
Sumatran-Malay	0.014
Malaysia-Malay	0.015
Filipino	0.015
Tai-Yuan	0.016
Vietnamese	0.017
Toraja	0.017
Dai	0.018
Tai-Yong	0.018
Toba	0.018
Tai-Khuen	0.019
Bajo	0.019
Karo	0.02
Murut	0.021
Burmese	0.022
Tai-Lue	0.022
Mon	0.022
Plang	0.022
Dusun	0.022
E-Sumba	0.023
Han	0.024
Wa	0.024
Kambera	0.024
Yao	0.025
C-Sumba	0.025
W-Sumba	0.025
Temuan	0.026

Populations	Ma'anyan
Bidayuh	0.026
Manobo	0.027
Manggarai	0.029
Taiwan	0.029
Karen	0.03
Hmong-CN	0.034
Lawa	0.034
Lebbo'	0.034
Mentawai	0.036
Lamaholot	0.038
Hmong-TH	0.041
Lembata	0.042
Igorot	0.042
Paluang	0.043
H'tin	0.043
Jehai	0.056
Ati	0.056
Agta	0.057
Alor	0.063
Kensiu	0.073
IN-Brahmin	0.075
Vezo	0.08
Temoro	0.081
Mikea	0.085
Iraya	0.087
Mamanwa	0.091
Aeta	0.094
French	0.108
Mlabri	0.154
Yoruba	0.164

Table S3. Results of f3-statistics, showing the lowest 50 Z-score combinations, on all populations in the low density dataset to test for Ma'anyan admixture. An f3-statistic is expected to be positive significant with a Z score >-2 if no admixture has taken place.

Target population	Population 1*	Population 2*	f3 value	SE	Z-score
Ma'anyan	Iraya	Mlabri	0,0012494	0,000643656	1,9411
Ma'anyan	Mlabri	Igorot	0,000964953	0,000388148	2,48604
Ma'anyan	Jehai	Igorot	0,000531593	0,00021105	2,51879
Ma'anyan	Mlabri	Taiwan	0,000980097	0,000315858	3,10296
Ma'anyan	Kensiu	Igorot	0,00100631	0,000251586	3,99986
Ma'anyan	Mlabri	Agta	0,00216078	0,000539446	4,00555
Ma'anyan	Manobo	Mlabri	0,00146867	0,000360337	4,07582
Ma'anyan	H'tin	Igorot	0,000966092	0,000236618	4,08291
Ma'anyan	Aeta	Mlabri	0,00309591	0,000719123	4,30511
Ma'anyan	Sumatran-Malay	Mlabri	0,00241122	0,000508989	4,73727
Ma'anyan	Yoruba	Igorot	0,00206575	0,000392053	5,26905
Ma'anyan	Ati	Mlabri	0,00227472	0,000409701	5,55215
Ma'anyan	Filipino	Mlabri	0,00159951	0,000285164	5,60909
Ma'anyan	Mamanwa	Mlabri	0,00186683	0,000306453	6,09173
Ma'anyan	Iraya	Jehai	0,00209051	0,000343169	6,0918
Ma'anyan	Dayak	Mlabri	0,00207119	0,000331207	6,25346
Ma'anyan	Mikea	Igorot	0,00222687	0,000353957	6,29137
Ma'anyan	Iraya	Kensiu	0,00263222	0,000415146	6,34046
Ma'anyan	Mentawai	Mlabri	0,00214312	0,000337205	6,35555
Ma'anyan	Yoruba	Mentawai	0,00242817	0,000378587	6,41376
Ma'anyan	Temuan	Igorot	0,00127519	0,000198279	6,43128
Ma'anyan	Yoruba	Murut	0,00219002	0,000336151	6,515
Ma'anyan	Mentawai	H'tin	0,00184931	0,000281734	6,56402
Ma'anyan	Iraya	H'tin	0,00217233	0,000324486	6,69467
Ma'anyan	Mlabri	E-Sumba	0,00266682	0,000382994	6,96309
Ma'anyan	Hmong-TH	Mlabri	0,00314585	0,000443036	7,10068
Ma'anyan	Dusun	Mlabri	0,00233471	0,000323972	7,20654
Ma'anyan	H'tin	Taiwan	0,00154934	0,000214634	7,21854
Ma'anyan	Vezo	Igorot	0,00229219	0,000316639	7,23912
Ma'anyan	Yoruba	Lebbo'	0,00283852	0,000388072	7,31441
Ma'anyan	Temoro	Igorot	0,00223352	0,000293669	7,60558
Ma'anyan	C-Sumba	Mlabri	0,00269372	0,000352449	7,64286
Ma'anyan	Yoruba	Mlabri	0,0036162	0,000471247	7,67367
Ma'anyan	Mlabri	Toraja	0,00196584	0,000254647	7,71987
Ma'anyan	Sumatran-Malay	Yoruba	0,00334507	0,000430849	7,76391
Ma'anyan	French	Mlabri	0,00425063	0,000535803	7,93319
Ma'anyan	H'tin	Toraja	0,00188538	0,000236766	7,96307
Ma'anyan	Mlabri	Tai-Lue	0,00356898	0,000443113	8,05434
Ma'anyan	Mlabri	Murut	0,00249194	0,00030931	8,05645
Ma'anyan	French	Igorot	0,00228156	0,000282665	8,07162

Ma'anyan	Iraya	Bidayuh	0,00209877	0,000258412	8,1218
Ma'anyan	H'tin	Agta	0,00258133	0,000316796	8,14823
Ma'anyan	Jehai	Mentawai	0,00198762	0,000243679	8,1567
Ma'anyan	Kensiu	Taiwan	0,00151931	0,000184827	8,22014
Ma'anyan	Banjar	Mlabri	0,00253314	0,000308105	8,22168
Ma'anyan	Jehai	Taiwan	0,00131213	0,000159384	8,23251
Ma'anyan	Alor	Mlabri	0,00314537	0,000380185	8,27324
Ma'anyan	Mlabri	Toba	0,00246328	0,000295879	8,32529
Ma'anyan	Hmong-TH	Yoruba	0,00404823	0,000485567	8,33712
Ma'anyan	H'tin	Dayak	0,0021435	0,000257094	8,3374

* This table shows the 50 lowest Z-score combinations of populations from all combinations produced in the f3-statistics analysis

Table S4. Pairwise F_{ST} values from the dataset of Asian SNPs

	Mikea	**Temoro**	**Vezo**	**Malagasy (average)**
Mikea	0	0,018	0,014	
Temoro	0,018	0	0,013	
Vezo	0,014	0,013	0	
Han	0,04	0,029	0,033	0,034
Dai	0,035	0,025	0,027	0,029
Cambodian	0,033	0,024	0,025	0,027
Burmese	0,035	0,025	0,029	0,030
Vietnamese	0,035	0,024	0,026	0,028
Dusun	0,04	0,031	0,031	0,034
Murut	0,039	0,029	0,03	0,033
Ma'anyan	0,032	0,024	0,025	0,027
SK-Dayak	0,027	0,018	0,02	0,022
Lebbo'	0,048	0,04	0,043	0,044
Bajo	0,033	0,023	0,026	0,027
C-Sumba	0,035	0,026	0,028	0,030
E-Sumba	0,041	0,029	0,033	0,034
W-Sumba	0,038	0,027	0,031	0,032
Filipino	0,03	0,02	0,023	0,024
Igorot	0,057	0,051	0,051	0,053

Annex F. Supplementary information from article entitled "The last sea nomads of the Indonesian archipelago: Genomic origins and dispersal"

(Submitted to European Journal of Human Genetics, under review)

Pradiptajati Kusuma[1,2], Nicolas Brucato[1], Murray P Cox[3], Thierry Letellier[1], Abdul Manan[4], Chandra Nuraini[5], Philippe Grangé[5], Herawati Sudoyo[2,6], François-Xavier Ricaut[1*]

Affiliations:
[1]Equipe de Médecine Evolutive, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse UMR-5288, Université de Toulouse, Toulouse, France
[2]Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular Biology, Jakarta, Indonesia
[3]Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
[4]University of Halu Oleo, Kendari, Indonesia
[5]UFR des Lettres, Langues, Arts et Sciences Humaines, Université de La Rochelle, La Rochelle, France
[6]Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia

Figure S1. Principal component analysis of the three Bajo communities (red symbols) within East Asia genetic diversity.

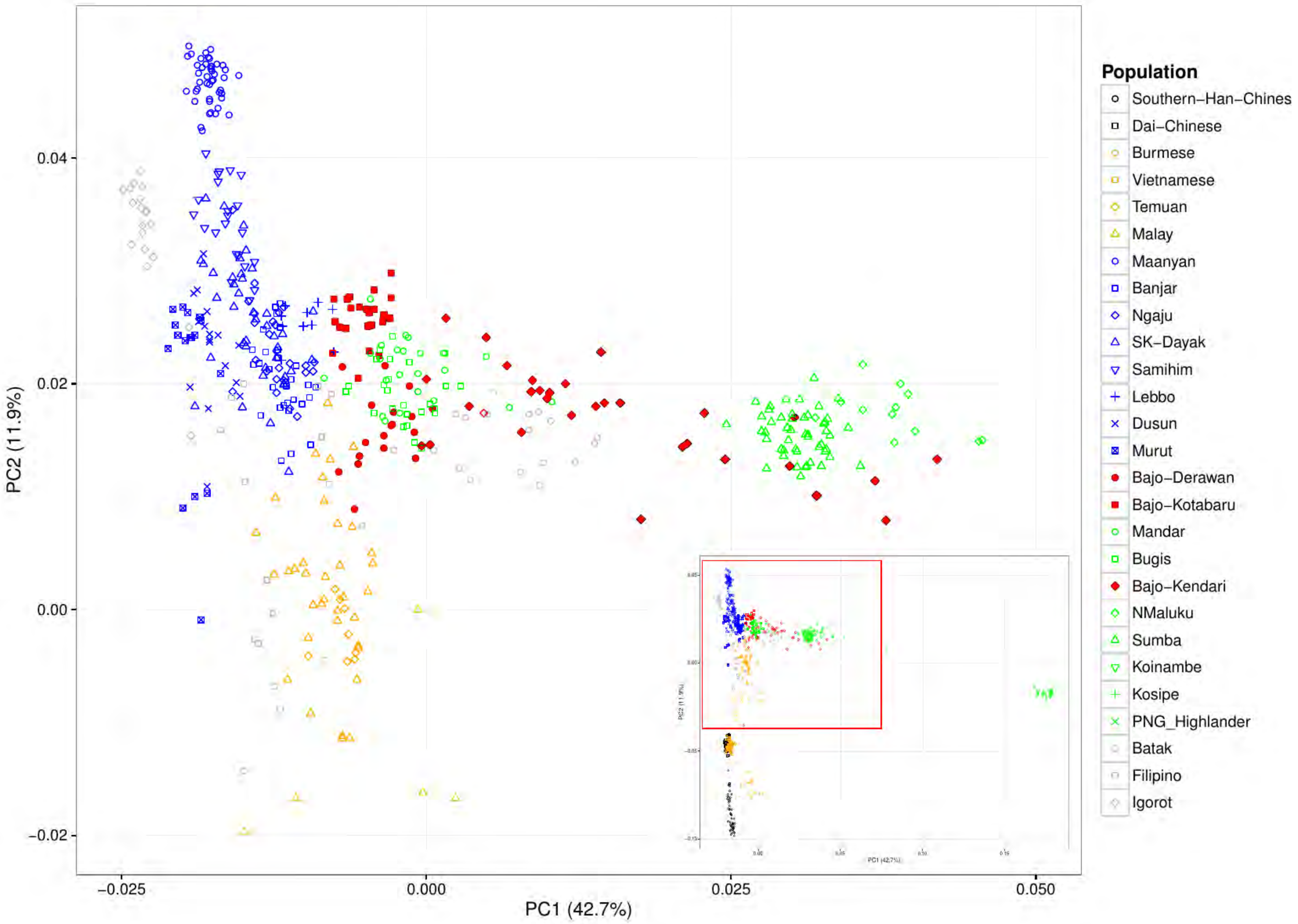


Figure S2. Clustering analysis using fineSTRUCTURE showing Bajo individuals clustered within each community (blue boxes) and no overlap between communities.

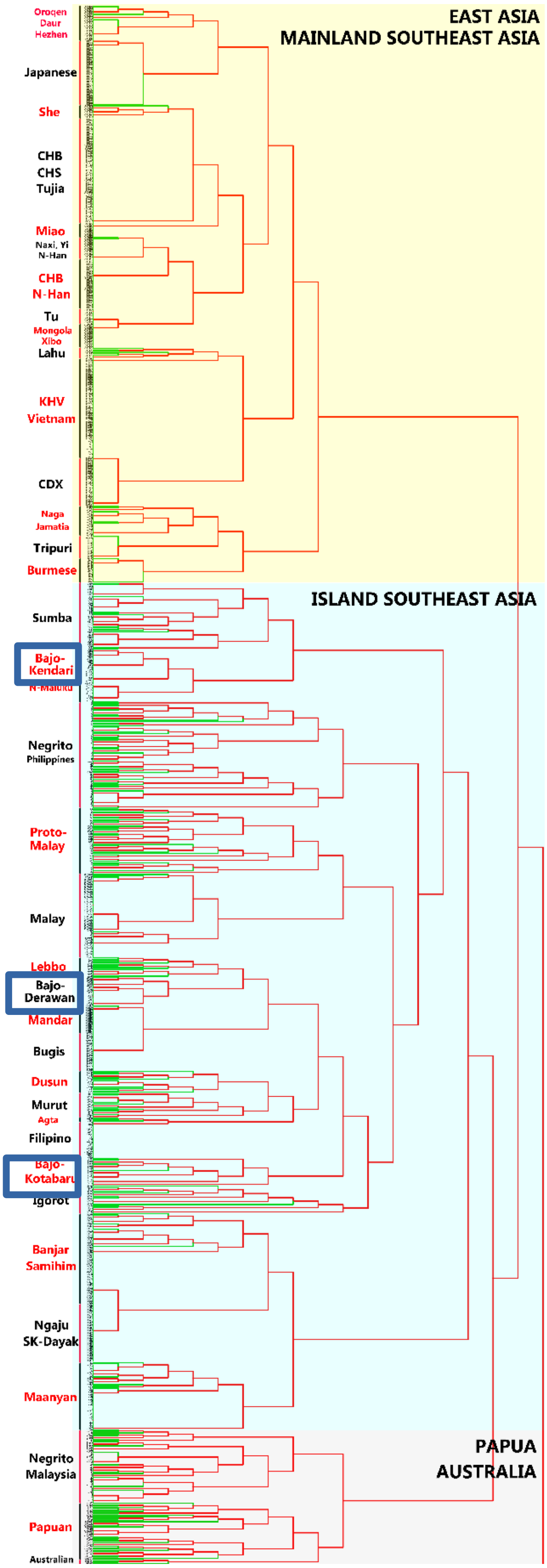


Figure S3. Admixture plot from K=11 - K=20 for all populations in the studied dataset. The lowest CV value is at K=17. The three Bajo groups are marked by asterisks.

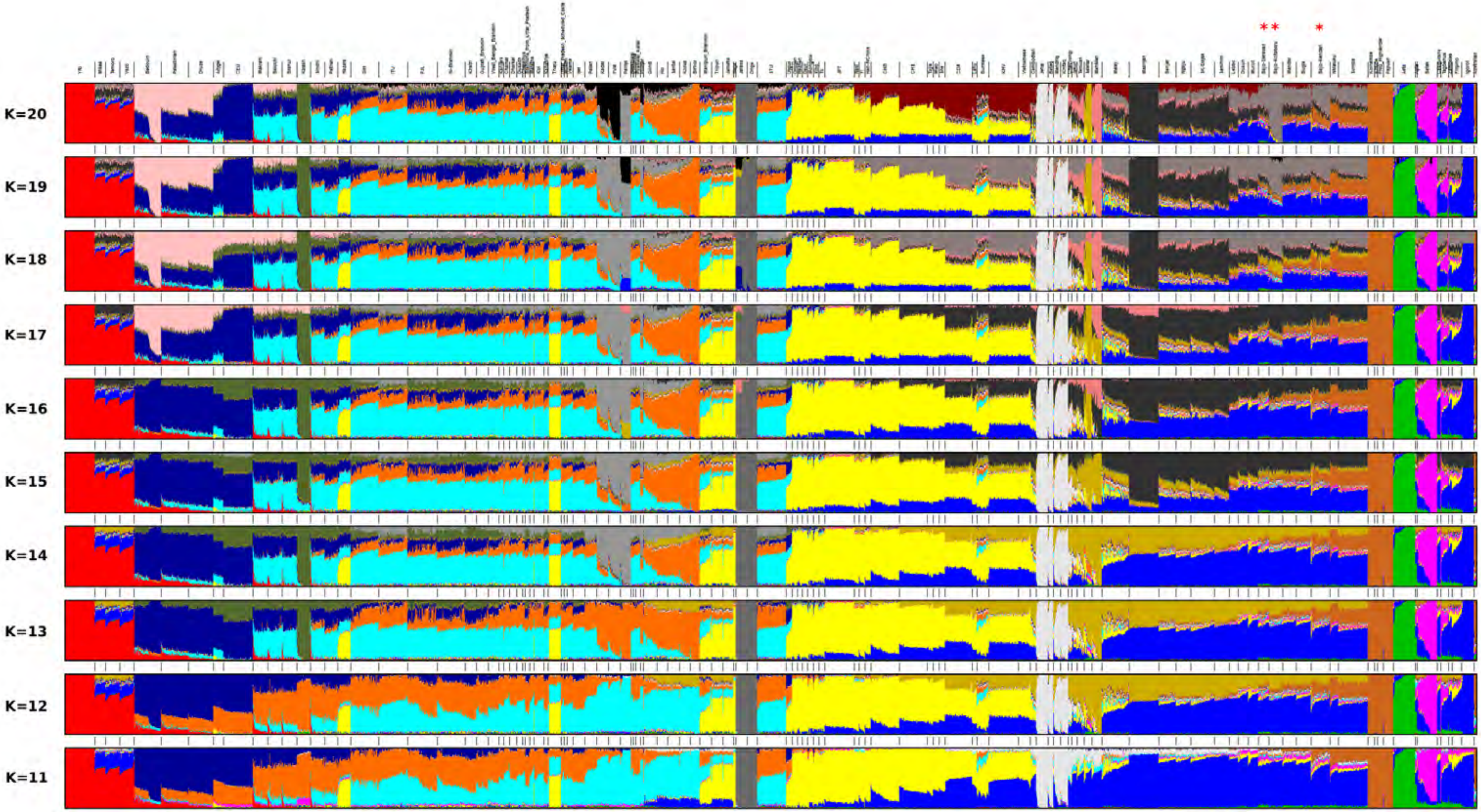


Figure S4. Cross validation analysis of ADMIXTURE runs from K=11 to K=25 with 10 iterations showing K=17 as the most representative plot.

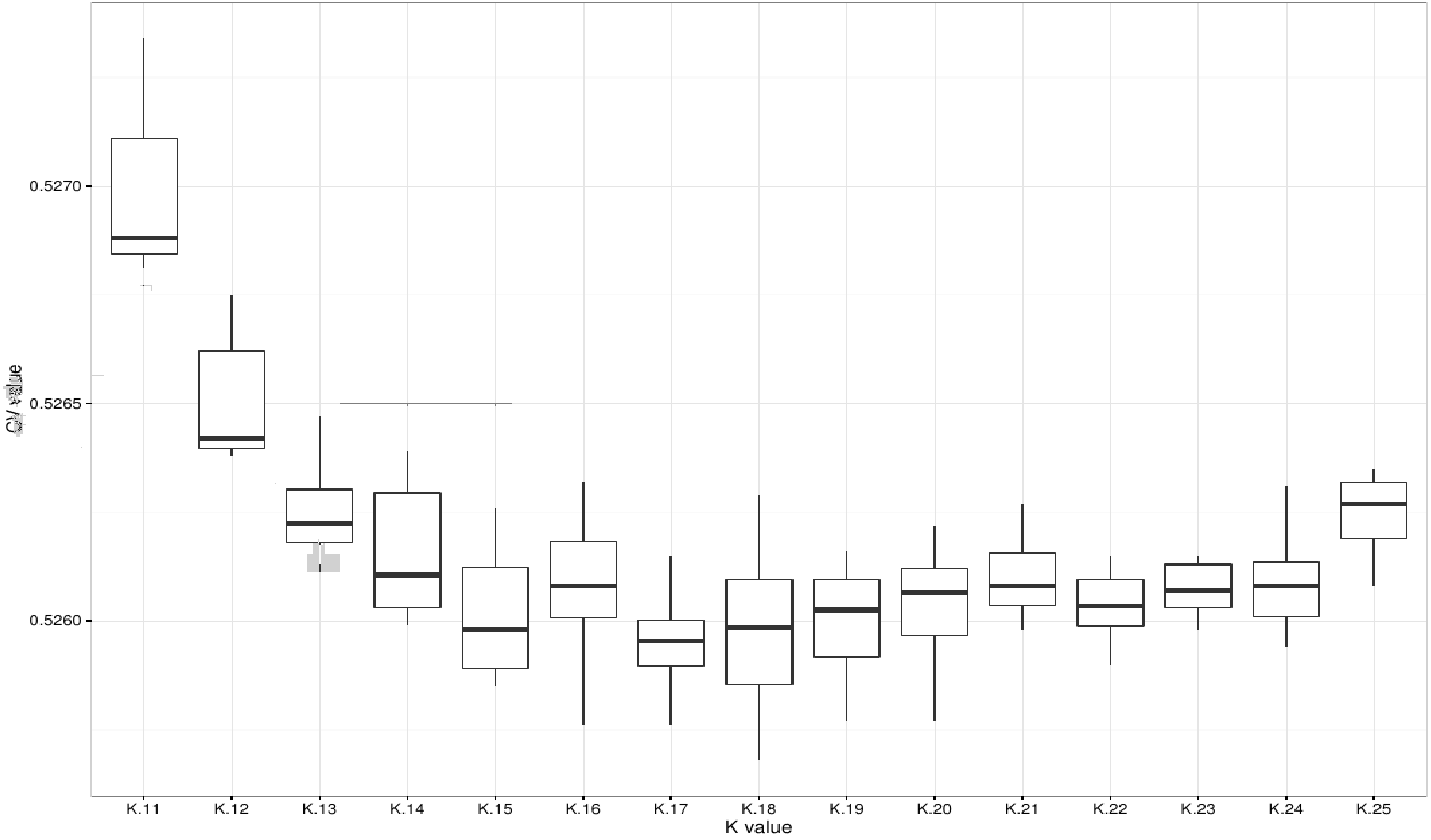


Figure S5. Comparisons of estimated autosomal and X ancestry on the basis of unsupervised clustering using ADMIXTURE. There is no significant difference between autosomal and X-chromosome of East Asian/Papuan ancestry in Kendari Bajo and Bugis (one-tailed Wilcoxon test P > 0.05), but there is significance difference, leading to potential Bajo ancestral female-bias admixture, in Derawan and Kotabaru Bajo (one tailed Wilcoxon test P < 0.01), traced by the Papuan component brought by Bajo ancestral women from Sulawesi to Borneo.

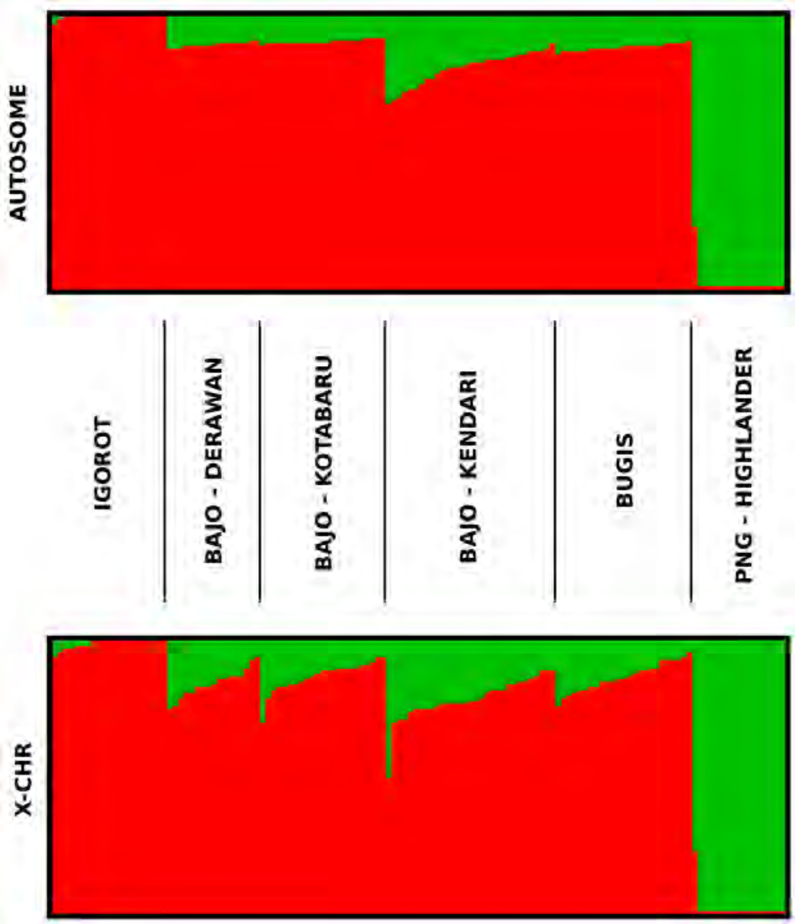


Figure S6. Runs of Homozygosity analyses on Island Southeast Asian populations compared to other representative populations from Africa, Europe, South Asia and East Asia.

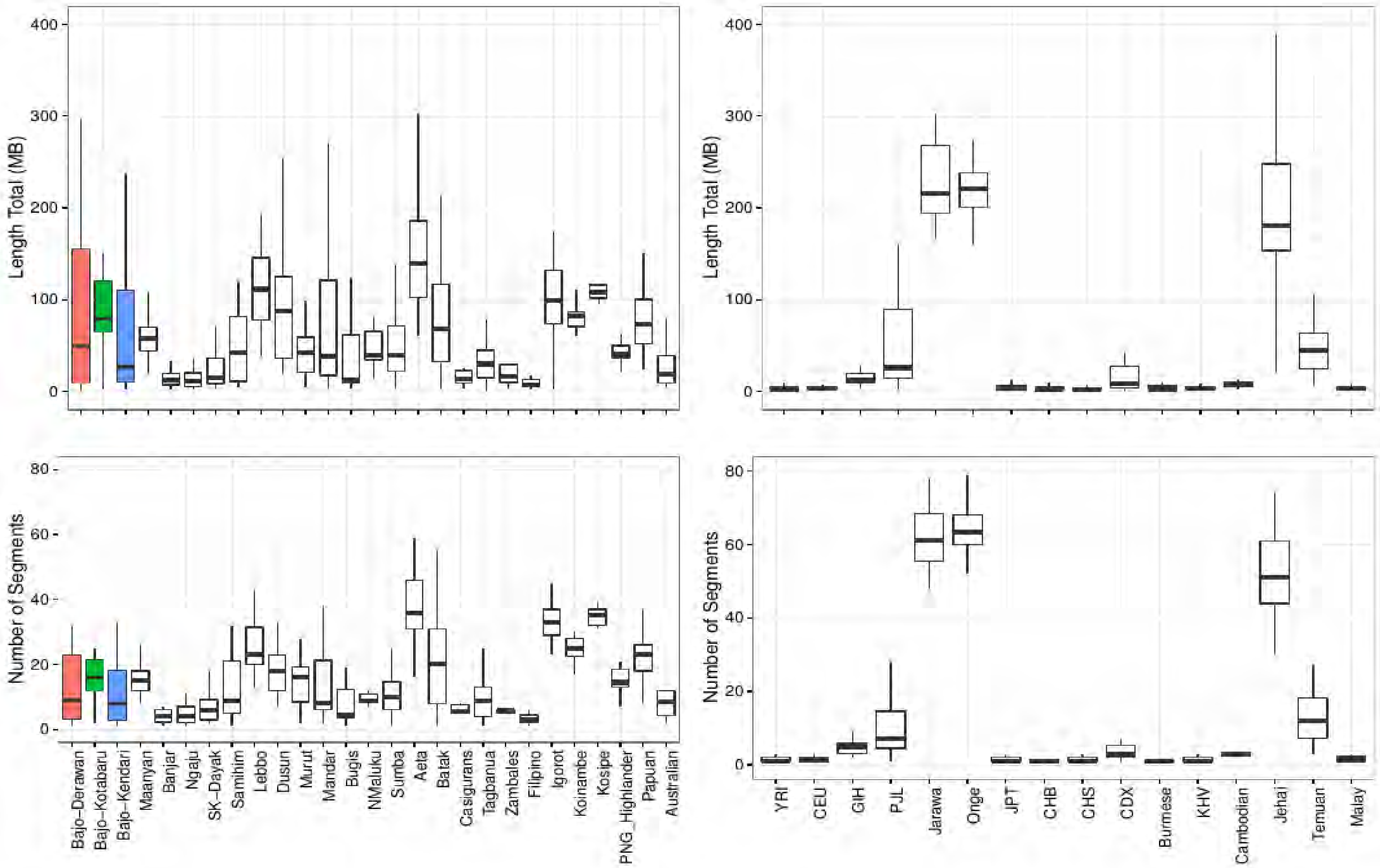
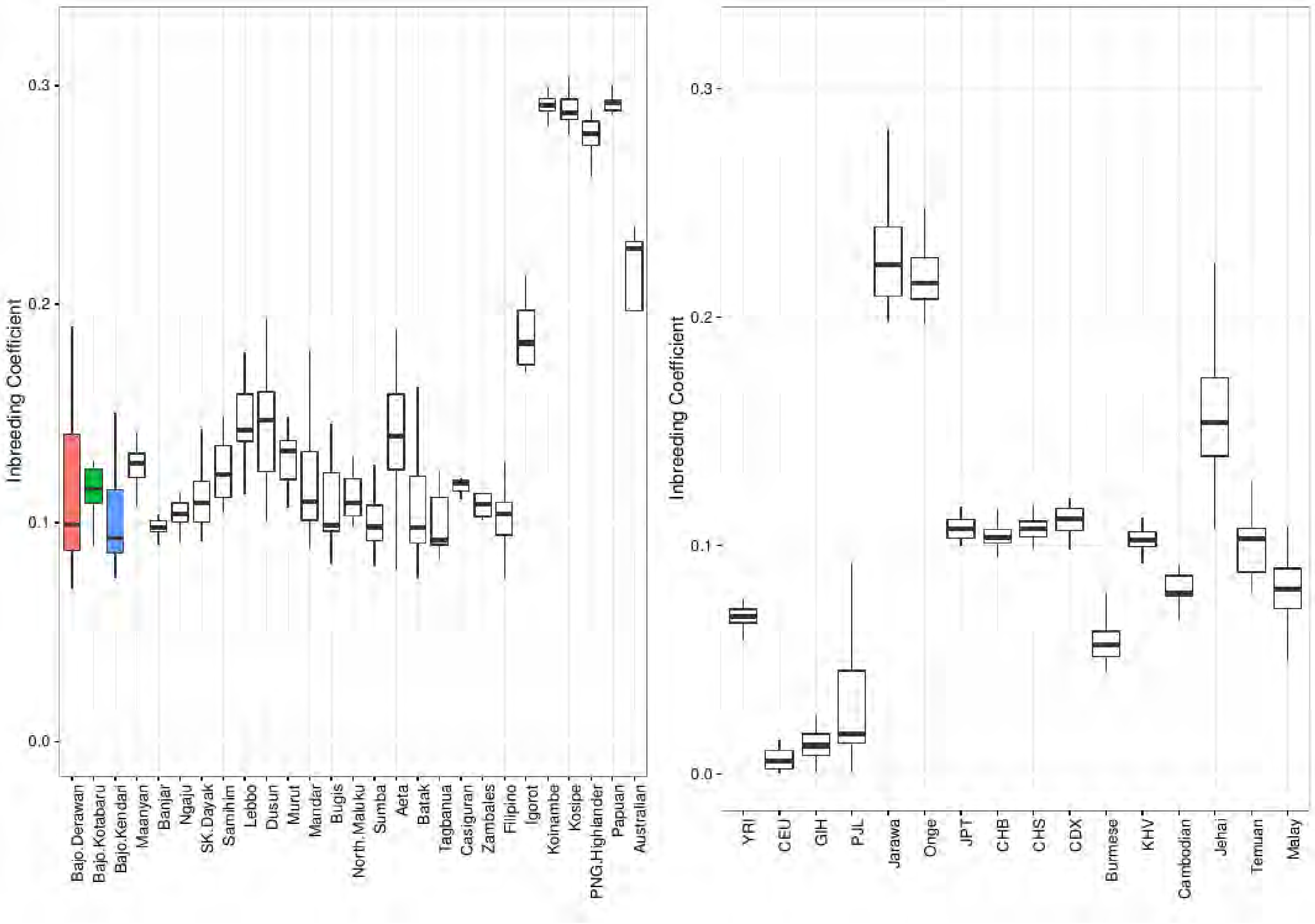


Figure S7. Inbreeding coefficient analyses on Island Southeast Asian populations compared to other representative populations from Africa, Europe, South Asia and East Asia.



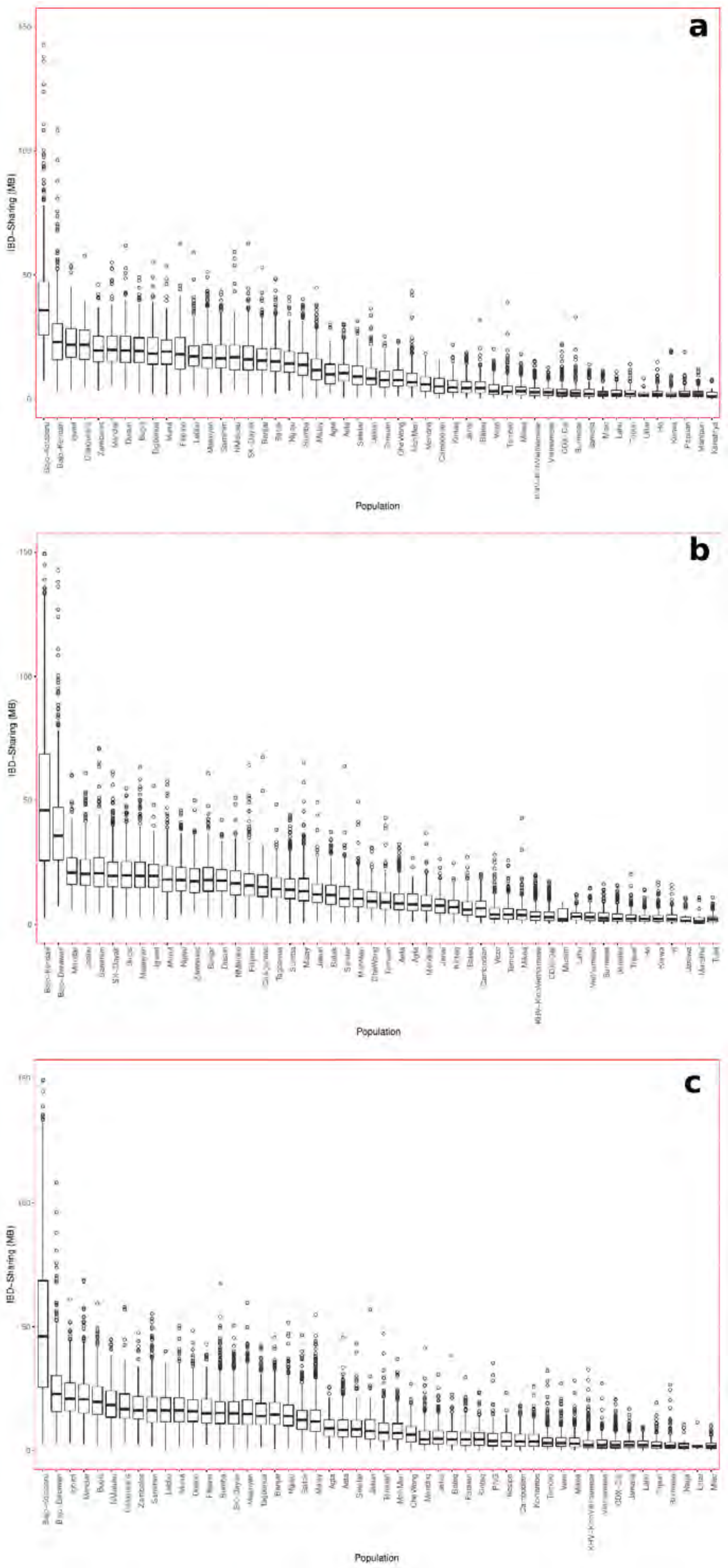


Figure S8. Shared IBD results from refinedIBD showing long chunk IBD sharing between Bajo communities with (a) Derawan Bajo (b) Kotabaru Bajo (c) Kendari Bajo as target populations.

Figure S9. TreeMix analysis on 13 migration nodes showing migrations from Papuan cluster to the three Bajo communities.

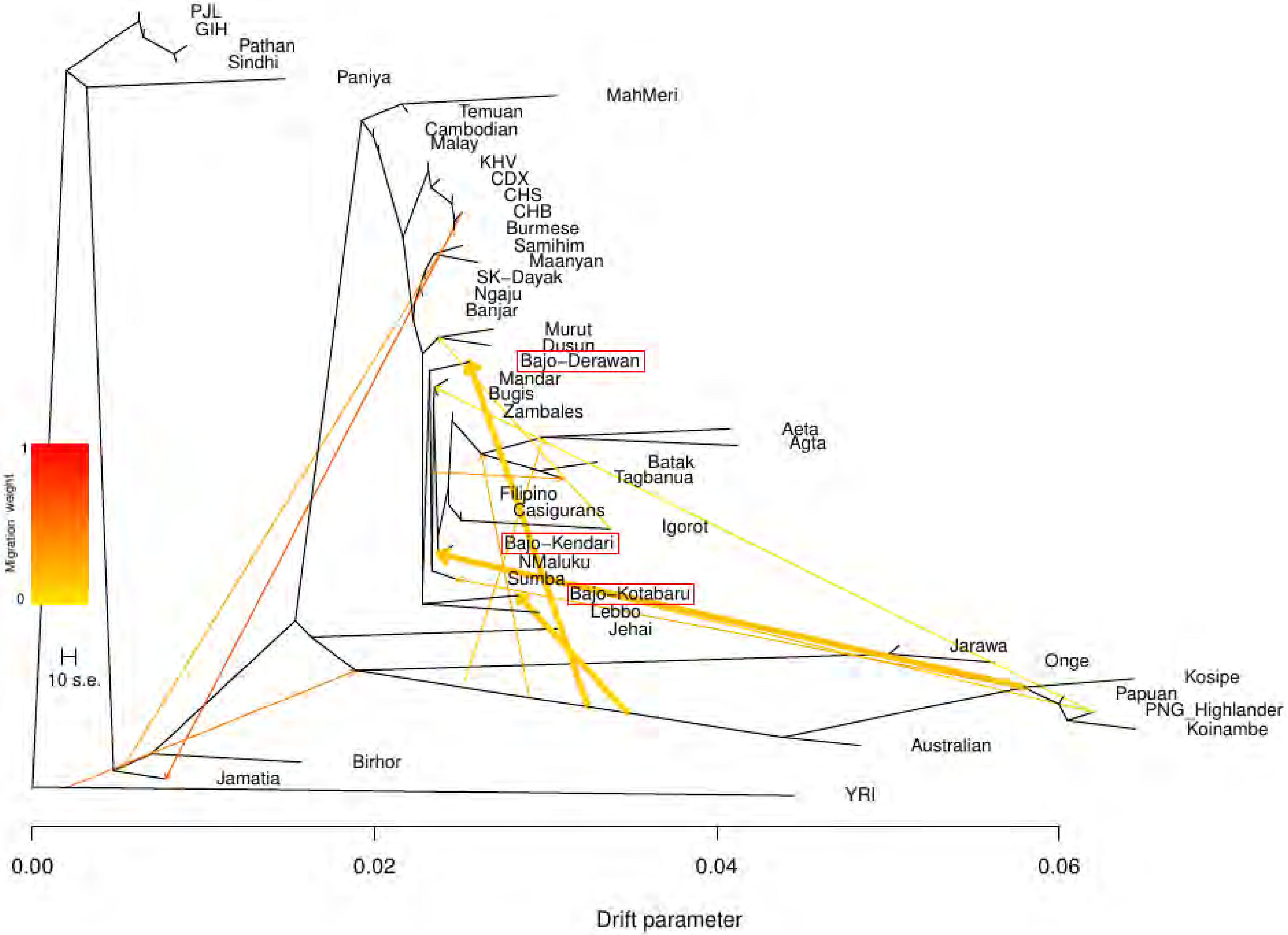


Table S1. Dataset of populations used in this study.

Population	Region	N	Reference
YRI	Africa	50	1
Mikea	Madagascar	18	2
Temoro	Madagascar	24	2
Vezo	Madagascar	24	2
Bedouin	Near/Middle East	46	3
Palestinian	Near/Middle East	46	3
Druze	Near/Middle East	42	3
Adygei	Near/Middle East	17	3
CEU	European	50	1
Makrani	Central/South Asia	25	3
Balochi	Central/South Asia	24	3
Brahui	Central/South Asia	25	3
Kalash	Central/South Asia	23	3
Sindhi	Central/South Asia	24	3
Pathan	Central/South Asia	24	3
Hazara	Central/South Asia	22	3
GIH	Central/South Asia	50	1
ITU	Central/South Asia	50	1
PJL	Central/South Asia	50	1
IN-Brahmin	Central/South Asia	47	4
Khatri	Central/South Asia	19	5
Gujrati_Brahmin	Central/South Asia	20	5
West_Bengal_Brahmin	Central/South Asia	18	5
Kanjars	Central/South Asia	8	6
Chamar	Central/South Asia	10	6
Dharkar	Central/South Asia	12	6
Dusadh	Central/South Asia	10	6
Chenchus	Central/South Asia	4	6
Brahmins_from_Uttar_Pradesh	Central/South Asia	8	6
Maratha	Central/South Asia	7	5
Kol	Central/South Asia	17	6
Kshatriya	Central/South Asia	7	7
Tharu	Central/South Asia	22	8
Uttar_Pradesh_Scheduled_Caste	Central/South Asia	5	6
Muslim	Central/South Asia	5	6
Velama	Central/South Asia	10	8
Iyer	Central/South Asia	20	5

Pallan	Central/South Asia	20	5
Kadar	Central/South Asia	20	5
Irula	Central/South Asia	20	8
Paniya	Central/South Asia	18	9
Hakkipikki	Central/South Asia	4	6
Kurumba	Central/South Asia	4	6
Piramalai_Kallar	Central/South Asia	8	6
Pulliyar	Central/South Asia	5	6
Gond	Central/South Asia	24	6
Ho	Central/South Asia	18	10
Santal	Central/South Asia	20	5
Korwa	Central/South Asia	18	5
Birhor	Central/South Asia	16	11
Manipuri_Brahmin	Central/South Asia	20	5
Tripuri	Central/South Asia	19	5
Jamatia	Central/South Asia	18	5
Naga	Central/South Asia	4	6
Jarawa	Central/South Asia	19	5
Onge	Central/South Asia	17	8
STU	Central/South Asia	50	1
Uygur	East Asia	10	3
Oroqen	East Asia	9	3
Hezhen	East Asia	9	3
Daur	East Asia	10	3
Mongola	East Asia	10	3
Xibo	East Asia	9	3
Tu	East Asia	10	3
JPT	East Asia	50	1
Naxi	East Asia	9	3
Yi	East Asia	10	3
Han-NChina	East Asia	10	3
CHB	East Asia	50	1
CHS	East Asia	50	1
Tujia	East Asia	10	3
Miao	East Asia	10	3
She	East Asia	10	3
CDX	East Asia	50	1
Lahu	East Asia	8	3
Burmese	Mainland Southeast Asia	20	4
KHV	Mainland Southeast Asia	50	1
Vietnamese	Mainland Southeast Asia	20	4

Jehai	Mainland Southeast Asia	20	12
Bateq	Mainland Southeast Asia	9	12
Mendriq	Mainland Southeast Asia	12	12
Kintaq	Mainland Southeast Asia	13	12
CheWong	Mainland Southeast Asia	6	12
Jakun	Mainland Southeast Asia	9	12
Temuan	Mainland Southeast Asia	12	12
Seletar	Mainland Southeast Asia	13	12
MahMeri	Mainland Southeast Asia	17	12
Malay	Mainland Southeast Asia	46	13
Maanyan	Island Southeast Asia	50	14
Banjar	Island Southeast Asia	30	15
Ngaju	Island Southeast Asia	25	15
SK-Dayak	Island Southeast Asia	40	14
Samihim	Island Southeast Asia	25	this study
Lebbo	Island Southeast Asia	15	2
Dusun	Island Southeast Asia	17	4
Murut	Island Southeast Asia	17	4
Bajo-Derawan	Island Southeast Asia	18	this study
Bajo-Kotabaru	Island Southeast Asia	23	this study
Mandar	Island Southeast Asia	23	this study
Bugis	Island Southeast Asia	25	this study
Bajo-Kendari	Island Southeast Asia	32	2
NMaluku	Island Southeast Asia	14	this study
Sumba	Island Southeast Asia	50	16
Koinambe	Papuan	11	17
Kosipe	Papuan	5	17
PNG_Highlander	Papuan	10	17
Papuan	Papuan	17	17
Aeta	Island Southeast Asia	37	17
Agta	Island Southeast Asia	3	17
Batak	Island Southeast Asia	34	17
Casigurans	Island Southeast Asia	6	17
Tagbanua	Island Southeast Asia	13	17
Zambales	Island Southeast Asia	6	17
Filipino	Island Southeast Asia	16	4
Igorot	Island Southeast Asia	21	4
Australian Aborigines	Australia	4	4

Supplementary Table S2. Pairwise FST genetic distance values between populations in East Asia including the three Bajo groups. Yellow-highlighted values are the lowest distance value between each Bajo group and other populations.

	Southern_Han	**Dai_Chinese**	**Burmese**	**Vietnamese**	**Temuan**	**Malay**	**Maanyan**	**Banjar**	**Ngaju**	**SK-Dayak**	**Samihim**	**Lebbo**	**Dusun**	**Murut**	**Bajo_Derawan**	**Bajo_Kotabaru**	**Mandar**	**Bugis**	**Bajo_Kendari**	**NMaluku**	**Sumba**	**Koinambe**	**Kosipe**	**PNG_Highland**	**Batak**	**Filipino**	**Igorot**
Southern_Han	0																										
Dai_Chinese	0.008	0																									
Burmese	0.008	0.01	0																								
Vietnamese	0.006	0.002	0.008	0																							
Temuan	0.029	0.02	0.019	0.018	0																						
Malay	0.014	0.008	0.008	0.006	0.012	0																					
Maanyan	0.029	0.02	0.025	0.019	0.026	0.013	0																				
Banjar	0.018	0.01	0.013	0.008	0.016	0.002	0.009	0																			
Ngaju	0.02	0.011	0.014	0.01	0.016	0.004	0.01	0.003	0																		
SK-Dayak	0.019	0.01	0.015	0.009	0.016	0.004	0.009	0.002	0.002	0																	
Samihim	0.026	0.018	0.022	0.016	0.024	0.011	0.014	0.007	0.009	0.006	0																
Lebbo	0.041	0.033	0.036	0.032	0.039	0.025	0.035	0.026	0.027	0.026	0.033	0															
Dusun	0.028	0.021	0.026	0.02	0.031	0.016	0.025	0.015	0.017	0.015	0.022	0.036	0														
Murut	0.027	0.02	0.026	0.019	0.031	0.015	0.025	0.015	0.016	0.015	0.021	0.036	0.021	0													
Bajo_Derawan	0.024	0.018	0.02	0.017	0.027	0.011	0.022	0.012	0.014	0.013	0.019	0.033	0.022	0.022	0												
Bajo_Kotabaru	0.035	0.028	0.029	0.027	0.034	0.02	0.031	0.021	0.023	0.022	0.029	0.043	0.033	0.033	0.025	0											
Mandar	0.02	0.015	0.017	0.013	0.024	0.007	0.018	0.007	0.009	0.008	0.015	0.029	0.017	0.017	0.012	0.023	0										
Bugis	0.018	0.012	0.015	0.011	0.021	0.005	0.016	0.005	0.007	0.006	0.013	0.026	0.016	0.015	0.01	0.021	0.003	0									
Bajo_Kendari	0.023	0.018	0.017	0.016	0.024	0.009	0.022	0.011	0.013	0.012	0.019	0.031	0.022	0.022	0.014	0.022	0.009	0.007	0								
NMaluku	0.034	0.031	0.028	0.029	0.037	0.021	0.034	0.022	0.024	0.024	0.031	0.042	0.033	0.034	0.022	0.035	0.015	0.014	0.012	0							
Sumba	0.028	0.023	0.021	0.022	0.029	0.014	0.028	0.016	0.017	0.017	0.025	0.037	0.028	0.029	0.019	0.03	0.012	0.011	0.01	0.008	0						
Koinambe	0.182	0.183	0.161	0.179	0.177	0.164	0.187	0.171	0.174	0.176	0.184	0.191	0.191	0.193	0.168	0.179	0.161	0.16	0.14	0.107	0.121	0					
Kosipe	0.182	0.182	0.161	0.179	0.176	0.164	0.187	0.171	0.173	0.175	0.184	0.191	0.19	0.193	0.167	0.179	0.161	0.16	0.141	0.11	0.123	0.05	0				
PNG_Highland	0.174	0.175	0.153	0.171	0.168	0.156	0.18	0.163	0.166	0.168	0.176	0.183	0.184	0.185	0.159	0.171	0.153	0.152	0.132	0.1	0.113	0.022	0.038	0			
Batak	0.046	0.042	0.04	0.041	0.049	0.034	0.047	0.036	0.038	0.037	0.044	0.057	0.046	0.046	0.038	0.05	0.035	0.034	0.035	0.04	0.038	0.173	0.173	0.165	0		
Filipino	0.014	0.011	0.015	0.009	0.024	0.007	0.017	0.007	0.009	0.008	0.013	0.028	0.014	0.014	0.011	0.023	0.006	0.004	0.01	0.019	0.016	0.175	0.174	0.167	0.032	0	
Igorot	0.043	0.039	0.047	0.038	0.056	0.036	0.044	0.035	0.037	0.035	0.041	0.055	0.04	0.038	0.038	0.051	0.032	0.03	0.039	0.048	0.046	0.215	0.214	0.207	0.062	0.025	0

Note. All pairwise FST values have P<1x10^{-4} after 10,000 bootstrap runs.

Table S3. The f3-statistics returning the top-50 significant admixture results (Z < -2) involving Bajo-Derawan and Bajo-Kendari as target populations. Bajo-Kotabaru is not shown as admixed population because of recent genetic drift.

Source1	Source2	Target	f3	StdError	Z-Score
Igorot	PNG_Highlander	Bajo-Derawan	-0.008757	0.000704	-12.441
Igorot	Koinambe	Bajo-Derawan	-0.00855	0.000696	-12.287
Igorot	Papuan	Bajo-Derawan	-0.008303	0.000692	-11.993
Igorot	Kosipe	Bajo-Derawan	-0.007729	0.000716	-10.789
Australian	Igorot	Bajo-Derawan	-0.005525	0.000683	-8.096
Murut	PNG_Highlander	Bajo-Derawan	-0.004745	0.000616	-7.702
Kosipe	Murut	Bajo-Derawan	-0.004595	0.000642	-7.153
Murut	Papuan	Bajo-Derawan	-0.004224	0.000594	-7.111
Koinambe	Murut	Bajo-Derawan	-0.004196	0.000611	-6.872
Dusun	PNG_Highlander	Bajo-Derawan	-0.003748	0.000622	-6.026
Dusun	Koinambe	Bajo-Derawan	-0.003341	0.000608	-5.491
Igorot	Iyer	Bajo-Derawan	-0.002456	0.000463	-5.309
Dusun	Papuan	Bajo-Derawan	-0.003101	0.000586	-5.29
Igorot	Maratha	Bajo-Derawan	-0.002691	0.000513	-5.247
Dusun	Kosipe	Bajo-Derawan	-0.003283	0.000628	-5.232
Igorot	PJL	Bajo-Derawan	-0.002215	0.000432	-5.128
Balochi	Igorot	Bajo-Derawan	-0.002382	0.000474	-5.022
Igorot	Velama	Bajo-Derawan	-0.002323	0.000469	-4.953
Chenchus	Igorot	Bajo-Derawan	-0.002568	0.000527	-4.876
Igorot	Makrani	Bajo-Derawan	-0.002415	0.000495	-4.876
Igorot	Muslim	Bajo-Derawan	-0.00237	0.000501	-4.729
Igorot	Pulliyar	Bajo-Derawan	-0.002771	0.000587	-4.717
Igorot	Pallan	Bajo-Derawan	-0.002064	0.000439	-4.705
Brahui	Igorot	Bajo-Derawan	-0.002308	0.000492	-4.694
GIH	Igorot	Bajo-Derawan	-0.002064	0.000446	-4.629
Igorot	Kadar	Bajo-Derawan	-0.002135	0.000466	-4.579
Igorot	STU	Bajo-Derawan	-0.001965	0.000436	-4.509
Bedouin	Igorot	Bajo-Derawan	-0.002313	0.000517	-4.477
Igorot	ITU	Bajo-Derawan	-0.001997	0.000447	-4.464
Igorot	Pathan	Bajo-Derawan	-0.002081	0.000466	-4.464
Igorot	Sindhi	Bajo-Derawan	-0.002089	0.000468	-4.464
Druze	Igorot	Bajo-Derawan	-0.002369	0.000534	-4.44
Igorot	Palestinian	Bajo-Derawan	-0.002236	0.000505	-4.429
Gujrati_Brahmin	Igorot	Bajo-Derawan	-0.002051	0.00047	-4.362
Igorot	West_Bengal_Brahmin	Bajo-Derawan	-0.00195	0.000448	-4.351

CEU	Igorot	Bajo-Derawan	-0.002229	0.000513	-4.348
Igorot	Piramalai_Kallar	Bajo-Derawan	-0.002088	0.000494	-4.229
Chamar	Igorot	Bajo-Derawan	-0.001963	0.000468	-4.19
Igorot	IN-Brahmin	Bajo-Derawan	-0.001846	0.000442	-4.181
Brahmins_from_Uttar_Pradesh	Igorot	Bajo-Derawan	-0.00204	0.00049	-4.166
Igorot	Khatri	Bajo-Derawan	-0.001922	0.000475	-4.044
Igorot	Kanjars	Bajo-Derawan	-0.001899	0.000478	-3.976
Gond	Igorot	Bajo-Derawan	-0.00168	0.000425	-3.956
Igorot	Kol	Bajo-Derawan	-0.001784	0.000454	-3.928
Igorot	Kalash	Bajo-Derawan	-0.001974	0.000509	-3.881
Igorot	Irula	Bajo-Derawan	-0.001771	0.000473	-3.741
Igorot	Kurumba	Bajo-Derawan	-0.002123	0.000579	-3.669
Adygei	Igorot	Bajo-Derawan	-0.00193	0.000527	-3.665
Igorot	Uttar_Pradesh_Scheduled_Caste	Bajo-Derawan	-0.001875	0.000519	-3.615
Igorot	Kshatriya	Bajo-Derawan	-0.001791	0.000496	-3.607
Igorot	Papuan	Bajo-Kendari	-0.023809	0.00059	-40.329
Igorot	Koinambe	Bajo-Kendari	-0.024159	0.000619	-39.039
Igorot	PNG_Highlander	Bajo-Kendari	-0.023977	0.000616	-38.937
Murut	Papuan	Bajo-Kendari	-0.020158	0.000532	-37.868
CDX	Papuan	Bajo-Kendari	-0.016897	0.00045	-37.523
Papuan	SK-Dayak	Bajo-Kendari	-0.015492	0.000415	-37.295
Filipino	Papuan	Bajo-Kendari	-0.016045	0.000431	-37.209
Maanyan	Papuan	Bajo-Kendari	-0.016681	0.000458	-36.427
Murut	PNG_Highlander	Bajo-Kendari	-0.020392	0.000561	-36.329
Dusun	Papuan	Bajo-Kendari	-0.018956	0.000522	-36.295
Igorot	Kosipe	Bajo-Kendari	-0.022329	0.000619	-36.058
CDX	Koinambe	Bajo-Kendari	-0.016939	0.000473	-35.787
PNG_Highlander	SK-Dayak	Bajo-Kendari	-0.015621	0.000437	-35.754
Koinambe	Murut	Bajo-Kendari	-0.020238	0.000569	-35.556
Koinambe	SK-Dayak	Bajo-Kendari	-0.01581	0.000445	-35.534
KHV	Papuan	Bajo-Kendari	-0.015879	0.00045	-35.315
Filipino	PNG_Highlander	Bajo-Kendari	-0.015803	0.000448	-35.283
Filipino	Koinambe	Bajo-Kendari	-0.016254	0.000461	-35.262
Maanyan	PNG_Highlander	Bajo-Kendari	-0.016968	0.000483	-35.158
CDX	PNG_Highlander	Bajo-Kendari	-0.01673	0.000479	-34.953
Dusun	Koinambe	Bajo-Kendari	-0.019299	0.000553	-34.927
Kosipe	Murut	Bajo-Kendari	-0.01961	0.000566	-34.631
Dusun	PNG_Highlander	Bajo-Kendari	-0.019313	0.00056	-34.504
KHV	Koinambe	Bajo-Kendari	-0.016078	0.000469	-34.26
Koinambe	Maanyan	Bajo-Kendari	-0.016776	0.00049	-34.257
Papuan	Samihim	Bajo-Kendari	-0.016718	0.000493	-33.896
Dusun	Kosipe	Bajo-Kendari	-0.018222	0.000541	-33.658

Kosipe	Maanyan	Bajo-Kendari	-0.015935	0.000475	-33.536
KHV	PNG_Highlander	Bajo-Kendari	-0.015686	0.00047	-33.408
Kosipe	SK-Dayak	Bajo-Kendari	-0.014739	0.000444	-33.192
CDX	Kosipe	Bajo-Kendari	-0.015984	0.000484	-33.003
CHS	Papuan	Bajo-Kendari	-0.016026	0.000491	-32.635
CHS	Koinambe	Bajo-Kendari	-0.01619	0.000498	-32.525
Filipino	Kosipe	Bajo-Kendari	-0.015136	0.000466	-32.468
Banjar	Papuan	Bajo-Kendari	-0.013645	0.000425	-32.092
PNG_Highlander	Samihim	Bajo-Kendari	-0.016453	0.000514	-32.017
Koinambe	Samihim	Bajo-Kendari	-0.01686	0.000531	-31.745
Banjar	Koinambe	Bajo-Kendari	-0.013876	0.000437	-31.733
Casigurans	Papuan	Bajo-Kendari	-0.018045	0.000579	-31.184
Ngaju	Papuan	Bajo-Kendari	-0.014021	0.00045	-31.163
Banjar	PNG_Highlander	Bajo-Kendari	-0.013669	0.000443	-30.884
KHV	Kosipe	Bajo-Kendari	-0.014921	0.000487	-30.639
CHS	PNG_Highlander	Bajo-Kendari	-0.015601	0.000512	-30.497
Javanese	Papuan	Bajo-Kendari	-0.01393	0.000458	-30.412
Kosipe	Samihim	Bajo-Kendari	-0.015795	0.000523	-30.181
Koinambe	Ngaju	Bajo-Kendari	-0.014207	0.000474	-29.984
Ngaju	PNG_Highlander	Bajo-Kendari	-0.01402	0.000474	-29.567
Casigurans	PNG_Highlander	Bajo-Kendari	-0.018117	0.000616	-29.401
Javanese	Koinambe	Bajo-Kendari	-0.014125	0.000481	-29.356
Casigurans	Koinambe	Bajo-Kendari	-0.017913	0.000618	-28.977

Table S4. Globetrotter results showing the admixture scenarios of studied populations.

Target Pop	Null inds.	Scenario	Date 1 (gens. ago	Population 1.1	%	Population 1.2	%	Population 1.3	%	Population 1.4	%	Date 2 (gens. Ago	Population 2.1	%	Population 2.2	%
B-DRW*	null0	One-date-multiway	24.02 (20-27)	CentralSouthAsia_Chamar	5.00%	ISEA_Filipino	45.00%	ISEA_Casigurans	24.00%	MSEA_Malay	26.00%	-	-	-	-	-
	null1	One-date	30.77 (26-35)	CentralSouthAsia_Chamar	5.00%	ISEA_Filipino	45.00%	ISEA_Casigurans	24.00%	MSEA_Malay	26.00%	-	-	-	-	-
B-KTBR*	null0	Uncertain	23.38 (19-27)	CentralSouthAsia_Kshatriya	5.00%	ISEA_Bugis	65.00%	MSEA_Malay	30.00%	-	-	-	-	-	-	-
	null1	One-date-multiway	33.29 (29-37)	CentralSouthAsia_Maratha	5.00%	ISEA_Bugis	69.00%	ISEA_Banjar	26.00%	-	-	-	-	-	-	-
B-KDR*	null0	Multiple-dates	62.84 (58-66)	PNG_Highlander	10.00%	ISEA_Bugis	90.00%	-	-	-	-	6.77 (3-10)	ISEA_Bugis	100.00%	-	-
	null1	Multiple-dates	65.93 (61-70)	PNG_Highlander	10.00%	ISEA_Bugis	90.00%	-	-	-	-	5.94 (3-9)	ISEA_Bugis	100.00%	-	-
B-DRW	null0	uncertain	26.65 (22-30)	CentralSouthAsia_Iyer	6.50%	ISEA_Bugis	43.50%	ISEA_Bajo-Kendari	11.50%	ISEA_Filipino	38.50%	-	-	-	-	-
	null1	uncertain	32.47 (28-38)	CentralSouthAsia_Iyer	6.50%	ISEA_Bugis	60.00%	ISEA_Filipino	33.50%	-	-	-	-	-	-	-
B-KTBR	null0	Multiple-dates	38.6 (32-43)	MSEA_Malay	27.00%	ISEA_Kendari	73.00%	-	-	-	-	6.2 (5-7)	MSEA_Malay	14.00%	ISEA_Kendari	86.00%
	null1	uncertain	13.8 (9-17)	ISEA_Banjar	8.00%	ISEA_Kendari	92.00%	-	-	-	-	-	-	-	-	-
B-KDR	null0	uncertain	43 (39-48)	PNG-Highlander	5.50%	ISEA_Bugis	68.50%	ISEA_Bajo-Kotabaru	26.00%	-	-	-	-	-	-	-
	null1	uncertain	47 (41-52)	PNG-Highlander	6.50%	ISEA_Bugis	85.50%	MSEA_Malay	8.00%	-	-	-	-	-	-	-
Bugis	null0	One-date-multiway	57.58 (54-61)	PNG-Highlander	5.00%	MSEA_Malay	45.00%	MSEA_Temuan	7.50%	ISEA_Filipino	42.50%	-	-	-	-	-
	null1	One-date-multiway	62.63 (59-65)	PNG-Highlander	14.00%	MSEA_Malay	45.00%	ISEA_Casigurans	41.00%	-	-	-	-	-	-	-
Mandar	null0	One-date	52.08 (48-56)	PNG-Highlander	6.00%	ISEA_Bugis	94.00%	-	-	-	-	-	-	-	-	-
	null1	One-date	53.6 (50-56)	PNG-Highlander	5.00%	ISEA_Bugis	95.00%	-	-	-	-	-	-	-	-	-
N-Maluku	null0	One-date	46.35	PNG-Highlander	28.00%	ISEA_Filipino	72.00%	-	-	-	-	-	-	-	-	-

			(44-47)													
	null1	One-date	51.2 (48-55)	PNG-Highlander	27.00%	ISEA_Casigurans	73.00%	-	-	-	-	-	-	-	-	-
Samihim	null0	One-date	2.4 (1-4)	ISEA_SK-Dayak	100%	-	-	-	-	-	-	-	-	-	-	-
	null1	One-date	10.84 (7-14)	ISEA_SK-Dayak	100%	-	-	-	-	-	-	-	-	-	-	-

*Note: Admixture scenario of target populations with asterisk (*) are estimated by excluding other Bajo communities as possible surrogate populations.

Table S5. MALDER analysis for each Bajo community using the parental populations inferred from GLOBETROTTER. For each estimated admixture date (t0 and t1), the amplitude of the weighted LD curve (Amp) and the time in generation (Gen) are given with their respective standard errors and Z-scores.

Target	**Source 1**	**Source 2**	**Amp$_{t0}$ (+/- SE)**	**Z$_{ampt0}$**	**Gen$_0$ (+/- SE)**	**Z$_{gent0}$**	**Ampt$_1$ (+/- SE)**	**Z$_{ampt1}$**	**Gen$_{t1}$ (+/- SE)**	**Z$_{gent1}$**
Bajo Derawan	Chamar	Malay	$6.35e^{-06}$ +/- $1.19e^{-06}$	5.34	26.61 +/- 5.76	4.62				
Bajo Derawan	Chamar	Filipino	$6.25e^{-05}$ +/- $1.17e^{-05}$	5.31	26.61 +/- 5.76	4.62				
Bajo Derawan	Filipino	Malay	$4.30e^{-05}$ +/- $8.09e^{-06}$	5.31	26.61 +/- 5.76	4.62				
Bajo Derawan	Bugis	Filipino	$4.85e^{-05}$ +/- $1.33e^{-05}$	5.43	26.61 +/- 5.76	4.62				
Bajo Derawan	Bugis	Iyer	$5.74e^{-05}$ +/- $1.15e^{-05}$	4.98	26.61 +/- 5.76	4.62				
Bajo Derawan	Filipino	Iyer	$7.16e^{-05}$ +/- $1.34e^{-05}$	5.33	26.61 +/- 5.76	4.62				
Bajo Kotabaru	Maratha	Banjar	$7.04e^{-05}$ +/- $9.02e^{-06}$	7.82	40.48 +/- 4.56	8.86				
Bajo Kotabaru	Maratha	Bugis	$6.52e^{-05}$ +/- $8.91e^{-06}$	7.31	40.48 +/- 4.56	8.86				
Bajo Kotabaru	Bugis	Banjar	*failed**	*failed**	*failed**	*failed**				
Bajo Kotabaru	Bajo-Kendari	Malay	*failed**	*failed**	*failed**	*failed**				
Bajo Kotabaru	Bajo-Kendari	Banjar	*failed**	*failed**	*failed**	*failed**				
Bajo Kendari	Bugis	PNG-Highlander	$6.51e^{-05}$ +/- $1.29e^{-05}$	5.029	78.64 +/- 11.41	6.895	$5.57e^{-05}$ +/- $1.47e^{-05}$	3.789	6.12 +/- 1.96	3.115

*Note: failed results are generated by population combinations (between Bajo Kotabaru with Banjar and Bajo-Kendari) that have long-range LD correlation, so that MALDER cannot estimate the admixture date

References

1 The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 2015; 526: 68–74.

2 Pierron D, Razafindrazaka H, Pagani L, Ricaut F-X, Antao T, Capredon M et al. Genome-wide evidence of Austronesian–Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. Proc Natl Acad Sci 2014; 111: 936–941.

3 Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S et al. Worldwide human relationships inferred from genome-wide patterns of variation. Science 2008; 319: 1100–1104.

4 Mörseburg A, Pagani L, Ricaut F-X, Yngvadottir B, Harney E, Castillo C et al. Multi-layered population structure in Island Southeast Asians. Eur J Hum Genet 2016; 24: 1605–1611.

5 Basu A, Sarkar-Roy N, Majumder PP. Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. Proc Natl Acad Sci 2016; 113: 1594–1599.

6 Metspalu M, Romero IG, Yunusbayev B, Chaubey G, Mallick CB, Hudjashov G et al. Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. Am J Hum Genet 2011; 89: 731–744.

7 Yunusbayev B, Metspalu M, Metspalu E, Valeev A, Litvinov S, Valiev R et al. The Genetic Legacy of the expansion of Turkic-speaking nomads across Eurasia. PLOS Genet 2015; 11: e1005068.

8 Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. Nature 2009; 461: 489–494.

9 Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rosset S, Parik J et al. The genome-wide structure of the Jewish people. Nature 2010; 466: 238–242.

10 Chaubey G, Metspalu M, Choi Y, Mägi R, Romero IG, Soares P et al. Population genetic structure in Indian Austroasiatic speakers: The role of landscape barriers and sex-specific admixture. Mol Biol Evol 2011; 28: 1013–1024.

11 Moorjani P, Thangaraj K, Patterson N, Lipson M, Loh P-R, Govindaraj P et al. Genetic evidence for recent population mixture in India. Am J Hum Genet 2013; 93: 422–438.

12 Aghakhanian F, Yunus Y, Naidu R, Jinam T, Manica A, Hoh BP et al. Unravelling the genetic history of Negritos and Indigenous populations of Southeast Asia. Genome Biol Evol 2015; 7: 1206–1215.

13 Teo Y-Y, Sim X, Ong RTH, Tan AKS, Chen J, Tantoso E et al. Singapore Genome Variation Project: A haplotype map of three Southeast Asian populations. Genome Res 2009; 19: 2154–2162.

14 Kusuma P, Brucato N, Cox MP, Pierron D, Razafindrazaka H, Adelaar A et al. Contrasting linguistic and genetic origins of the Asian source populations of Malagasy. Sci Rep 2016; 6: 26066.

15 Brucato N, Kusuma P, Cox MP, Pierron D, Purnomo GA, Adelaar A et al. Malagasy genetic ancestry comes from an historical Malay trading post in Southeast Borneo. Mol Biol Evol 2016; 33: 2396–2400.

16 Cox MP, Hudjashov G, Sim A, Savina O, Karafet TM, Sudoyo H et al. Small traditional human communities sustain genomic diversity over microgeographic scales despite linguistic isolation. Mol Biol Evol 2016; 33: 2273–2284.

17 Scholes C, Siddle K, Ducourneau A, Crivellaro F, Järve M, Rootsi S et al. Genetic diversity and evidence for population admixture in Batak Negritos from Palawan. Am J Phys Anthropol 2011; 146: 62–72.

Annex G. Supplementary information from article entitled "Malagasy genetic ancestry comes from an historical Malay trading post in Southeast Borneo"

(Published in Molecular Biology and Evolution, 33 (9): 2396-2400, July 5th 2016. doi:10.1093/molbev/msw117)

Nicolas Brucato[1,*], Pradiptajati Kusuma[1,2,*], Murray P. Cox[3,] Denis Pierron[1], Gludhug A. Purnomo[2], Alexander Adelaar[4], Toomas Kivisild[6,7], Thierry Letellier[1], Herawati Sudoyo[2,5] and François-Xavier Ricaut[1,†].

[1] Evolutionary Medicine Group, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse UMR 5288 CNRS, Université Toulouse III, Université de Toulouse. France, Toulouse, France
[2] Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular Biology, Jakarta, Indonesia
[3] Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
[4] Asia Institute, University of Melbourne, Melbourne, Australia
[5] Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia
[6] Department of Biological Anthropology, University of Cambridge, Cambridge, United Kingdom
[7] Estonian Biocentre, Tartu, Estonia

Figure S1. Locations of the sampled groups in Borneo island in Indonesia. The map was generated using Global Mapper v.15 software. (http://www.bluemarblegeo.com/products/global-mapper.php)

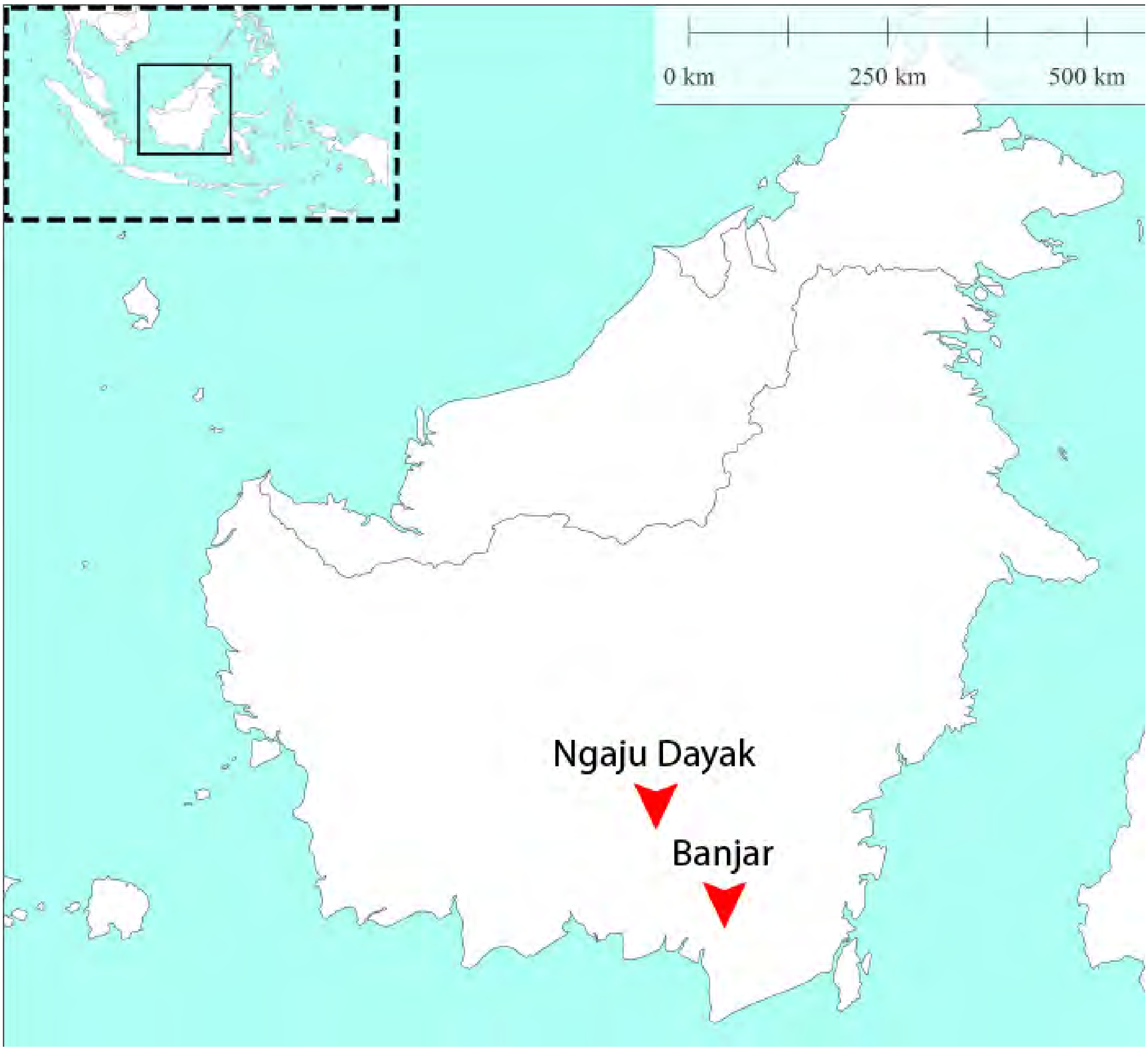


Figure S2. ADMIXTURE plot from K=2 to K=16 of the low SNP density dataset using ADMIXTURE software ##(Alexander, et al. 2009)##. The Banjar and Ngaju populations are identified by the red rectangle, respectively labelled 41 and 42. The studied populations are numerically labelled as follows: 1: Gambian; 2: Mende; 3: Beninese; 4: Yoruba; 5: Afar; 6: Amhara; 7: Tygray; 8: Oromo; 9: Wolayta; 10: Ari-Blacksmith; 11: Ari-Cultivator; 12: Somalian; 13: Sudanese; 14: Anuak; 15: Gumuz; 16: Luhya; 17: South African Bantu; 18: Vezo; 19: Mikea; 20: Temoro; 21: US. European descendant; 22: Onge; 23: Brahmin; 24: Dravidian; 25: Indian from Singapore; 26: Han; 27: Burmese; 28: Vietnamese; 29: Jehai; 30: Bateq; 31: Mendriq; 32: Kintaq; 33: MahMeri; 34: CheWong; 35: Jakun; 36: Temuan; 37: Seletar; 38: Malay; 39: Besemah; 40: Semende; 41: Bidayuh; 42: Banjar; 43: Ngaju; 44: South Kalimantan Dayak; 45: Ma'anyan; 46: Lebbo; 47: Murut; 48: Dusun; 49: Bajo; 50: Central Sumbanese; 51: East Sumbanese; 52: West Sumbanese; 53: Nusa Tenggara; 54: Moluccas; 55: Papua New Guinea Highlander; 56: Philippines Negrito; 57: Filipino; 58: Igorot; 59: Taiwan Aborigenes; 60: Fiji; 61: Polynesia.

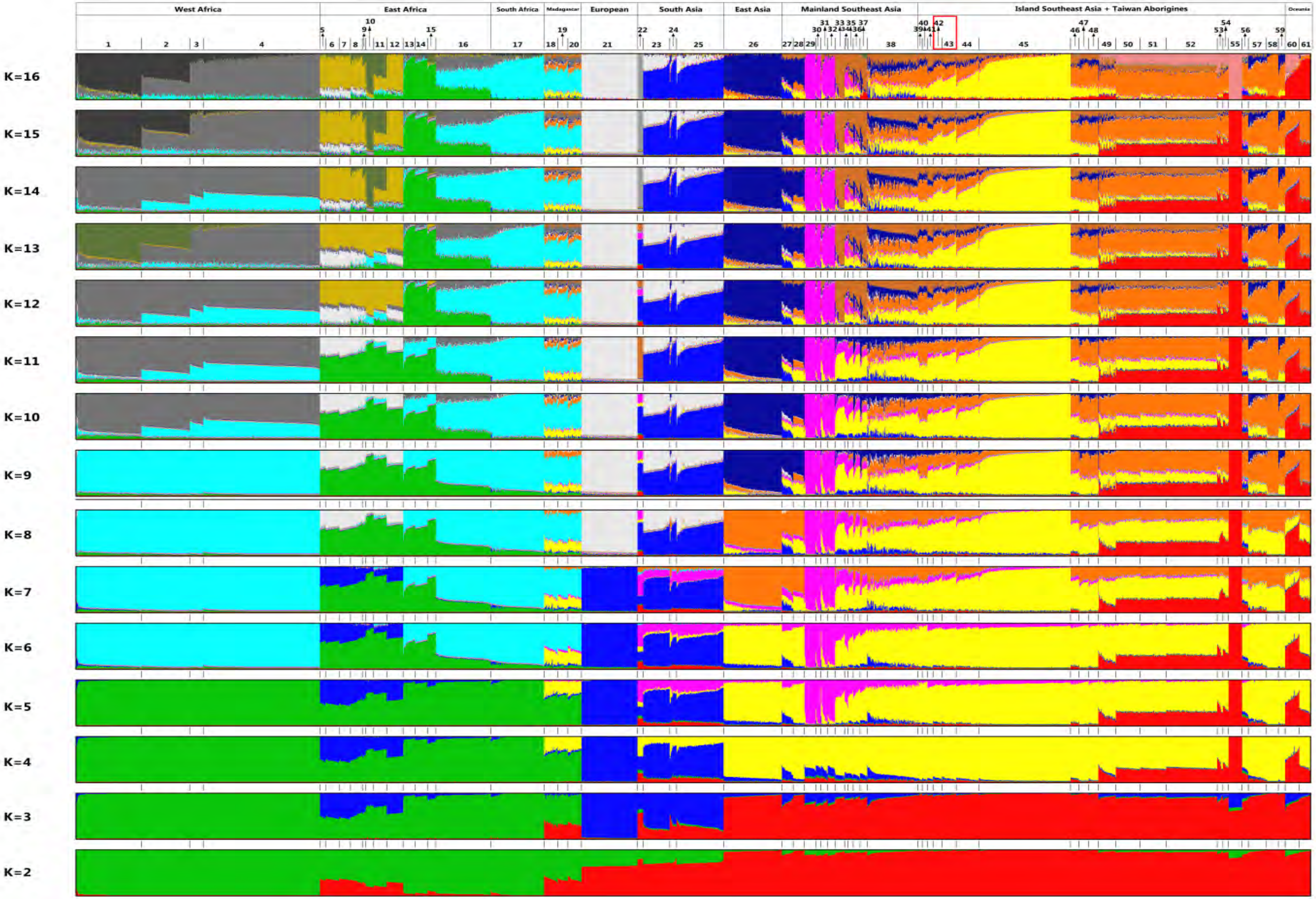

West Africa
East Africa
South Africa
Madagascar
European
South Asia
East Asia
Mainland Southeast Asia
Island Southeast Asia + Taiwan Aborigines
Oceania
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
K=16
K=15
K=14
K=13
K=12
K=11
K=10
K=9
K=8
K=7
K=6
K=5
K=4
K=3
K=2

Figure S3. Cross-validation plot generated from the ADMIXTURE analysis from K=2 to K=18 on the low density SNP dataset from 10 iterations. It shows that K=16 has the lowest median CV value thereby representing the strongly supported ADMIXTURE plot on the analysis.

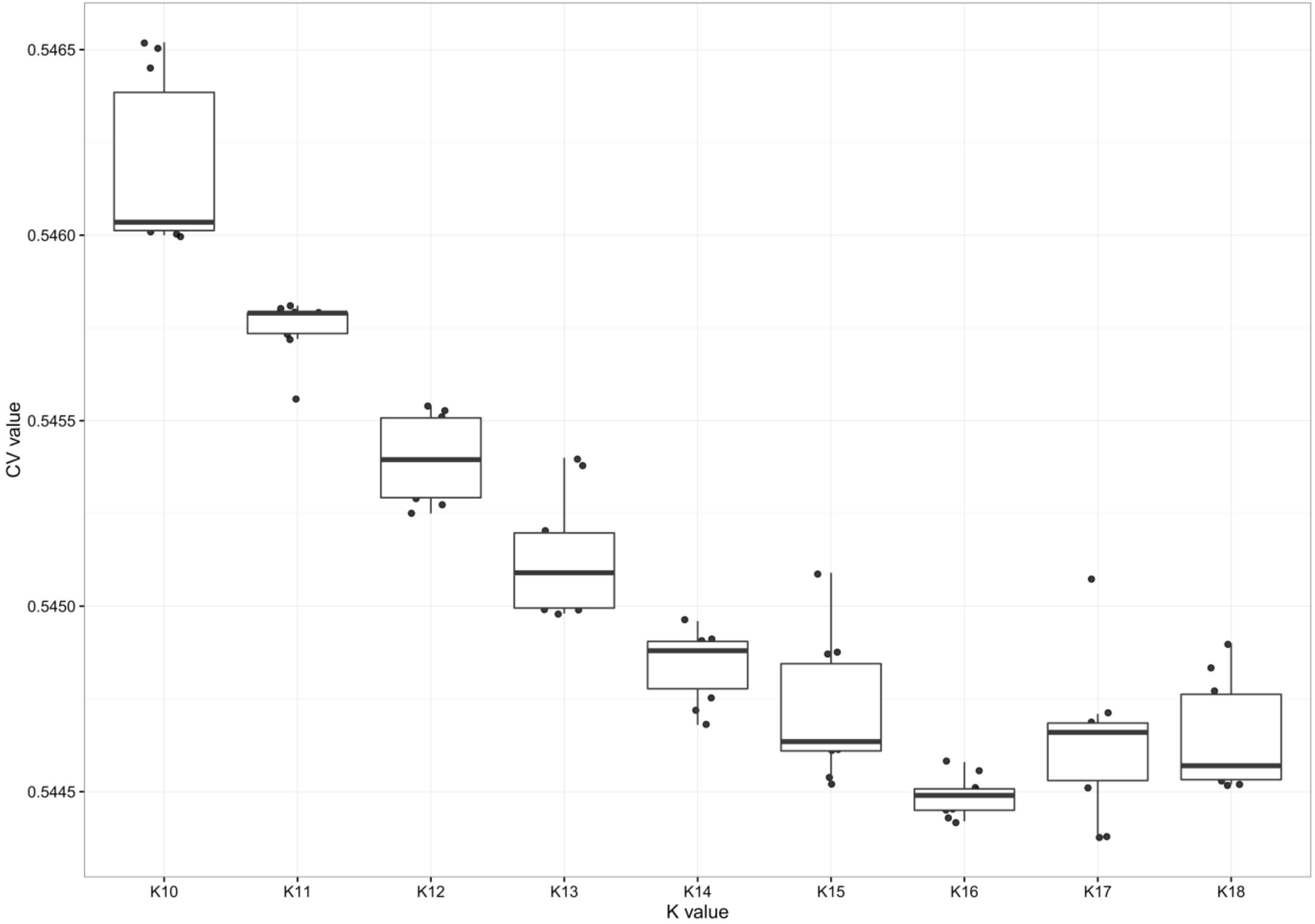


Figure S4. TreeMix analysis on the low density dataset with five migration nodes (99% variability). The arrows indicate the direction of the gene flow and their colors its intensity. A strong gene flow is identified from a node clustering Southeast Asian populations into the three Malagasy populations.

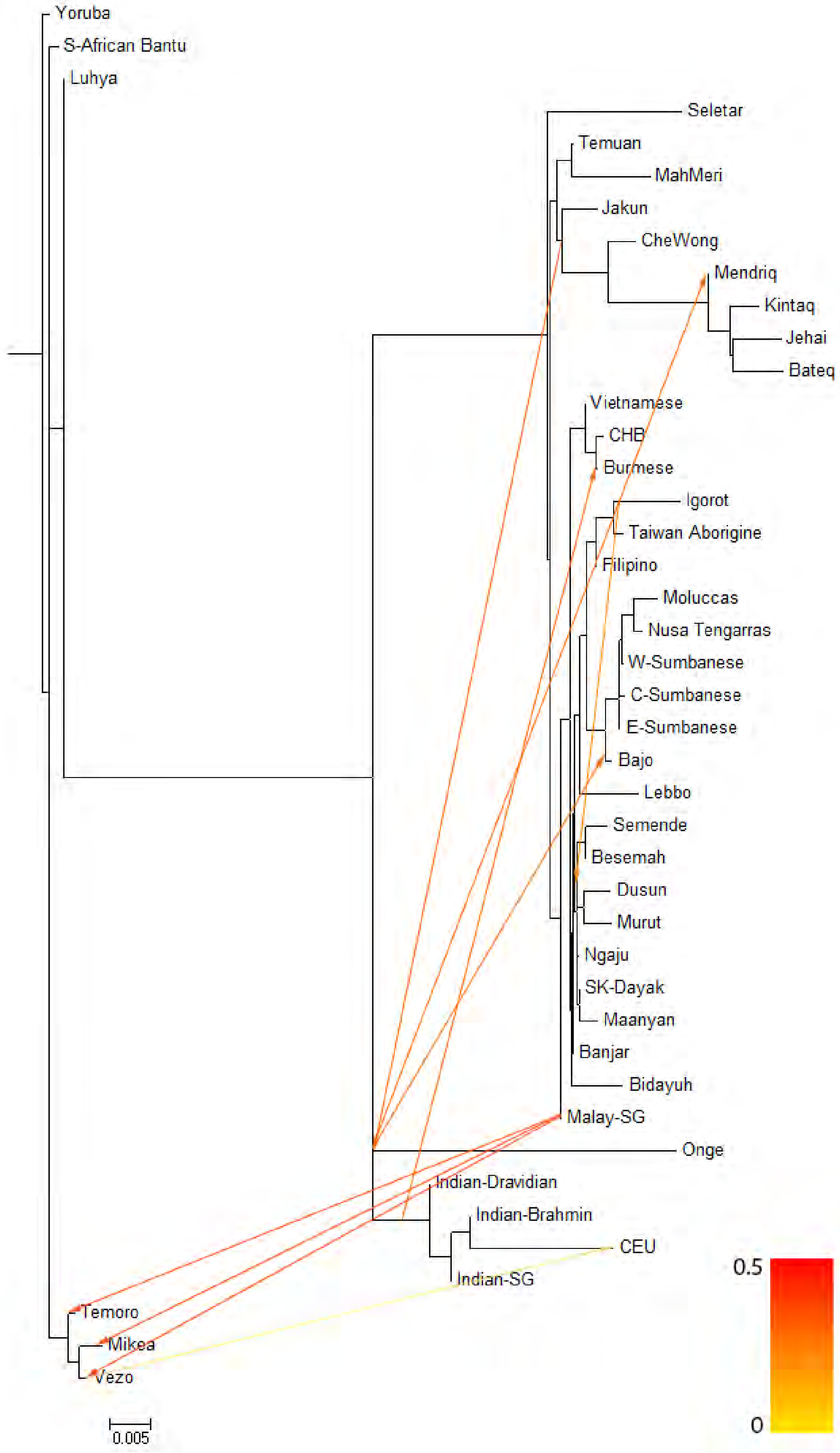


Figure S5. Ancestry-specific Principal Component Analysis based on masked SNPs from individuals of the high density dataset obtained after PCAdmix analysis and plot with EIGENSOFT v6.0.1.

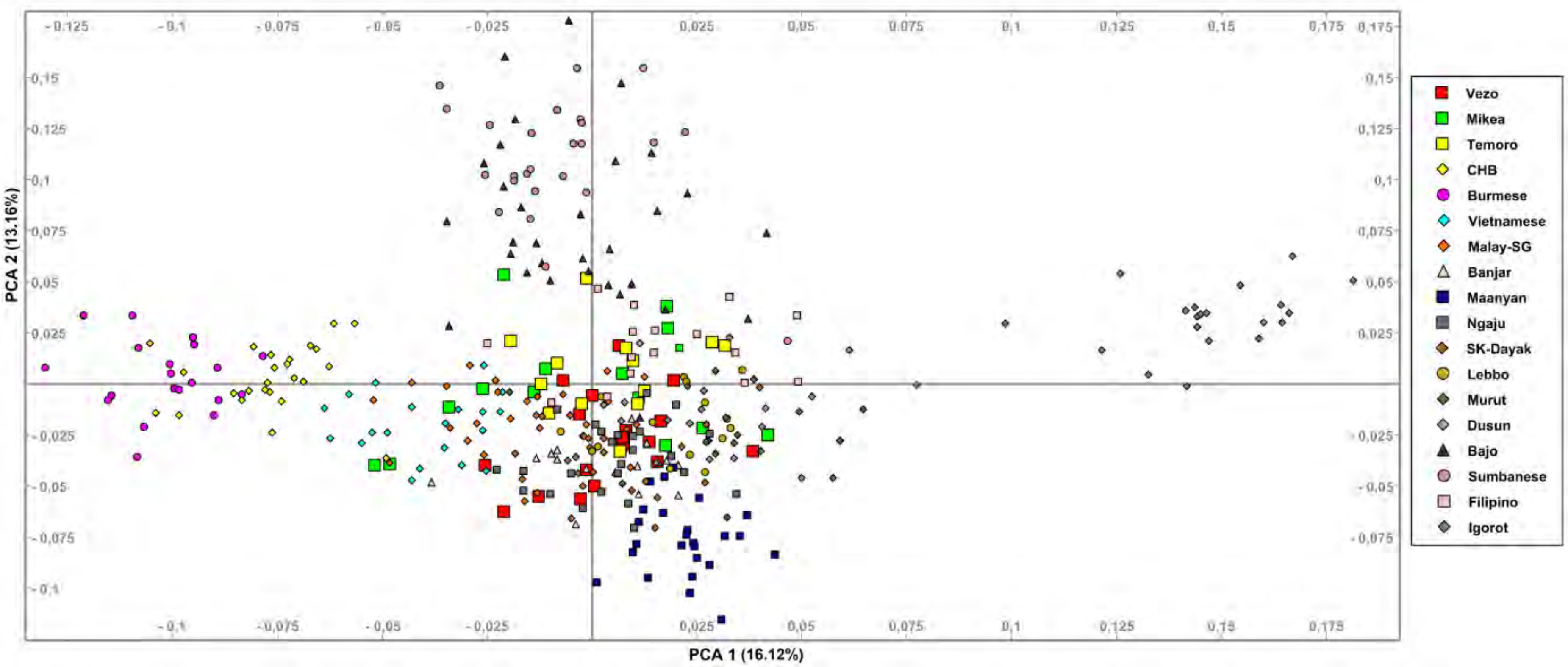


Figure S6. Shared Identity-By-Descent fragments between pairs of individuals in the high density dataset, filtering for different thresholds of shared IBD fragments (from IBD > 10cM to IBD >90cM), calculated with RefinedIBD. Populations are represented by a blue circle which diameter is proportional to the number of individuals (from n=4 to n=31). Shared IBD fragments between each pair of individuals are represented by a black line with Cytoscape.

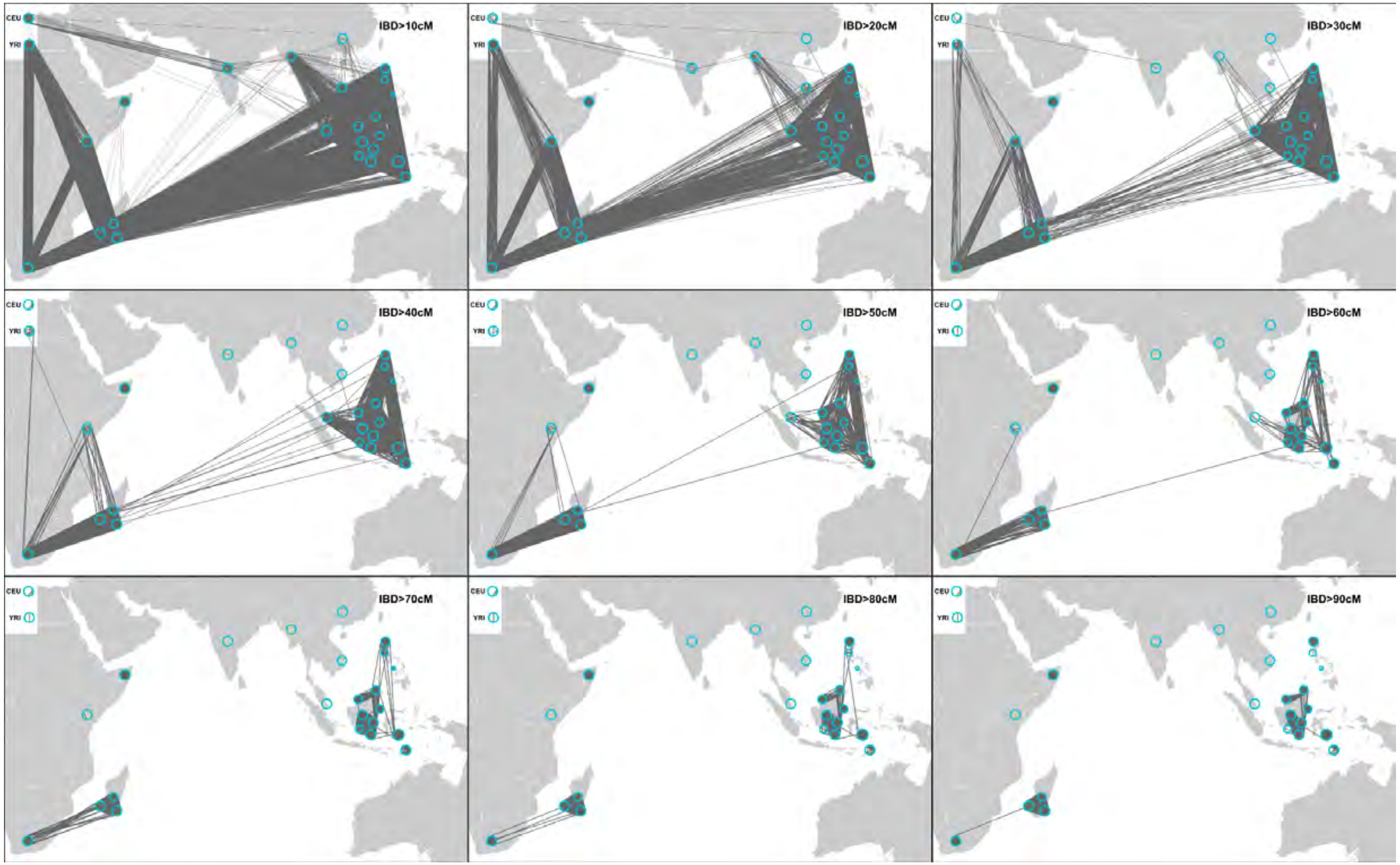
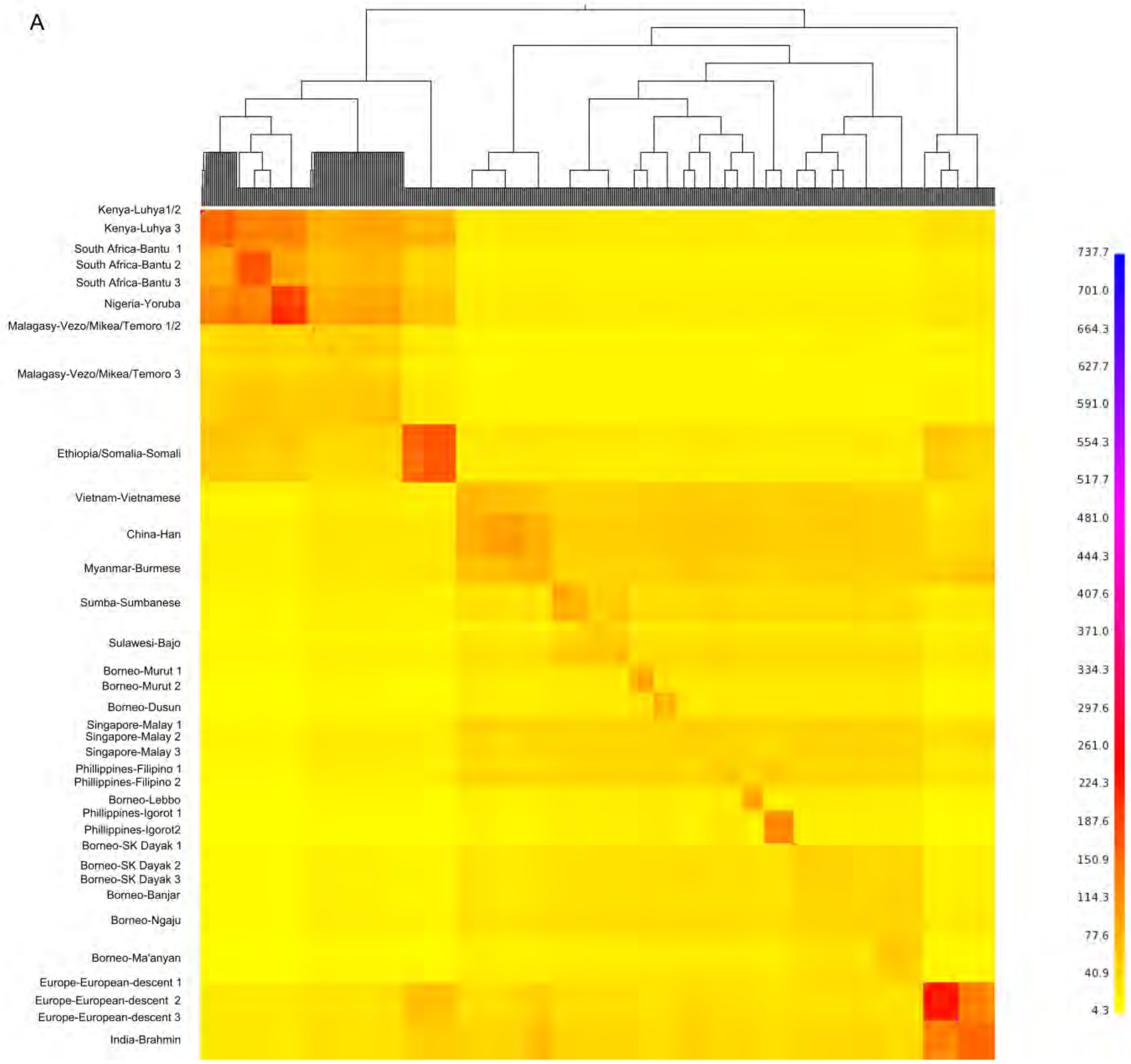


Figure S7. Coancestry heat map (A) and pairwise coincidence matrix (B) of the individuals of the high SNP density dataset generated by fineSTRUCTURE ##(Lawson, et al. 2012)##. The color scale in (A) indicates the chunk counts between each pair of individuals calculated by Chromopainter. The color scale in (B) indicates the pairwise coincidence between each pair of individuals. Nodes are collapsed below 60% probability. For convenience, close labels are merged (e.g., Kenya-Luhya 1/2 refers to population sub-structure in two groups of the Kenya-Luhya).



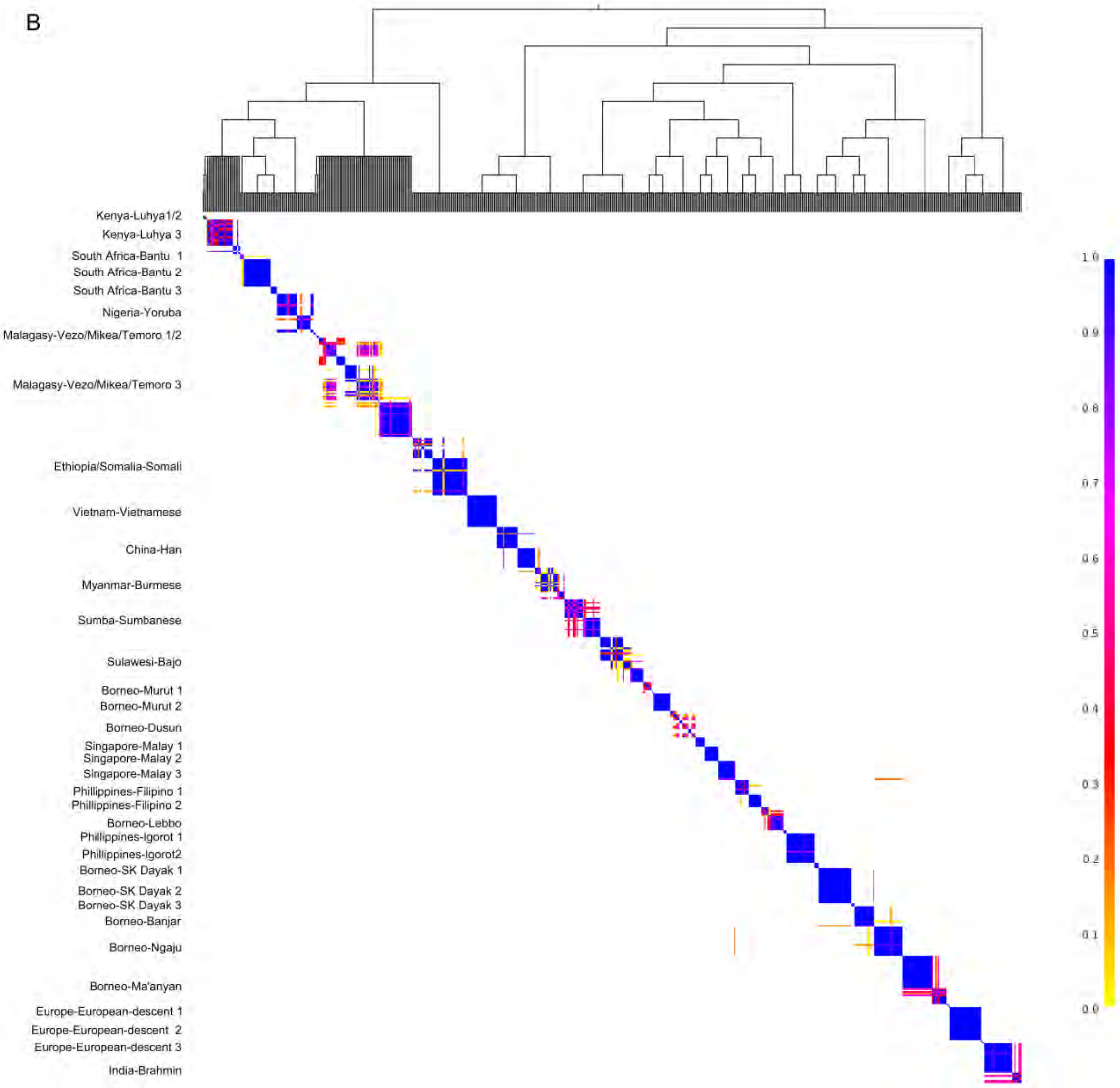

B
Kenya-Luhya1/2
Kenya-Luhya 3
South Africa-Bantu 1
South Africa-Bantu 2
South Africa-Bantu 3
Nigeria-Yoruba
Malagasy-Vezo/Mikea/Temoro 1/2
Malagasy-Vezo/Mikea/Temoro 3
Ethiopia/Somalia-Somali
Vietnam-Vietnamese
China-Han
Myanmar-Burmese
Sumba-Sumbanese
Sulawesi-Bajo
Borneo-Murut 1
Borneo-Murut 2
Borneo-Dusun
Singapore-Malay 1
Singapore-Malay 2
Singapore-Malay 3
Phillippines-Filipino 1
Phillippines-Filipino 2
Borneo-Lebbo
Phillippines-Igorot 1
Phillippines-Igorot2
Borneo-SK Dayak 1
Borneo-SK Dayak 2
Borneo-SK Dayak 3
Borneo-Banjar
Borneo-Ngaju
Borneo-Ma'anyan
Europe-European-descent 1
Europe-European-descent 2
Europe-European-descent 3
India-Brahmin
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

Table S1. Table of the populations included in both the low and high density SNP datasets used in the study.

Population	Country	N (LDD)	N (HDD)	Reference
Yoruba	Nigeria	206	25	(International HapMap, et al. 2010)
Mende	Senegal	85	-	(International HapMap et al. 2010)
Gambian	Gambia	115	-	(International HapMap et al. 2010)
European	USA	99	25	(International HapMap et al. 2010)
Han	China	103	25	(International HapMap et al. 2010)
Luhya	Kenya	97	25	(International HapMap et al. 2010)
Beninese	Benin	24	-	Fortes-Lima, personal communication
Afar	Ethiopia	10	-	(Pagani, et al. 2012)
Tygray	Ethiopia	20	-	(Pagani et al. 2012)
Amhara	Ethiopia	24	-	(Pagani et al. 2012)
Wolayta	Ethiopia	6	-	(Pagani et al. 2012)
Oromo	Ethiopia	21	-	(Pagani et al. 2012)
Ari-Cutivator	Ethiopia	21	-	(Pagani et al. 2012)
Ari-Blacksmith	Ethiopia	14	-	(Pagani et al. 2012)
Gumuz	Ethiopia	15	-	(Pagani et al. 2012)
Anuak	Ethiopia	22	-	(Pagani et al. 2012)
Sudanese	Sudan	21	-	(Pagani et al. 2012)
Somalian	Somalia	23	23	(Pagani et al. 2012)
Bantu	South Africa	94	25	(May, et al. 2013)
Mikea	Malagasy	18	18	(Pierron, et al. 2014)
Temoro	Malagasy	24	24	(Pierron et al. 2014)
Vezo	Malagasy	24	24	(Pierron et al. 2014)
Lebbo	Indonesia (Borneo)	15	15	(Pierron et al. 2014)
Bajo	Indonesia (Sulawesi)	31	31	(Pierron et al. 2014)
Indian	Singapore	83	-	(Teo, et al. 2009)
Malay	Singapore	89	25	(Teo et al. 2009)
Jehai	Malaysia	20	-	(Aghakhanian, et al. 2015)
Bateq	Malaysia	9	-	(Aghakhanian et al. 2015)
Mendriq	Malaysia	12	-	(Aghakhanian et al. 2015)
Kintaq	Malaysia	13	-	(Aghakhanian et al. 2015)
MahMeri	Malaysia	17	-	(Aghakhanian et al. 2015)
CheWong	Malaysia	6	-	(Aghakhanian et al. 2015)
Jakun	Malaysia	9	-	(Aghakhanian et al. 2015)
Temuan	Malaysia	12	-	(Aghakhanian et al. 2015)
Seletar	Malaysia	13	-	(Aghakhanian et al. 2015)

Onge	Andaman	10	-	(Reich, et al. 2011)
Dravidian	India	12	-	(Reich et al. 2011)
Besemah	Indonesia (Sumatra)	8	-	(Reich et al. 2011)
Semende	Indonesia (Sumatra)	9	-	(Reich et al. 2011)
Bidayuh	Malaysia (Borneo)	10	-	(Reich et al. 2011)
Nusa Tenggara	Indonesia	10	-	(Reich et al. 2011)
Moluccas	Indonesia	10	-	(Reich et al. 2011)
PNG Highlander	Papua New Guinea	24	-	(Reich et al. 2011)
Filipino**	Philippines	16	-	(Reich et al. 2011)
Philippines Negrito	Philippines	11	-	(Reich et al. 2011)
Taiwanese Aborigines	Taiwan	12	-	(Reich et al. 2011)
Fiji	Fiji	25	-	(Reich et al. 2011)
Polynesian	Polynesia	19	-	(Reich et al. 2011)
Sumbanese*	Indonesia (Sumba)	180	25	(Cox, et al. forthcoming)
Brahmin	India	47	25	(Mörseburg, et al. forthcoming)
Burmese	Myanmar	20	20	(Mörseburg et al. forthcoming)
Vietnamese	Vietnam	20	20	(Mörseburg et al. forthcoming)
Igorot	Philippines	21	21	(Mörseburg et al. forthcoming)
Filipino**	Philippines	16	16	(Mörseburg et al. forthcoming)
Dusun	Malaysia (Borneo)	17	17	(Mörseburg et al. forthcoming)
Murut	Malaysia (Borneo)	17	17	(Mörseburg et al. forthcoming)
Ma'anyan	Indonesia (Borneo)	162	25	(Kusuma, et al. 2016)
SK.Dayak	Indonesia (Borneo)	40	25	(Kusuma et al. 2016)
Ngaju	Indonesia (Borneo)	25	25	*this study*
Banjar	Indonesia (Borneo)	16	16	*this study*

Note: LDD: Low Density of SNP Dataset (40,272 SNPs); HDD: High Density of SNP Dataset (374,189 SNPs). *: Sumbanese of the LDD were sub-grouped in 3 locations of sampling for some analyses: West (W-Sumba), Central (C-Sumba), East (E-Sumba).; **: For some analyses Filipino individuals from Reich et al. (2011) and Mörseburg et al. (2016) were pooled.

Table S2. Results of f3-statistics showing the Z-score <-2 obtained for each combination on populations in the high density dataset to test for admixture. Only the 15 lowest Z-scores are shown for each tested group.

Tested group	Parental Pop A	Parental Pop B	f3	SD	Z-score
Bajo	Luhya	Igorot	-0,00074	0,00015	-4,83578
Bajo	S-African_Bantu	Igorot	-0,00075	0,00016	-4,79533
Bajo	Yoruba	Igorot	-0,00073	0,00015	-4,70864
Bajo	Eth-Somalian	Igorot	-0,00064	0,00015	-4,20585
Bajo	Somalian	Igorot	-0,00063	0,00015	-4,19374
Bajo	Gyan	Igorot	-0,00042	0,00014	-2,91720
Bajo	CEU	Igorot	-0,00038	0,00016	-2,34123
Banjar	Gyan	Igorot	-0,00117	0,00009	-12,53420
Banjar	Gyan	Maanyan	-0,00086	0,00007	-11,70650
Banjar	Somalian	Igorot	-0,00112	0,00010	-11,08290
Banjar	Eth-Somalian	Igorot	-0,00113	0,00010	-10,98400
Banjar	CEU	Maanyan	-0,00089	0,00008	-10,75080
Banjar	CEU	Igorot	-0,00113	0,00010	-10,74080
Banjar	Luhya	Igorot	-0,00110	0,00011	-10,40430
Banjar	S-African_Bantu	Igorot	-0,00110	0,00011	-10,30480
Banjar	Yoruba	Igorot	-0,00110	0,00011	-10,24280
Banjar	Eth-Somalian	Maanyan	-0,00077	0,00008	-9,87042
Banjar	Maanyan	Somalian	-0,00076	0,00008	-9,71601
Banjar	Gyan	Murut	-0,00074	0,00008	-9,13794
Banjar	S-African_Bantu	Maanyan	-0,00073	0,00008	-8,92003
Banjar	Maanyan	Luhya	-0,00069	0,00008	-8,65876
Banjar	CEU	Murut	-0,00081	0,00009	-8,63599
Burmese	Gyan	CHB	-0,00325	0,00006	-56,17750
Burmese	Vietnamese	Gyan	-0,00306	0,00006	-51,65090
Burmese	CEU	CHB	-0,00316	0,00007	-43,60350
Burmese	Vietnamese	CEU	-0,00311	0,00007	-42,84930
Burmese	CHB	Somalian	-0,00280	0,00007	-41,56630
Burmese	Eth-Somalian	CHB	-0,00283	0,00007	-40,63090
Burmese	Vietnamese	Somalian	-0,00260	0,00007	-37,29110
Burmese	Gyan	Murut	-0,00267	0,00007	-36,73700
Burmese	Vietnamese	Eth-Somalian	-0,00260	0,00007	-36,71400
Burmese	Gyan	Igorot	-0,00290	0,00008	-36,48650
Burmese	CHB	Luhya	-0,00257	0,00007	-35,51950
Burmese	S-African_Bantu	CHB	-0,00261	0,00007	-35,41820
Burmese	Yoruba	CHB	-0,00257	0,00007	-34,33380
Burmese	SK-Dayak	Gyan	-0,00220	0,00006	-34,22450
Burmese	Gyan	Maanyan	-0,00233	0,00007	-33,86420
Luzon	Gyan	Igorot	-0,00226	0,00008	-26,73450
Luzon	Somalian	Igorot	-0,00228	0,00009	-24,00030
Luzon	CEU	Igorot	-0,00233	0,00010	-23,76760
Luzon	Eth-Somalian	Igorot	-0,00227	0,00010	-23,51470
Luzon	Luhya	Igorot	-0,00229	0,00010	-22,86680
Luzon	Yoruba	Igorot	-0,00229	0,00010	-22,43360
Luzon	S-African_Bantu	Igorot	-0,00227	0,00010	-22,24460
Luzon	Vezo	Igorot	-0,00161	0,00008	-20,85780
Luzon	Mikea	Igorot	-0,00166	0,00008	-20,39340
Luzon	Temoro	Igorot	-0,00160	0,00008	-20,18860
Luzon	Burmese	Igorot	-0,00112	0,00006	-18,23300
Luzon	Igorot	Sumbanese	-0,00073	0,00006	-12,37340
Luzon	Malay-SG	Igorot	-0,00069	0,00006	-12,16120
Mikea	SK-Dayak	Yoruba	-0,00995	0,00010	-100,59000
Mikea	S-African_Bantu	Malay-SG	-0,00946	0,00009	-99,94540
Mikea	Ngaju	Yoruba	-0,00985	0,00010	-99,50600
Mikea	Yoruba	Maanyan	-0,01003	0,00010	-99,44410
Mikea	Yoruba	Murut	-0,01017	0,00010	-99,15970
Mikea	S-African_Bantu	Dusun	-0,01009	0,00010	-99,13510
Mikea	Vietnamese	S-African_Bantu	-0,00973	0,00010	-98,79300
Mikea	S-African_Bantu	Luzon	-0,00971	0,00010	-98,31260
Mikea	S-African_Bantu	Banjar	-0,00979	0,00010	-98,24890
Mikea	Malay-SG	Yoruba	-0,00939	0,00010	-97,83200
Mikea	Dusun	Yoruba	-0,01003	0,00010	-97,10260
Ngaju	Gyan	Maanyan	-0,00033	0,00006	-5,42691
Ngaju	CEU	Maanyan	-0,00030	0,00007	-4,34292
Ngaju	Gyan	Igorot	-0,00037	0,00009	-4,23838
Ngaju	Eth-Somalian	Maanyan	-0,00021	0,00006	-3,27339
Ngaju	Eth-Somalian	Igorot	-0,00030	0,00010	-3,01367
Ngaju	Maanyan	Somalian	-0,00019	0,00006	-2,98514
Ngaju	Luhya	Igorot	-0,00030	0,00010	-2,95500
Ngaju	Somalian	Igorot	-0,00028	0,00010	-2,92635
Ngaju	Yoruba	Igorot	-0,00029	0,00010	-2,85123
Ngaju	S-African_Bantu	Igorot	-0,00028	0,00010	-2,68118
Ngaju	CEU	Igorot	-0,00026	0,00010	-2,60884
Ngaju	S-African_Bantu	Maanyan	-0,00017	0,00007	-2,50506
Ngaju	Maanyan	Luhya	-0,00017	0,00007	-2,43966
Ngaju	Yoruba	Maanyan	-0,00016	0,00007	-2,32433
Ngaju	Gyan	Murut	-0,00011	0,00007	-1,43415
SK-Dayak	Maanyan	Burmese	-0,00031	0,00004	-7,90256
SK-Dayak	Gyan	Maanyan	-0,00044	0,00006	-7,43457
SK-Dayak	Maanyan	Sumbanese	-0,00030	0,00004	-7,41106
SK-Dayak	Vietnamese	Maanyan	-0,00027	0,00004	-7,27878
SK-Dayak	Malay-SG	Maanyan	-0,00023	0,00003	-7,08378
SK-Dayak	Maanyan	Bajo	-0,00035	0,00005	-7,02357
SK-Dayak	CEU	Maanyan	-0,00041	0,00007	-6,08851
SK-Dayak	Maanyan	CHB	-0,00024	0,00004	-5,80104
SK-Dayak	Gyan	Igorot	-0,00049	0,00008	-5,79169
SK-Dayak	Temoro	Maanyan	-0,00028	0,00005	-5,32602
SK-Dayak	Vezo	Maanyan	-0,00027	0,00005	-5,20714
SK-Dayak	Maanyan	Murut	-0,00021	0,00004	-5,04424
SK-Dayak	Maanyan	Luhya	-0,00034	0,00007	-4,91390
SK-Dayak	Luhya	Igorot	-0,00049	0,00010	-4,90414
SK-Dayak	S-African_Bantu	Maanyan	-0,00034	0,00007	-4,84829
Somalian	Yoruba	CEU	-0,00478	0,00010	-47,79240
Somalian	CEU	Luhya	-0,00448	0,00009	-47,43610
Somalian	S-African_Bantu	CEU	-0,00443	0,00010	-44,19020
Somalian	Gyan	Luhya	-0,00221	0,00009	-25,10260
Somalian	Gyan	Yoruba	-0,00237	0,00009	-24,97390
Somalian	S-African_Bantu	Gyan	-0,00205	0,00009	-21,81550
Somalian	Burmese	Luhya	-0,00036	0,00010	-3,47382
Somalian	Yoruba	Burmese	-0,00034	0,00011	-2,98567
Somalian	CHB	Luhya	-0,00013	0,00011	-1,19885

Luzon	CHB	Igorot	-0,00059	0,00006	-9,63134	Sumbanese	Luhya	Igorot	-0,00065	0,00011	-5,87579
Luzon	Vietnamese	Igorot	-0,00046	0,00006	-7,76522	Sumbanese	S-African_Bantu	Igorot	-0,00065	0,00011	-5,85895
Gyan	CEU	Burmese	-0,00234	0,00008	-30,74300	Sumbanese	Yoruba	Igorot	-0,00065	0,00011	-5,84816
Gyan	Malay-SG	CEU	-0,00235	0,00008	-28,72070	Sumbanese	Eth-Somalian	Igorot	-0,00047	0,00010	-4,53064
Gyan	CEU	Sumbanese	-0,00240	0,00009	-28,18800	Sumbanese	Somalian	Igorot	-0,00045	0,00010	-4,34370
Gyan	Ngaju	CEU	-0,00242	0,00009	-27,79990	Sumbanese	Gyan	Igorot	-0,00021	0,00009	-2,35079
Gyan	SK-Dayak	CEU	-0,00242	0,00009	-27,08320	Sumbanese	CEU	Igorot	-0,00012	0,00011	-1,08230
Gyan	Banjar	CEU	-0,00235	0,00009	-26,65450	Temoro	S-African_Bantu	Ngaju	-0,01130	0,00009	-124,99600
Gyan	Vietnamese	CEU	-0,00238	0,00009	-26,49640	Temoro	S-African_Bantu	SK-Dayak	-0,01141	0,00009	-124,52900
Gyan	Dusun	CEU	-0,00248	0,00009	-26,47620	Temoro	SK-Dayak	Yoruba	-0,01133	0,00009	-123,75300
Gyan	CEU	Maanyan	-0,00238	0,00009	-26,35860	Temoro	Ngaju	Yoruba	-0,01122	0,00009	-123,16300
Gyan	CEU	Murut	-0,00241	0,00009	-25,80560	Temoro	S-African_Bantu	Maanyan	-0,01147	0,00009	-122,94300
Gyan	CEU	Lebbo	-0,00240	0,00009	-25,68800	Temoro	S-African_Bantu	Banjar	-0,01118	0,00009	-122,91200
Gyan	CEU	Luzon	-0,00224	0,00009	-25,41830	Temoro	S-African_Bantu	Malay-SG	-0,01083	0,00009	-122,90400
Gyan	CEU	CHB	-0,00225	0,00009	-24,96840	Temoro	Banjar	Yoruba	-0,01112	0,00009	-121,74100
Gyan	CEU	Bajo	-0,00235	0,00010	-24,66920	Temoro	Yoruba	Maanyan	-0,01137	0,00009	-121,48900
Gyan	CEU	Igorot	-0,00231	0,00010	-23,36910	Temoro	Malay-SG	Yoruba	-0,01073	0,00009	-121,35500
Luhya	Eth-Somalian	Yoruba	-0,00072	0,00005	-15,55730	Temoro	Yoruba	Murut	-0,01149	0,00009	-121,07700
Luhya	Yoruba	Somalian	-0,00071	0,00005	-15,36080	Temoro	S-African_Bantu	Murut	-0,01157	0,00010	-120,65900
Luhya	Yoruba	CEU	-0,00101	0,00007	-14,90660	Temoro	Vietnamese	S-African_Bantu	-0,01114	0,00009	-120,07700
Luhya	S-African_Bantu	Somalian	-0,00062	0,00004	-13,93520	Temoro	S-African_Bantu	Luzon	-0,01108	0,00009	-119,93300
Luhya	Gyan	Yoruba	-0,00087	0,00006	-13,82580	Temoro	Yoruba	Luzon	-0,01099	0,00009	-118,51300
Luhya	S-African_Bantu	Eth-Somalian	-0,00059	0,00005	-12,47290	Vezo	S-African_Bantu	Ngaju	-0,01131	0,00009	-126,64300
Luhya	Yoruba	Burmese	-0,00068	0,00007	-10,42080	Vezo	S-African_Bantu	SK-Dayak	-0,01140	0,00009	-125,42300
Luhya	Malay-SG	Yoruba	-0,00068	0,00007	-10,24830	Vezo	S-African_Bantu	Maanyan	-0,01147	0,00009	-124,43700
Luhya	Yoruba	Sumbanese	-0,00067	0,00007	-10,19040	Vezo	S-African_Bantu	Malay-SG	-0,01079	0,00009	-123,28900
Luhya	Dusun	Yoruba	-0,00070	0,00007	-10,17090	Vezo	S-African_Bantu	Banjar	-0,01118	0,00009	-122,99600
Luhya	SK-Dayak	Yoruba	-0,00068	0,00007	-10,16430	Vezo	SK-Dayak	Yoruba	-0,01132	0,00009	-122,00400
Luhya	Ngaju	Yoruba	-0,00068	0,00007	-10,06140	Vezo	Ngaju	Yoruba	-0,01123	0,00009	-121,63800
Luhya	Yoruba	CHB	-0,00068	0,00007	-10,01850	Vezo	S-African_Bantu	Murut	-0,01161	0,00010	-121,03500
Luhya	Yoruba	Luzon	-0,00067	0,00007	-9,95678	Vezo	S-African_Bantu	Luzon	-0,01109	0,00009	-120,72900
Luhya	Vietnamese	Yoruba	-0,00068	0,00007	-9,92264	Vezo	Yoruba	Maanyan	-0,01138	0,00009	-120,61700
Malay-SG	Gyan	Maanyan	-0,00165	0,00006	-26,43780	Vezo	S-African_Bantu	Dusun	-0,01146	0,00010	-120,42300
Malay-SG	Gyan	Murut	-0,00190	0,00007	-26,38870	Vezo	Vietnamese	S-African_Bantu	-0,01113	0,00009	-120,37100
lMalay-SG	SK-Dayak	Gyan	-0,00144	0,00005	-26,33230	Vezo	Malay-SG	Yoruba	-0,01070	0,00009	-119,67100
Malay-SG	Gyan	Igorot	-0,00202	0,00008	-24,03160	Vezo	Banjar	Yoruba	-0,01112	0,00009	-119,60800
Malay-SG	CEU	Maanyan	-0,00168	0,00007	-23,49810	Vezo	Yoruba	Murut	-0,01153	0,00010	-118,59900
Malay-SG	SK-Dayak	CEU	-0,00151	0,00007	-23,18220	Vietnamese	Ngaju	CHB	-0,00062	0,00004	-16,10290
Malay-SG	CEU	Murut	-0,00196	0,00008	-23,08730	Vietnamese	SK-Dayak	CHB	-0,00054	0,00004	-14,28810
Malay-SG	Dusun	Gyan	-0,00159	0,00007	-22,56590	Vietnamese	Banjar	CHB	-0,00053	0,00004	-12,78690
Malay-SG	Ngaju	Gyan	-0,00121	0,00006	-21,92340	Vietnamese	Maanyan	CHB	-0,00052	0,00004	-12,52290
Malay-SG	SK-Dayak	Somalian	-0,00130	0,00006	-21,53510	Vietnamese	Malay-SG	CHB	-0,00043	0,00004	-11,57120
Malay-SG	Eth-Somalian	Murut	-0,00172	0,00008	-21,37210	Vietnamese	CHB	Murut	-0,00042	0,00005	-9,27405
Malay-SG	SK-Dayak	Eth-Somalian	-0,00129	0,00006	-21,28580	Vietnamese	CHB	Lebbo	-0,00047	0,00005	-9,09139
Malay-SG	Somalian	Murut	-0,00169	0,00008	-21,09710	Vietnamese	Dusun	CHB	-0,00033	0,00005	-7,09284
Malay-SG	Eth-Somalian	Maanyan	-0,00138	0,00007	-20,51370	Vietnamese	Burmese	Igorot	-0,00036	0,00005	-6,54163
Malay-SG	Maanyan	Somalian	-0,00137	0,00007	-20,33400	Vietnamese	CHB	Bajo	-0,00025	0,00005	-4,59526
Mikea	S-African_Bantu	SK-Dayak	-0,01002	0,00010	-103,48400	Vietnamese	CHB	Sumbanese	-0,00013	0,00004	-2,85187
Mikea	S-African_Bantu	Ngaju	-0,00991	0,00010	-102,49300	Vietnamese	Gyan	Igorot	-0,00020	0,00009	-2,18635
Mikea	S-African_Bantu	Maanyan	-0,01012	0,00010	-102,06900	Vietnamese	CEU	Igorot	-0,00012	0,00010	-1,17054
Mikea	S-African_Bantu	Murut	-0,01023	0,00010	-100,83700						

Table S3. f3-statistics table obtained for the tree configuration (Malagasy population; Asian population; Yoruba) using the Asian-SNP dataset.

Source 1	Source 2	Target	f3	SE	Z-score	SNPs
Han	Temoro	Yoruba	0.114	0.005	21.092	1595
Burmese	Temoro	Yoruba	0.108	0.005	21.899	1596
Vietnamese	Temoro	Yoruba	0.119	0.006	20.305	1596
Jehai	Temoro	Yoruba	0.109	0.006	18.594	1592
Bateq	Temoro	Yoruba	0.106	0.006	16.485	1590
Mendriq	Temoro	Yoruba	0.11	0.006	17.136	1593
Kintaq	Temoro	Yoruba	0.108	0.006	18.793	1594
Mah-Meri	Temoro	Yoruba	0.111	0.006	19.042	1591
CheWong	Temoro	Yoruba	0.114	0.006	17.852	1585
Jakun	Temoro	Yoruba	0.117	0.006	19.377	1590
Temuan	Temoro	Yoruba	0.115	0.005	21.534	1589
Seletar	Temoro	Yoruba	0.115	0.007	17.066	1593
Malay	Temoro	Yoruba	0.115	0.005	22.628	1597
Besemah	Temoro	Yoruba	0.119	0.006	21.528	1591
Semende	Temoro	Yoruba	0.123	0.006	21.331	1590
Bidayuh	Temoro	Yoruba	0.118	0.006	21.085	1589
Banjar	Temoro	Yoruba	0.121	0.006	21.923	1593
Ngaju	Temoro	Yoruba	0.12	0.005	22.258	1595
SK-Dayak	Temoro	Yoruba	0.121	0.005	23.247	1596
Ma'anyan	Temoro	Yoruba	0.122	0.006	21.937	1597
Lebbo	Temoro	Yoruba	0.117	0.005	21.457	1593
Murut	Temoro	Yoruba	0.119	0.005	22.637	1593
Dusun	Temoro	Yoruba	0.119	0.006	20.81	1594
Bajo	Temoro	Yoruba	0.116	0.005	21.905	1597
C-Sumbanese	Temoro	Yoruba	0.117	0.005	21.529	1597
E-Sumbanese	Temoro	Yoruba	0.113	0.005	20.879	1597
W-Sumbanese	Temoro	Yoruba	0.114	0.005	21.931	1597
Nusa_Tenga	Temoro	Yoruba	0.107	0.006	18.523	1590
Moluccas	Temoro	Yoruba	0.111	0.006	19.981	1587
PNG highlander	Temoro	Yoruba	0.086	0.008	11.34	1578
Philippine Negrito	Temoro	Yoruba	0.109	0.006	19.665	1587
Filipino	Temoro	Yoruba	0.12	0.006	20.97	1596
Igorot	Temoro	Yoruba	0.127	0.006	22.68	1593
Taiwan Aboriginese	Temoro	Yoruba	0.119	0.006	20.335	1594
Fiji	Temoro	Yoruba	0.103	0.005	19.196	1594
Polynesia	Temoro	Yoruba	0.116	0.006	20.188	1591
Han	Vezo	Yoruba	0.12	0.005	25.6	1620

Burmese	Vezo	Yoruba	0.112	0.005	24.617	1620
Vietnamese	Vezo	Yoruba	0.123	0.005	24.012	1621
Jehai	Vezo	Yoruba	0.111	0.005	20.357	1616
Bateq	Vezo	Yoruba	0.108	0.006	18.361	1614
Mendriq	Vezo	Yoruba	0.113	0.006	20.44	1619
Kintaq	Vezo	Yoruba	0.11	0.006	18.975	1615
Mah-Meri	Vezo	Yoruba	0.118	0.005	21.917	1616
CheWong	Vezo	Yoruba	0.121	0.006	19.101	1605
Jakun	Vezo	Yoruba	0.119	0.006	21.084	1612
Temuan	Vezo	Yoruba	0.119	0.005	23.297	1612
Seletar	Vezo	Yoruba	0.114	0.006	18.853	1615
Malay	Vezo	Yoruba	0.12	0.005	24.098	1622
Besemah	Vezo	Yoruba	0.124	0.005	24.584	1614
Semende	Vezo	Yoruba	0.126	0.005	24.929	1614
Bidayuh	Vezo	Yoruba	0.124	0.006	21.471	1613
Banjar	Vezo	Yoruba	0.127	0.005	25.264	1617
Ngaju	Vezo	Yoruba	0.125	0.005	24.353	1620
SK-Dayak	Vezo	Yoruba	0.126	0.005	26.05	1622
Ma'anyan	Vezo	Yoruba	0.128	0.005	25.407	1622
Lebbo	Vezo	Yoruba	0.122	0.006	21.471	1616
Murut	Vezo	Yoruba	0.124	0.005	22.842	1619
Dusun	Vezo	Yoruba	0.127	0.005	23.517	1619
Bajo	Vezo	Yoruba	0.117	0.005	24.57	1620
C-Sumbanese	Vezo	Yoruba	0.119	0.005	23.606	1621
E-Sumbanese	Vezo	Yoruba	0.116	0.005	23.617	1622
W-Sumbanese	Vezo	Yoruba	0.115	0.005	24.193	1622
Nusa_Tenga	Vezo	Yoruba	0.112	0.006	20.193	1615
Moluccas	Vezo	Yoruba	0.115	0.005	22	1613
PNG highlander	Vezo	Yoruba	0.082	0.008	10.777	1603
Philippine Negrito	Vezo	Yoruba	0.113	0.005	21.12	1609
Filipino	Vezo	Yoruba	0.123	0.005	24.999	1619
Igorot	Vezo	Yoruba	0.13	0.005	23.754	1618
Taiwan Aboriginese	Vezo	Yoruba	0.125	0.006	21.543	1618
Fiji	Vezo	Yoruba	0.106	0.005	19.416	1616
Polynesia	Vezo	Yoruba	0.121	0.006	21.102	1614
Han	Mikea	Yoruba	0.116	0.005	22.42	1621
Burmese	Mikea	Yoruba	0.11	0.005	22.506	1622
Vietnamese	Mikea	Yoruba	0.118	0.006	21.511	1621
Jehai	Mikea	Yoruba	0.106	0.005	20.259	1619
Bateq	Mikea	Yoruba	0.103	0.006	16.336	1616
Mendriq	Mikea	Yoruba	0.107	0.006	18.533	1619

Kintaq	Mikea	Yoruba	0.108	0.006	18.301	1616
Mah-Meri	Mikea	Yoruba	0.113	0.005	20.671	1615
CheWong	Mikea	Yoruba	0.115	0.006	19.484	1612
Jakun	Mikea	Yoruba	0.118	0.006	20.018	1616
Temuan	Mikea	Yoruba	0.115	0.005	21.211	1615
Seletar	Mikea	Yoruba	0.109	0.006	19.324	1617
Malay	Mikea	Yoruba	0.115	0.005	22.626	1622
Besemah	Mikea	Yoruba	0.122	0.006	21.96	1615
Semende	Mikea	Yoruba	0.12	0.006	20.337	1616
Bidayuh	Mikea	Yoruba	0.121	0.006	21.17	1616
Banjar	Mikea	Yoruba	0.122	0.005	22.223	1619
Ngaju	Mikea	Yoruba	0.119	0.006	21.055	1622
SK-Dayak	Mikea	Yoruba	0.123	0.005	23.21	1622
Ma'anyan	Mikea	Yoruba	0.123	0.006	21.426	1622
Lebbo	Mikea	Yoruba	0.121	0.006	21.713	1618
Murut	Mikea	Yoruba	0.118	0.005	21.946	1621
Dusun	Mikea	Yoruba	0.124	0.006	20.771	1621
Bajo	Mikea	Yoruba	0.116	0.005	21.556	1622
C-Sumbanese	Mikea	Yoruba	0.116	0.006	19.959	1622
E-Sumbanese	Mikea	Yoruba	0.113	0.006	19.793	1621
W-Sumbanese	Mikea	Yoruba	0.115	0.005	21.267	1621
Nusa_Tenga	Mikea	Yoruba	0.114	0.006	18.352	1618
Moluccas	Mikea	Yoruba	0.116	0.006	17.925	1614
PNG highlander	Mikea	Yoruba	0.089	0.008	10.992	1603
Philippine Negrito	Mikea	Yoruba	0.109	0.006	17.881	1613
Filipino	Mikea	Yoruba	0.121	0.006	21.286	1621
Igorot	Mikea	Yoruba	0.126	0.006	20.052	1618
Taiwan Aboriginese	Mikea	Yoruba	0.12	0.006	20.126	1620
Fiji	Mikea	Yoruba	0.107	0.006	18.074	1619
Polynesia	Mikea	Yoruba	0.12	0.006	18.604	1618

Table S4. F_{ST} genetic distances (lower diagonal) between 'Asian-SNP' Malagasy and data from the Asian populations of the high density dataset, and the corresponding standard deviation (upper diagonal).

F_{ST} \ S.D.	Vezo	Mikea	Temoro	CHB	Burmese	Vietnamese	Malay-SG	Banjar	Ma'anyan	Ngaju	SK-Dayak	Lebbo	Murut	Dusun	Bajo	Sumbanese	Filipino	Igorot
Vezo	0	0.005	0.006	0.004	0.004	0.003	0.004	0.003	0.004	0.003	0.003	0.004	0.004	0.004	0.003	0.003	0.003	0.003
Mikea	0.01	0	0.006	0.005	0.004	0.005	0.004	0.005	0.004	0.004	0.004	0.005	0.005	0.005	0.004	0.004	0.005	0.005
Temoro	0.007	0.016	0	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.005	0.004	0.005	0.004	0.004	0.004	0.005
CHB	0.035	0.041	0.031	0	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.001	0.001	0.001	0.002
Burmese	0.032	0.034	0.024	0.009	0	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.002
Vietnamese	0.026	0.032	0.021	0.007	0.009	0	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.002	0.001	0.001	0.001	0.002
Malay-SG	0.019	0.028	0.015	0.015	0.009	0.007	0	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.001	0.001	0.001	0.002
Banjar	0.021	0.029	0.016	0.02	0.015	0.009	0.004	0	0.001	0.001	0.001	0.002	0.002	0.001	0.001	0.002	0.001	0.002
Ma'anyan	0.026	0.036	0.021	0.028	0.024	0.017	0.009	0.008	0	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.002
Ngaju	0.021	0.028	0.016	0.021	0.015	0.011	0.003	0.004	0.008	0	0.001	0.002	0.001	0.002	0.001	0.001	0.001	0.002
SK-Dayak	0.018	0.025	0.015	0.019	0.014	0.01	0.002	0.002	0.005	0.001	0	0.002	0.001	0.001	0.001	0.001	0.001	0.002
Lebbo	0.043	0.046	0.036	0.042	0.036	0.034	0.024	0.027	0.034	0.027	0.025	0	0.002	0.002	0.002	0.002	0.002	0.003
Murut	0.033	0.041	0.027	0.028	0.028	0.021	0.016	0.018	0.024	0.016	0.015	0.035	0	0.002	0.001	0.002	0.001	0.002
Dusun	0.033	0.041	0.026	0.028	0.026	0.02	0.017	0.016	0.022	0.019	0.015	0.036	0.023	0	0.001	0.002	0.001	0.002
Bajo	0.028	0.032	0.018	0.02	0.016	0.014	0.008	0.011	0.019	0.011	0.009	0.028	0.02	0.019	0	0.001	0.001	0.002
Sumbanese	0.031	0.037	0.019	0.025	0.02	0.019	0.011	0.015	0.021	0.014	0.012	0.032	0.025	0.024	0.008	0	0.001	0.002
Filipino	0.023	0.028	0.017	0.013	0.015	0.008	0.006	0.008	0.013	0.009	0.006	0.028	0.013	0.012	0.007	0.012	0	0.002
Igorot	0.051	0.056	0.043	0.047	0.05	0.039	0.037	0.038	0.042	0.038	0.037	0.057	0.04	0.041	0.038	0.044	0.026	0

Table S5. GLOBETROTTER inferred dates (in both generations from present and years, bootstrap 95% CIs given in parenthesis; YBP = years before present), admixing sources, and percentage (%) of admixture for each population of interest. We assume a generation time of 25 years. R2 corresponds to the goodness-of-fit of the tested model. FQ1 and FQ2 correspond respectively to the fit of a single admixture event and the fit of the first two principal components capturing the admixture events. M corresponds to the additional R2 explained by adding a second date versus assuming only a single date of admixture (M>0.35 to infer multiple dates event). The admixture models presented in the table correspond to the best fit models considering the 'best guess'. P: p-value of evidence of any detectable admixture event obtained after 100 bootstrap resamplings of the NULL procedure. Date confidence intervals are based on 100 bootstrap replicates of the date inference.

Target Group	**Analysis**	**P**	**R^2**	**FQ_1**	**FQ_2**	**M**	**Best-guess**	**1-date (gen.)**	**1-date (YBP)**	**1-date Best Surrogate 1**	**1-date Best Surrogate 2**	**2-date (gen.)**	**2-date (YBP)**	**2-date Best Surrogate 1**	**2-date Best Surrogate 2**
Vezo	main	< 0.01	0.996	1	1	0.095	one-date	25 (24-27)	625 (600-675)	Banjar (39%)	S.A.Bantu (61%)	-	-	-	-
	main.null	< 0.01	0.996	1	1	0.012	one-date	25 (24-27)	625 (600-675)	Banjar (38%)	S.A.Bantu (62%)	-	-	-	-
Mikea	main	< 0.01	0.993	1	1	0.142	one-date	27 (25-29)	675 (625-725)	Banjar (36%)	S.A.Bantu (64%)	-	-	-	-
	main.null	< 0.01	0.991	1	1	0.255	one-date	27 (25-29)	675 (625-725)	Banjar (36%)	S.A.Bantu (64%)	-	-	-	-
Temoro	main	< 0.01	0.996	1	1	0.076	one-date	30 (29-31)	750 (725-775)	Banjar (37%)	S.A.Bantu (63%)	-	-	-	-
	main.null	< 0.01	0.995	1	1	0.102	one-date	30 (28-31)	750 (725-775)	Banjar (37%)	S.A.Bantu (63%)	-	-	-	-
Banjar	main	< 0.01	0.621	1	1	0.078	one-date	17 (11-20)	425 (275-500)	Ma'anyan (23%)	Malay (77%)	-	-	-	-
	main.null	< 0.01	0.615	1	1	0.042	one-date	23 (14-32)	575 (350-800)	Ma'anyan (24%)	Malay (76%)	-	-	-	-
Ngaju	main	< 0.01	0.521	1	1	0.082	one-date	16 (9-19)	400 (225-475)	Ma'anyan (19%)	Malay (81%)	-	-	-	-
	main.null	< 0.01	0.554	1	1	0.045	one-date	20 (17-23)	500 (425-575)	Ma'anyan (43%)	Malay (57%)	-	-	-	-
SK Dayak	main	< 0.01	0.463	1	1	0.127	one-date	12 (7-16)	300 (175-400)	Malay (21%)	Ma'anyan (79%)	-	-	-	-
	main.null	< 0.01	0.615	1	1	0.084	one-date	19 (11-35)	475 (275-875)	Malay (27%)	Ma'anyan (73%)	-	-	-	-
Malay	main	< 0.01	0.959	0.972	0.989	0.481	multiple-dates	7 (2-10)	175 (50-250)	Burmese (24%)	Filipino (76%)	42 (29-58)	1050 (725-1450)	Brahmin (14%)	Filipino (86%)
	main.null	< 0.01	0.94	0.979	0.992	0.514	multiple-dates	9 (4-15)	225 (100-375)	Burmese (16%)	Filipino (84%)	66 (43-111)	1650 (1075-2775)	Brahmin (19%)	Filipino (81%)
Bajo	main	< 0.01	0.946	0.882	0.998	0.22	one-date-multiway	35 (32-40)	875 (800-1000)	Sumbanese (23%), Brahmin (12%)	Banjar (25%), Filipino (25%), Malay (11%), Ngaju (6%)	-	-	-	-
	main.null	< 0.01	0.953	0.96	0.999	0.22	one-date-multiway	47 (43-54)	1175 (1075-1350)	Sumbanese (27%), Brahmin (15%)	Banjar (23%), Filipino (27%), Ngaju (8%)	-	-	-	-

Table S6. Estimated dates of admixture in Malagasy ethnic groups using Banjar and South African Bantu data as parental populations with ALDER.

Test Pop	**Ref A**	**Ref B**	**p-value**	**2-ref z-score**	**1-ref z-score A**	**1-ref z-score B**	**max decay diff (%)**	**2-ref decay (gen.)**	**2-ref amp_exp**
Temoro	Banjar	South Africa Bantu	5.1e-81	19.34	17.87	22.90	0.03	29.16 +/- 1.08	0.0018 +/- 0.0001
Mikea	Banjar	South Africa Bantu	6.9e-85	19.79	18.70	11.12	0.09	25.21 +/- 1.27	0.0017 +/- 0.0001
Vezo	Banjar	South Africa Bantu	2.6e-69	17.89	18.88	15.81	0.01	22.29 +/- 1.25	0.0017 +/- 0.0001

Supplementary Methods

Sample collection and ethics

A total of 41 DNA samples were analyzed from two groups in Southeast Borneo: the Banjar (n=16) and the Ngaju (n=25) that were collected in Banjarmasin city (Annex E: Figure S1). Blood samples were collected from healthy adult donors, all of whom provided written informed consent. DNA was extracted using a standard salting-out procedure. All participants were surveyed for language affiliation, current residence, familial birthplaces, and a genealogy of four generations to establish ancestry. This study was approved by the Research Ethics Commission of the Eijkman Institute for Molecular Biology (Jakarta, Indonesia). Genome-wide SNP genotyping for individuals from the two groups were performed using the Illumina Human Omni Express Bead Chip-24 v1.0 (Illumina Inc., San Diego, CA), which characterizes 730,525 single nucleotide markers regularly spaced across the genome. The accession number for the Southeast Borneo samples genotyped for this study is EGA: EGAS00001001841.

Dataset

We gathered data from previously published studies on populations from Madagascar, Southeast Asia, South Asia, East Asia, East Africa, South Africa, and Europe (Annex E: Table S1). Two datasets were compiled respective to their analytical use: a low SNP density dataset of populations covering a large geographical area; and a high SNP density dataset of populations composed by a subset of populations of the latter dataset. To avoid any statistical bias that could be induced by a size effect due to populations over-represented in the high SNP density dataset, we randomly selected a maximum of 31 individuals in each group, such as each population has a number of individuals between 15 and 31. Quality controls were applied using Plink v1.9 (Chang et al., 2015) to filter for i) close relatives, using an Identity-by-Descent (IBD) estimation with upper threshold of 0.25 (second degree relatives); ii) SNPs that failed the Hardy-Weinberg exact (HWE) test ($P < 10^{-6}$) were excluded; iii) samples with a call rate <0.99 and displayed missing rates >0.05 across all samples in each population were excluded; and iv) variants in high linkage disequilibrium ($r^2 > 0.5$) were also removed for the low density dataset. After filtering, the low SNP density dataset included 2,183 individuals from 61 populations genotyped for 40,272; and the SNPs high SNP density dataset was composed of 551 individuals from 24 populations genotyped for 374,189 SNPs. All genotypes of the high SNP density dataset were then phased together with SHAPEIT v2.r790

(Delaneau et al., 2012) using the 1000 Genomes phased data (Delaneau et al., 2014) as reference panel and the HapMap phase 2 genetic map (International HapMap, 2005).

Statistical Analyses

The low density dataset was described by the following analyses. Principal Components Analysis was computed with EIGENSOFT v6.0.1 (Patterson et al., 2006). ADMIXTURE v1.23 (Alexander et al., 2009) was used to estimate the profile of individual genomic ancestries using maximum likelihood for components K = 2 to K = 18. Ten replicates were run at each value of K with different random seeds, then merged and assessed for clustering quality using CLUMPP (Jakobsson & Rosenberg, 2007), and the cross-validation value was calculated to determine the optimal number of genomic components (here, K = 16). ADMIXTURE and PCA plots were generated with Genesis (Buchmann & Hazelhurst, 2014). Three-population (f3) statistics (Patterson et al., 2012) were computed for each trios of populations of the low SNP density dataset to identify groups showing potential recent admixture event. TreeMix v1.12 analysis (Pickrell & Pritchard, 2012) was performed with all Asian populations and three representative of African groups (Yoruba, Luhya, South African Bantu), to estimate gene flows from Asian groups and their relative intensity, with blocks of 200 SNPs to account for linkage disequilibrium and migration edges added sequentially until the model explained 99% of the variance (the TreeMix outputs in Newick format were visualized with MEGA v7.0 (Tamura et al., 2013)).

Population structure of the phased high density dataset was evaluated using the fineSTRUCTURE v2.07 package (Lawson et al., 2012). It uses the detection of shared IBD fragments between each pair of individuals, without self-copying, calculated with Chromopainter v2.0 (Lawson et al., 2012) to perform a model-based Bayesian clustering of genotypes. From the results, a coancestry heat map and a dendrogram were inferred to visualize the number of clusters statistically defined that would describe relevantly the data. FineSTRUCTURE v2 identified 37 groups of individuals that can be statistically defined as natural populations (Lawson et al., 2012), according to their shared Identity-By-Descent (IBD) (bootstrapped nodes > 60%; Annex E: Figure S7). Most of these correspond to anthropologically defined populations, such as the Banjar or the Ngaju, with few exceptions (e.g., the sub-structuration of the Kenyan Luhya group). Although each Malagasy individual has its closest connections with other Malagasy, regardless of their ethnicity, these ethnic groups were treated separately for anthropological interest. Haplotype sharing between pairs of individuals was estimated from the phased high SNP density dataset by the Refined IBD algorithm of Beagle v4.0 (Browning & Browning, 2007; Browning & Browning, 2013), filtering for detected fragments with a logarithm of odds ratio (LOD)>3. Ten iterations were performed, randomizing the seed number for each run, and the overlapping shared fragments were merged to favour the detection of long shared IBD. Detected fragments between the same pairs of individuals were summed up and visualized with Cytoscape v3.2.1 (Shannon et al., 2003). IBD sharing data with Malagasy individuals was also averaged per Asian population to geographically plot a gradient map with Surfer v12.0 using the Kriging method. All maps used in the present study were generated using Global Mapper v.15 software. (http://www.bluemarblegeo.com/products/global-mapper.php). Local ancestry analysis in Malagasy individuals was performed with PCAdmix v1.0 (Brisbin et al., 2012) using two parental metapopulations of 100 individuals of African ancestry (randomly selected in Yoruba, South African Bantu, Kenyan Luhya and Somali groups), and of Asian ancestry (randomly selected in Han, Igorot, Ma'anyan and Malay groups). The phased Malagasy data were screened using the linkage disequilibrium information to define the probability of common ancestry of each Malagasy haplotype with each 'parental' metapopulation. The Viterbi algorithm was then used to mask all haplotypes according to one or the other ancestry in the Malagasy individuals. Ancestry-specific PCAs and F_{ST} calculations were realized with EIGENSOFT v6.0.1 (Patterson et al., 2006). Three-population (f3) statistics (Patterson et al., 2012) and TreeMix v1.12 analysis (Pickrell & Pritchard, 2012) were also performed on this masked dataset (1,664 SNPs for 18 populations). Haplotype 'painting' with Chromopainter v2 (Lawson et al., 2012) was realized on the high density of SNP dataset, defining each cluster of populations as target or donor/surrogate according to the anthropological question addressed. Mutational rates and Ne parameters were first estimated with an Estimation-Maximization (EM) algorithm running Chromopainter v2 on all 22 autosomes for the entire dataset with 10 iterations (Lawson et al., 2012). The weighted average of these parameters, according to the SNP coverage of each used chromosomes and the number of individuals, were then used to compute the chromosomal painting. Each cluster of populations has been successively identified as target and the others as surrogates (at the exclusion of the Malagasy cluster which has not been used as surrogate). The obtained painted chromosomes for each cluster were used in GLOBETROTTER v1.0 (Hellenthal et al., 2014) to estimate the ratios and the dates of the potential admixture events characterizing them. Coancestry curves were estimated with and without standardization with a 'NULL' individual, and consistency between each estimated parameters was checked. A hundred bootstrap resamplings were performed to estimate the p-value of the admixture events (considering the 'NULL' individual) and the 95% confidence interval for the obtained dates (without the 'NULL' individual). The 'best-guess' scenario given by GLOBETROTTER v1.0 (Hellenthal et al., 2014) was considered for each target population. Using the parental populations given by GLOBETROTTER v1.0 (Hellenthal et al., 2014), dates of admixture were also estimated by ALDER v1 (Loh et al., 2013).

REFERENCES

Aghakhanian F, Yunus Y, Naidu R, Jinam T, Manica A, Hoh BP, Phipps ME 2015. Unravelling the genetic history of Negritos and indigenous populations of Southeast Asia. Genome Biol Evol 7: 1206-1215. doi: 10.1093/gbe/evv065

Alexander DH, Novembre J, Lange K 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19: 1655-1664. doi: 10.1101/gr.094052.109

Brisbin A, Bryc K, Byrnes J, Zakharia F, Omberg L, Degenhardt J, Reynolds A, Ostrer H, Mezey JG, Bustamante CD 2012. PCAdmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. Human Biology 84: 343-364. doi: 10.3378/027.084.0401

Browning BL, Browning SR 2013. Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics 194: 459-471. doi: 10.1534/genetics.113.150029

Browning SR, Browning BL 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. The American Journal of Human Genetics 81: 1084-1097. doi: 10.1086/521987

Buchmann R, Hazelhurst S. 2014. Genesis Manual. University of the Witwatersrand, Johannesburg.

Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4: 7. doi: 10.1186/s13742-015-0047-8

Cox MP, Hudjashov G, Sim A, Savina O, Karafet TM, Sudoyo H, Lansing JS forthcoming. Small traditional human communities sustain genomic diversity over microgeographic sscales despite linguistic isolation. Molecular Biology and Evolution.

Delaneau O, Marchini J, Genomes Project C, Genomes Project C 2014. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nature Communications 5: 3934. doi: 10.1038/ncomms4934

Delaneau O, Marchini J, Zagury JF 2012. A linear complexity phasing method for thousands of genomes. Nat Methods 9: 179-181. doi: 10.1038/nmeth.1785

GoldenSoftware. 2014. Surfer v.12.01.

Hellenthal G, Busby GB, Band G, Wilson JF, Capelli C, Falush D, Myers S 2014. A genetic atlas of human admixture history. Science 343: 747-751. doi: 10.1126/science.1243518

International HapMap C 2005. A haplotype map of the human genome. Nature 437: 1299-1320. doi: 10.1038/nature04226

International HapMap C, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Gibbs RA, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarrol SA, Nemesh J, Dermitzakis E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghori MJ, McGinnis R, McLaren W, Pollack S, Price AL, Schaffner SF, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE 2010. Integrating common and rare genetic variation in diverse human populations. Nature 467: 52-58. doi: 10.1038/nature09298

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: 1801-1806. doi: 10.1093/bioinformatics/btm233

Kusuma P, Brucato N, Cox MP, Pierron D, Razafindrazaka H, Adelaar A, Sudoyo H, Letellier T, Ricaut F-X 2016. Contrasting Linguistic and Genetic Influences during the Austronesian Settlement of Madagascar. Scientific Reports 6:26066. doi: doi: 10.1038/srep26066

Lawson DJ, Hellenthal G, Myers S, Falush D 2012. Inference of population structure using dense haplotype data. PLoS Genet 8: e1002453. doi: 10.1371/journal.pgen.1002453

Loh PR, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, Berger B 2013. Inferring admixture histories of human populations using linkage disequilibrium. Genetics 193: 1233-1254. doi: 10.1534/genetics.112.147330

May A, Hazelhurst S, Li Y, Norris SA, Govind N, Tikly M, Hon C, Johnson KJ, Hartmann N, Staedtler F, Ramsay M 2013. Genetic diversity in black South Africans from Soweto. BMC Genomics 14: 644. doi: 10.1186/1471-2164-14-644

Mörseburg A, Pagani L, Ricaut F-X, Yngvadottir B, Harney E, Castillo C, Antao T, Hoogervorst T, Kusuma P, Razafindrazaka H, Cardona A, Pierron D, Letellier T, Wee J, Abdullah M, Metspalu M, Kivisild T forthcoming. Multi-layered population structure and local adaptations in Island Southeast Asians. European Journal of Human Genetics.

Pagani L, Kivisild T, Tarekegn A, Ekong R, Plaster C, Gallego Romero I, Ayub Q, Mehdi SQ, Thomas MG, Luiselli D, Bekele E, Bradman N, Balding DJ, Tyler-Smith C 2012. Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am J Hum Genet 91: 83-96. doi: 10.1016/j.ajhg.2012.05.015

Patterson N, Price AL, Reich D 2006. Population structure and eigenanalysis. PLoS Genet 2: e190. doi: 10.1371/journal.pgen.0020190

Patterson NJ, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D 2012. Ancient admixture in human history. Genetics 192: 1065-1093. doi: 10.1534/genetics.112.145037

Pickrell JK, Pritchard JK 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8: e1002967. doi: 10.1371/journal.pgen.1002967

Pierron D, Razafindrazaka H, Pagani L, Ricaut F-X, Antao T, Capredon M, Sambo C, Radimilahy C, Rakotoarisoa J-A, Blench RM, Letellier T, Kivisild T 2014. Genome-wide evidence of Austronesian–Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. Proceedings of the National Academy of Sciences 111: 936-941. doi: 10.1073/pnas.1321860111

Reich D, Patterson N, Kircher M, Delfin F, Nandineni MR, Pugach I, Ko AM, Ko YC, Jinam TA, Phipps ME, Saitou N, Wollstein A, Kayser M, Paabo S, Stoneking M 2011. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. Am J Hum Genet 89: 516-528. doi: 10.1016/j.ajhg.2011.09.005

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research 13: 2498-2504. doi: 10.1101/gr.1239303

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30: 2725-2729. doi: 10.1093/molbev/mst197

Teo YY, Sim X, Ong RT, Tan AK, Chen J, Tantoso E, Small KS, Ku CS, Lee EJ, Seielstad M, Chia KS 2009. Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations. Genome Res 19: 2154-2162. doi: 10.1101/gr.095000.109

Annex H. Frequency of mitochondrial Polynesian motif (B4a1a1) and Malagasy motif (B4a1a1b) haplogroup in the populations involved in this thesis. Malagasy motif detection was performed using both RFLP on two diagnostic SNPs (1473 and 3423), and through whole mitogenome sequencing.

No.	Population	Polynesian motif	Malagasy motif
1	Ma’anyan	0 (0%)	0 (0%)
2	SK Dayak	0 (0%)	0 (0%)
3	Ngaju	0 (0%)	0 (0%)
4	Banjar	1 (1%)	0 (0%)
5	Samihim	0 (0%)	0 (0%)
6	Bajo Kotabaru	0 (0%)	0 (0%)
7	Bajo Derawan	0 (0%)	0 (0%)
8	Lebbo’	0 (0%)	0 (0%)
9	Bajo Kendari	1 (4%)	0 (0%)
10	Bugis	2 (4%)	0 (0%)
11	Mandar	0 (0%)	0 (0%)
12	North Maluku	8 (21%)	0 (0%)

Annex I. Table of GLOBETROTTER admixture analyses on Malagasy populations by involving all Indonesian populations generated data of genome-wide SNPs on 231,800 SNPs dataset.

Target Pop	**Null inds.**	**Scenario**	**Date 1 (gens. ago**	**Population 1.1**	**%**	**Population 1.2**	**%**
Vezo	null0	One-date	24.80	SAfrican_Bantu	61%	ISEA_Banjar	39%
	null1	One-date	25.60	SAfrican_Bantu	61%	ISEA_Banjar	39%
Mikea	null0	One-date	27.20	SAfrican_Bantu	64%	ISEA_Banjar	36%
	null1	One-date	26.90	SAfrican_Bantu	64%	ISEA_Banjar	36%
Temoro	null0	One-date	30.1	SAfrican_Bantu	62%	ISEA_Banjar	38%
	null1	One-date	31.0	SAfrican_Bantu	62%	ISEA_Banjar	38%

Annex J. Short Curriculum Vitae of the author, comprising the list of publications.

Name : Pradiptajati Kusuma
Date of Birth : Surabaya, 4 May 1989
Occupation : Research Assistant
Office : Eijkman Institute for Molecular Biology
Jalan Diponegoro 69, Jakarta 10430, Indonesia
Phone +62-21-3917131; +62-21-3148695
Email : pradipta@eijkman.go.id;
pradiptajati.kusuma@univ-tlse3.fr
Education :
1. Bachelor Degree (B.Sc.) in Biological Science, Bandung Institute of Technology, Bandung, Indonesia (2010)
2. Master Degree (M.Biomed.) in Biomedical Science, University of Indonesia, Jakarta, Indonesia (2013)

Research Focus :
1. Molecular Anthropology in Indonesia archipelago and Southeast Asia
2. Human dispersal through Indian Ocean in both directions: Africa to Indonesia, and Indonesia to Africa/Madagascar

Languages : Bahasa Indonesia (native), English (fluent), French (basic), Arabic (basic reading)

Past and present appointments :

1. Research Assistant (Dec 2010 – now), Lab of Human Genome Diversity and Diseases. Eijkman Institute for Molecular Biology, Jakarta, Indonesia
2. Doctoral Student (Feb 2014 – now), Laboratoire *Anthropologie Moléculaire et Imagerie de Synthèse* (AMIS), Université Paul Sabatier – Toulouse III, France
3. Master Student (Feb 2011 – Jan 2013), Lab of Human Genetic Diversity and Diseases (in Eijkman Institute Jakarta). University of Indonesia, Jakarta, Indonesia
4. Research Assistant (Jan 2010 – Jul 2010), Institute of Research and Community Development. Bandung Institute of Technology, Bandung, Indonesia
5. Undergraduate Research Student (Jul 2009 – Oct 2010), Lab of Physiology and Biomedical Science. Bandung Institute of Technology, Bandung, Indonesia
6. Technical Assistant (Jul 2009 – Jan 2010), Lab of Developmental Biology, Bandung Institute of Technology. Bandung, Indonesia

Awards and academic achievements :

1. Scholarship award for Doctoral Study from Ministry of Education, Indonesia, and the Embassy of France in Indonesia, France. Feb 2014 - now
2. Graduated from Master Program in Biomedical Science, Faculty of Medicine, University of Indonesia, with mention: Very Good. Feb 2013
3. Scholarship award for Master Study from Ministry of Education, Indonesia. Jan 2011 – Feb 2013
4. Graduated from Bachelor Program in Biological Science, School of Life Science and Technology, Bandung Institute of Technology, with mention: Cum Laude. Oct 2010
5. Dean's List for Academic Achievement, Bachelor Program in Biological Science, School of Life Science and Technology, Bandung Institute of Technology. 2007 – 2010
6. Scholarship award for Undergraduate Study from Ministry of Education, Indonesia. Jul 2006 – Oct 2010

Publications, conferences, and workshops :

Theses

Study of North Maluku Genetic Background using Mitochondrial DNA.
Master Thesis for Graduate Study in University of Indonesia.

DNA-Based Identification of Flesh Flies Larvae (Diptera: Sarcophagidae) in Jakarta using Cytochrome Oxidase subunit I DNA Barcode.
Thesis for Undergraduate Study in Bandung Institute of Technology.

Publications

1. **Kusuma, P.**, N. Brucato, M.P. Cox, T. Letellier, A. Manan, C. Nuraini, P. Grangé, H. Sudoyo, F.-X Ricaut. 2017. The Last Sea Nomads of the Indonesian Archipelago: Genomic Origins and Dispersal. *European Journal of Human Genetics*. AOP: 17 May 2017. doi: 10.1038/ejhg.2017.88
2. **Kusuma, P**., N. Brucato, M.P. Cox, C. Nuraini, T. Letellier, P. Grangé, H. Sudoyo and F.-X. Ricaut. 2017. A Genomic Perspective of the Origin and Dispersal of the Bajo Sea Nomads in Indonesia. In B. Bellina, R. Blench and J.-C. Gallipaud, *Sea Nomads' Past*. National University of Singapore Press: Singapore. In Press.

3. Brucato, N., **P. Kusuma**, P. Beaujard, H. Sudoyo, M.P. Cox and F.-X. Ricaut. 2017. Genomic Admixture Tracks Pulses of Economic Activity over 2,000 Years in the Indian Ocean Trading Network. *Scientific Reports*. In Press.
4. Skoglund, P., C. Posth, K. Sirak, M. Spriggs, F. Valentin, S. Bedford, G.R. Clark, C. Reepmeyer, F. Petchey, D. Fernandes, Q. Fu, E. Harney, M. Lipson, S. Mallick, M. Novak, N. Rohland, K. Stewardson, S. Abdullah, M.P. Cox, F.R. Friedlaender, J.S. Friedlaender, T. Kivisild, G. Koki, **P. Kusuma**, D.A. Merriwether, F.-X. Ricaut, J.T.S. Wee, N. Patterson, J. Krause, R. Pinhasi and D. Reich. 2016. Genomic Insights into the Peopling of the Southwest Pacific. *Nature.* 538: 510-513. doi:10.1038/nature19844
5. Brucato, N.*, **P. Kusuma***, M.P. Cox, D. Pierron, G. Purnomo, T. Letellier, T. Kivisild, A. Adelaar, H. Sudoyo and F.-X. Ricaut. 2016. Malagasy Genetic Ancestry Comes from an Historical Malay Trading Post in Southeast Borneo. *Molecular Biology and Evolution*. 33: 2396-2400. doi: 10.1093/molbev/msw117
6. **Kusuma, P**., N. Brucato, M.P. Cox, D. Pierron, H. Razafindrazaka, A. Adelaar, H. Sudoyo, T. Letellier and F.-X. Ricaut. Contrasting Linguistic and Genetic Influences during the Austronesian Settlement of Madagascar. *Scientific Reports.* 6: 26066. doi: 10.1038/srep26066
7. Mörseburg A., L. Pagani, F-X Ricaut, B. Yngvadottir, E. Harney, C. Castillo, T. Hoogervorst, T. Antao, **P. Kusuma**, N. Brucato, A. Cardona, D. Pierron, T. Letellier, J. Wee, S. Abdullah, M. Metspalu, T. Kivisild. Multi-Layered Population Structure in Island Southeast Asians. *European Journal of Human Genetics.* 24: 1605-1611. doi: 10.1038/ejhg.2016.60
8. **Kusuma, P**., M.P. Cox, N. Brucato, H. Sudoyo, T. Letellier and F.-X. Ricaut. 2015. Western Eurasian Genetic Influences in the Indonesian Archipelago. *Quaternary International*. 416: 243-248. doi: 10.1016/j.quaint.2015.06.048
9. **Kusuma, P.**, M.P. Cox, D. Pierron, H. Razafindrazaka, N. Brucato, L. Tonasso, H.L. Suryadi, T. Letellier, H. Sudoyo and F.-X. Ricaut. 2015. Mitochondrial DNA and the Y Chromosome Suggest the Settlement of Madagascar by Indonesian Sea Nomad Populations. *BMC Genomics.* 16: 191. doi: 10.1186/s12864-015-1394-7
10. Karmin, M., L. Saag, M. Vicente, M.A.W. Sayres, M. Järve, U.G. Talas, S. Rootsi, A.-M. Ilumäe, R. Mägi, M. Mitt, L. Pagani, T. Puurand, Z. Faltyskova, F. Clemente, A. Cardona, E. Metspalu, H. Sahakyan, B. Yunusbayev, G. Hudjashov, M. DeGiorgio, E.-L. Loogväli, C. Eichstaedt, M. Eelmets, G. Chaubey, K. Tambets, S. Litvinov, M. Mormina, Y. Xue, Q. Ayub, G. Zoraqi, T.S. Korneliussen, F. Akhatova, J. Lachance, S. Tishkoff, K.

Momynaliev, F.-X. Ricaut, **P. Kusuma**, H. Razafindrazaka, D. Pierron, M.P. Cox, G.N.N. Sultana, R. Willerslev, C. Muller, M. Westaway, D. Lambert, V. Skaro, L. Kovačević, S. Turdikulova, D. Dalimova, R. Khusainova, N. Trofimova, V. Akhmetova, I. Khidiyatova, D.V. Lichman, J. Isakova, E. Pocheshkhova, Z. Sabitov, N.A. Barashkov, P. Nymadawa, E. Mihailov, J.W.T. Seng, I. Evseeva, A.B. Migliano, S. Abdullah, G. Andriadze, D. Primorac, L. Atramentova, O. Utevska, L. Yepiskoposyan, D. Marjanović, A. Kushniarevich, D.M. Behar, C. Gilissen, L. Vissers, J. Veltman, E. Balanovska, M. Derenko, B. Malyarchuk, A. Metspalu, S. Fedorova, A. Eriksson, A. Manica, F. Mendez, T.M. Karafet, K. Veeramah, N. Bradman, M.F. Hammer, L.P. Osipova, O. Balanovsky, E.K. Khusnutdinova, K. Johnsen, M. Remm, M.G. Thomas, C. Tyler-Smith, P.A. Underhill, E. Willerslev, R. Nielsen, M. Metspalu, R. Villems and T. Kivisild. 2015. A Recent Bottleneck of Y Chromosome Diversity Coincides with a Global Change in Culture. *Genome Research*. 25: 459-466. doi: 10.1101/gr.186684.114

Conferences

1. **Kusuma P**, Brucato N, Cox MP, Letellier T, Grangé P, Sudoyo H, Ricaut F-X. The last sea nomads of the Indonesian archipelago: Genomic origins and dispersal. *Abstract to the 6th Conference on DNA Polymorhisms in Human Populations*. National History Museum, Paris, France, Nov 2016. Oral presentation
2. Brucato N, **Kusuma P**, Pierron D, Letellier T, Sudoyo H, Cox MP, Ricaut F-X. Genetic legacy of the Indian Ocean trading network. *Abstract to the 6th Conference on DNA Polymorhisms in Human Populations*. National History Museum, Paris, France, Nov 2016. Poster and oral presentation
3. **Kusuma P**, Brucato N, Cox MP, Pierron D, Grangé P, Leavesley M, Gessain A, Letellier T, Sudoyo H, Ricaut F-X. The Austronesian dispersal ove the Indo-Pacific region. *Abstract to the International Symposium on Austronesian Diaspora*. Bali, Indonesia, Jul 2016. Poster presentation
4. Brucato N, **Kusuma P**, Cox MP, Razafindrazaka H, Pierron D, Letellier T, Sudoyo H, Ricaut FX. À la recherche de l'ancestre Asiatique de Malgache en Indonésie. *Abstract to the 32ème Colloque du Groupement des Anthropologistes de Langue Française.*Toulouse, France, Jun 2016. Oral presentation
5. **Kusuma P**, Brucato N, Cox MP, Pierron D, Grangé P, Leavesley M, Gessain A, Letellier T, Sudoyo H, Ricaut F-X. The Austronesian dispersal ove the Indo-Pacific region.

*Abstract to the 4th Toulouse Economics and Biology Workshop. Evolution: Transmission Mechanisms, Population Structure.*Toulouse, France, May 2016. Poster presentation

6 **Kusuma P**. Malagasy genetic ancestry comes from an historical Malay trading post in Southeast Borneo. *Journée Axe Transversal « Biology-Culture ». AMIS Lab*. Toulouse, France, May 2016. Oral presentation

7. **Kusuma P**, Cox MP, Brucato N, Razafindrazaka H, Pierron D, Letellier T, Sudoyo H, Ricaut F-X. Austronesian westward expansion : Settlement of Madagascar by Indonesian sea nomad groups. *Abstract to 15th International Conference of the European Association of Southeast Asian Archaeologists*. Nanterre, France, Jul 2015. Oral presentation
8. **Kusuma P**, Cox MP, Brucato N, Razafindrazaka H, Pierron D, Letellier T, Sudoyo H, Ricaut F-X. In search of Malagasy's ancestor in Indonesia. *Abstract to 9th ISABS Conference on Forensic and Anthropologic Genetics and Mayo Clinic Lectures in Individualized Medicine*. Bol, Croatia, Jun 2015. Oral presentation
9. **Kusuma P**, Cox MP, Razafindrazaka H, Pierron D, Letellier T, Sudoyo H, Ricaut F-X. Austronesian westward expansion : settlement of Madagascar by Indonesian sea nomad groups. *Abstract to Journées de l'École Doctorale BSB, Université de Toulouse.* Toulouse, France, Apr 2015. Oral presentation
10. **Kusuma P**, Cox MP, Razafindrazaka H, Pierron D, Letellier T, Sudoyo H, Ricaut F-X. Austronesian westward expansion : Ma'anyan Indonesian groups and the settlement of Madagascar. *Abstract to XVII UISPP Congress.* Burgos, Spain, Sep 2014. Oral presentation
11. Sudoyo H*, **Kusuma P***, Suryadi HL. Molecular Anthropological Study of the Dayak Ma'anyan: In Search of the Malagasy Origin. *Abstract to 20th IPPA Congress.* Seam Reap, Cambodia, Jan 2014. Oral presentation
12. **Kusuma, P**. Mitochondrial DNA and modern human migration: cases from population of Indonesia archipelago. *Abstract to 3rd Asian Workshop on Genomics and Community Genetics.* Dehradun (India), Nov 2011. Oral presentation
13. Megawati AAD*, **Kusuma P***, Purnomo GA, Harahap AR, Setianingsih I, Sudoyo H. Does distribution of Southeast Asian Ovalocytosis reflect Austronesian dispersal? : A study of North Maluku and North Sumatran population. *Abstract to 5th International Eijkman Conference*. Jakarta (Indonesia), Nov 2011. Oral presentation
14. **Kusuma P**, Tan MI, Anggraeni T, Sudoyo H. DNA-Based Identification of Flesh Flies Larvae (Diptera: Sarcophagidae) in Jakarta using Cytochrome Oxidase subunit I DNA

Barcode. *Abstract to International Conference on Mathematics and Natural Sciences.* Bandung (Indonesia), Nov 2010. Oral presentation

Workshops

1. EMBO workshop on Population Genomics: background and tools. in Naples, Italy. May 2017. *Participant*
2. Workshop on Population genetics tools. Eijkman Institute. in Jakarta, Indonesia. Nov 2016. *Lecturer*
3. Hands-on workshop on molecular detection of HTLV-1 and HHV-8 viruses in human saliva, serum, and blood samples. Pasteur Institute Paris. in Paris, France. Jun 2016. *Participant*
4. Hands-on workshop on Capture method on whole mitochondrial genome sequencing. in Copenhagen, Denmark. Oct 2014. *Participant*
5. The 3rd Asian Workshop on Genomics and Community Genetics. Pasteur Institute, Paris, and Murdoch University, Australia. in Dehradun, India. Nov 2011. *Participant*