

Episodes of diversification and isolation in Island Southeast Asian and Near Oceanian male lineages

Monika Karmin, Rodrigo Flores, Lauri Saag, Georgi Hudjashov, Nicolas Brucato, Chelzie Crenna-Darusallam, Maximilian Larena, Phillip Endicott, Mattias Jakobsson, J Stephen Lansing, et al.

▶ **To cite this version:**

Monika Karmin, Rodrigo Flores, Lauri Saag, Georgi Hudjashov, Nicolas Brucato, et al.. Episodes of diversification and isolation in Island Southeast Asian and Near Oceanian male lineages. Molecular Biology and Evolution, 2022, 39 (3), pp.msac045,. 10.1093/molbev/msac045 . hal-03753750

HAL Id: hal-03753750
https://cnrs.hal.science/hal-03753750

Submitted on 18 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article
Discoveries

Episodes of diversification and isolation in Island Southeast Asian and Near Oceanian male lineages

Monika Karmin*[1,2], Rodrigo J. Flores[2,3], Lauri Saag[2], Georgi Hudjashov[1,2], Nicolas Brucato[4], Chelzie Crenna-Darusallam[5], Maximilian Larena[6], Phillip L. Endicott[2,7], Mattias Jakobsson[6], J. Stephen Lansing[8], Herawati Sudoyo[5,9,10], Matthew Leavesley[9,10], Mait Metspalu[2], François-Xavier Ricaut[4] and Murray P. Cox*[1]

[1] School of Natural Sciences, Massey University, Palmerston North 4442, New Zealand
[2] Institute of Genomics, University of Tartu, Tartu 51010, Estonia
[3] Institute of Computer Science, University of Tartu, Tartu 51009, Estonia
[4] Laboratoire Evolution et Diversité Biologique (EDB UMR 5174), Université de Toulouse Midi-Pyrénées, CNRS, IRD, UPS, Toulouse,
[5] Genome Diversity and Disease Laboratory, Eijkman Institute for Molecular Biology, Jakarta 10430, Indonesia
[6] Department of Organismal Biology, University of Uppsala, Uppsala 75236, Sweden
[7] Department Hommes Natures Societies, Musée de l'Homme, 75016, Paris, Ile de France, France
[8] Complexity Science Hub Vienna, Vienna 1080, Austria
[9] School of Humanities and Social Sciences, University of Papua New Guinea, National Capital District, Papua New Guinea
[10] CABAH and College of Arts, Society and Education, James Cook University, Cairns, QLD 4870, Australia

***Corresponding authors:** m.p.cox@massey.ac.nz, monika.karmin@ut.ee

Abstract

Island Southeast Asia and Oceania host one of the world's richest assemblages of human phenotypic, linguistic and cultural diversity. Despite this, the region's male genetic lineages are globally among the last to remain unresolved. We compiled ~9.7 Mb of Y chromosome sequence from a diverse sample of over 380 men from this region, including 152 first reported here. The granularity of this dataset allows us to fully resolve and date the regional Y chromosome phylogeny. This new high-resolution tree confirms two main population bursts: multiple rapid diversifications following the region's initial settlement ~50 kya, and extensive expansions <6 kya. Notably, ~40-25 kya the deep rooting local lineages of C-

M130, M-P256, and S-B254 show almost no further branching events in Island Southeast Asia, New Guinea and Australia, matching a similar pause in diversification seen in maternal mitochondrial DNA lineages. The main local lineages start diversifying ~25 kya, at the time of the Last Glacial Maximum. This improved Y chromosome topology highlights localized events with important historical implications, including pre-Holocene contact between Mainland and Island Southeast Asia, potential interactions between Australia and the Papuan world, and a sustained period of diversification following the flooding of the ancient Sunda and Sahul continents as the insular landscape observed today formed. The high-resolution phylogeny of the Y chromosome presented here thus enables a detailed exploration of past isolation, interaction and change in one of the world's least understood regions.

Keywords: Y chromosome, human population genetics, Island Southeast Asia, phylogeography, migration

Introduction

Island Southeast Asia (ISEA) and Near Oceania form a diverse interlocking geographical and cultural region with a population of over 400 million people. Comprising the Indonesia, Philippine and Taiwan archipelagoes, ISEA hosts three of the five largest island states in the world, with Indonesia alone covering an area equivalent to that from Ireland to the Caspian Sea. Oceania is of course even larger, stretching from the eastern edge of ISEA, to New Guinea, and out to the furthest islands of the remote Pacific. Within Oceania, this study focuses on the Near Oceania region, particularly New Guinea and the Bismarck Archipelago, including their potential connections to Australia.

Over time, climatic changes have substantially transformed the physical landscape of this region from islands to continents and back. During the Pleistocene, long periods ~65-15 thousand years ago (kya) saw many of today's islands connected into substantial land masses, with the two largest called Sunda and Sahul (Figure 1) (Voris 2000). Deep sea trenches around Wallacean Islands kept these land masses separate (Hanebuth et al. 2000), but within

them, lakes and river systems likely served as potential pathways for communication and movement, as they still do today. The Sunda and Sahul continents reached their maximum extent when sea levels were lowest during the Last Glacial Maximum (LGM) ~25 kya. Water levels subsequently rose with bursts of flooding taking place ~14 kya, and the land links between modern islands and mainland Asia were largely drowned by ~9 kya (Hanebuth et al. 2000; Voris 2000).

Archaeology shows that modern humans reached Australia ~65-52 kya (O'Connell and Allen 2015; Clarkson et al. 2017), mainland New Guinea ~49-40 kya (Summerhayes et al. 2010; Groube et al. 1986), with the easternmost fringes of the Bismarck Archipelago being settled soon after, ~45-35 kya (Leavesley et al. 2004; Torrence et al. 2004). Much later during the Holocene (<10 kya), the advent and widespread adoption of farming in mainland Asia triggered influential movements of people. This led to the spread of Austronesian languages and Neolithic practices through ISEA and out into Oceania, starting around 6 kya in Taiwan and reaching eastern Indonesia by ~3.5 kya (Bellwood 2006; Xu et al. 2012; Kirch 2017; Deng et al. 2020; Kusuma et al. 2015; Brucato et al. 2016). Movements within the region, in part stimulated by independent instances of plant cultivation (Denham and Haberle 2015), were also extensive (Hudjashov et al. 2017). The result is that Indonesia and Papua New Guinea today host the richest linguistic diversity in the world, with over 700 and 800 languages spoken in those two modern nations. Coupled with this linguistic richness is an astonishing diversity of socio-cultural practices – post marital residence patterns, kinship systems, social norms, and community structures, many of which are only now transitioning away from traditional practices (Godelier 1982; Lansing et al. 2008; Guillot et al. 2013; Tumonggor et al. 2014).

Despite this extraordinary human diversity, knowledge of genetic diversity has lagged substantially behind other global regions. Of particular note is the Y chromosome, whose phylogeny has been resolved elsewhere in the world (Hallast et al. 2015; Karmin et al. 2015; Poznik et al. 2016), but is still only partially known in ISEA and Oceania (Mona et al. 2007; Karafet et al. 2010; Bergström et al. 2016). Despite the trend towards whole genome data (Hudjashov et al. 2017; Jacobs et al. 2019; Brucato et al. 2021; Larena et al. 2021), the lack

of a robust regional Y chromosome phylogeny has limited the identification and molecular dating of sex-specific historical processes. By undertaking extensive new regional sampling and chromosomal-level sequencing, this study presents the first fully resolved regional Y chromosome phylogeny. Here, we fit newly identified lineages into the broader global phylogeny of Y chromosome diversity and characterize patterns that reflect the specific regional history of Island Southeast Asia and Near Oceania.

Results and Discussion

High resolution Y chromosome lineages for one of the last understudied global regions

To develop a fully resolved picture of Y chromosome (chrY) diversity in this region, we extracted chrY sequences from complete human genomes in a set of 152 samples sequenced on the Illumina platform that have not been analyzed before for their chrY diversity. This includes 14 newly reported samples (3 from New Guinea, 7 from Mentawai and 4 from Sumba), 112 men from ISEA (Jacobs et al. 2019) and 26 men from Papua New Guinea (Brucato et al. 2021) (Table S1). Samples were chosen explicitly to fill gaps in earlier geographical coverage and lineage distributions. We then combined this new dataset with previously published chrY sequences from the target region, as well as from broader geographical and phylogenetic contexts. The final global dataset consisted of 795 full chrY sequences extracted from complete human genomes or from targeted sequencing, including samples from neighboring East Asia, Mainland Southeast Asia (MSEA) and other world regions (Figure 1, Table S1).

A fully resolved Y chromosome phylogeny with robust dates

Paternal lineages from ISEA and New Guinea typically fall into the major chrY haplogroups C-M130, M-P256, S-B254 and O-M175, respectively sampled here from 77, 50, 66 and 103 individuals (Figure 1, Table S1). Distinct sub-lineages of C-M130 and O-M175 that occur in ISEA, New Guinea and Australia largely fall within previously described diversity (Kayser et al. 2001; Karafet et al. 2010; Bergström et al. 2016; Nagle et al. 2016; Bergström et al. 2017). In contrast, M-P256 and S-B254 are typically restricted to New Guinea and neighboring islands and have much less well known phylogenies. We identify new phylogenetic structures and sub-haplogroups, and resolve a substantial number of lineages that were previously

classified only as paragroup K*-M526 (Kayser et al. 2006; Bergström et al. 2016; Nagle et al. 2016; Bergström et al. 2017). We also provide additional resolution to the rare lineage NO-M2313, previously classified as K2a1*-M2313 (Poznik et al. 2016). The result is a fully resolved phylogeny of regional Y chromosome diversity, with all individuals uniquely assigned and no paragroups (Figure 1, Figures S1-S3). Sequence data allows dating of diversifications and expansions. Coalescence times were inferred using the Bayesian algorithms of BEAST (Drummond et al. 2012) and dates were calibrated using previously published split times and clock rates (Table S2) (Karmin et al. 2015).

Previous studies have highlighted the critical importance of robust filtering due to the complex repeat structure of the Y chromosome (Karmin et al. 2015; Poznik et al. 2016). We employed a series of filters and sequence masks to limit the dataset to high quality variant calls (detailed in Methods and Table S9), resulting in ~9.7 MB of complete chrY sequence per sample. Across the whole dataset, there is an average of 648 (SD = 18) mutations from tips to the common BT node, with a unimodal and non-skewed distribution (Figure S1). Each chrY lineage was assigned to a haplogroup according to the alternating letter/number rules of the Y Chromosome Consortium (Hammer 2002), taking account of additional modifications proposed subsequently to accommodate the volume of variant data from full chrY sequencing (Karmin et al. 2015). Because our study resolves several previously unresolved paragroups, we define a number of new clades in the global Y chromosome tree, and necessarily re-label some previously known clades while keeping as close as possible to historic nomenclature.

Major radiations around 50 kya reflect first settlement in the region

In concordance with previous findings from global chrY phylogenies (Karmin et al. 2015; Poznik et al. 2016), rapid diversification of lineages C-M130, M-P256 and S-B254 indicate human settlement of ISEA, New Guinea and Australia over a very short time frame close to 50 kya. These major haplogroups arose in the region with few distinguishing mutations, indicating rapid diversification (Figure 2, Figures S2-S4). Even today many of these branches have limited geographic distributions. Indeed, some remain rare single lineages with no further branching events, reflecting long-term isolation and small population size subsequent to this initial expansion event. Examples include S12-Z41926 in Australia, C9-B68, S2-

MY68 and S3-MY69 in Indonesia, and M4-MY62 in New Britain (Figure 2, Figures S2-S3). In addition, rare lineages F-MY1500 (N = 3), P2-B525 (N = 3) and NO2-MY1533 (N = 5) present in MSEA and ISEA show a similar pattern of very early splitting events.

C-M130, M-P256 and S-B254 diversified rapidly and early

A small number of C-M130, M-P256 and S-B254 lineages from the early dispersal event dominate the eastern regions of ISEA and Near Oceania, while some are geographically widespread today. Within our extended dataset, eight C-M130 sub-haplogroups arose between 50-40 kya, of which five are found today in ISEA, New Guinea and/or Australia (Figure 1, Figure S2). With 32 individuals carrying C2-M38, this is the lineage with the widest geographical distribution, reaching from the Wallacean Islands to as far east as Polynesia (Figure 1, Figure S2). Lineage C4-MY1088 has been sequenced in six Australian men and is restricted to Australia since its separation from C2-M38 ~45 kya (95% CI: 41.6-49.6 kya) (Figure S2, File S1) (Bergström et al. 2016). Sequenced in seven and 35 men, C7-B67 and C8-F725 stem from the early dispersal and became distributed more recently across both Sunda and the Wallacean Islands in modern western and eastern Indonesia, respectively (Figure S2).

Before splitting ~49 kya (95% CI: 45.9-51.3 kya), haplogroups M-P256 and S-B254 (Figure 2, Figure S3) shared five common mutations (Table S7). Subsequently, both lineages accumulated only a few haplogroup-defining variants, instead rapidly diversifying further. Before ~40 kya, M-P256 split into three lineages and S-B254 into eight (Figure 2, Figure S3). All lineages that we have sampled from this early diversification further split much later, some even after tens of thousands of years. Today, all sampled M-P256 lineages are present in New Guinea and the Bismarck Archipelago, with two lineages restricted to that area. The most numerous lineage, M1-M106 with 46 individuals, also has the widest geographic distribution, with individuals from the Wallacean Islands of eastern Indonesia and the Torres Strait nested within New Guinean diversity (Figure 3, Figure S3, File S2).

Many of the eight haplogroup S-B254 lineages (Figure 2, inset of Figure 3, Figure S3, File S2, Table S4), which also arose during this same time period of 50-40 kya, are geographically

restricted. Of these, four Australian individuals belong to S7'8-FGC38729 and S12-Z41926, while S1-B273 is found in two Philippine men, but S9-P202 and S10-SK1807 are spread widely across eastern Indonesia and New Guinea with 31 and 16 individuals, respectively (Figure 3, Figure S3). S4'6-PR495.2, present in 16 men, is mostly specific to New Britain and first split around 30 kya (95% CI: 28.0-33.2 kya), well after the time of first settlement of New Britain (Pavlides and Gosden 1994; Leavesley et al. 2004), and today has sub-lineages distributed as far as Australia and Polynesia. Strictly Australian lineages, such as S7'8-FGC38729, diverged around this time, with evidence of only limited later contact (Figure S3). The overall picture ~50-40 kya is therefore one of extensive early population movements across the region associated with rapid lineage radiation, in agreement with archaeological evidence showing that early settlers spread quickly across Sahul within a few thousand years (Bowler 2003; Leavesley et al. 2004; Summerhayes et al. 2010; Clarkson et al. 2017).

A long pause followed initial settlement

Regional chrY lineages show a striking lack of new diversity between 40 and 25 kya (Figure 2, Figures S2 and S3, Tables S3 and S4). Existing lineages in C-M130, M-P256 and S-B254 survived through this gap, but most show no evidence of further radiation during this time. This striking lack of new diversity between 40 and 25 kya is also observed in the regional mitochondrial DNA (mtDNA) lineages (Pedro et al. 2020). ChrY haplogroup S4'6-PR495.2, which is mostly found in New Britain (eastern Papua New Guinea), is the only lineage that split during the hiatus, ~30 kya (95% CI: 28.0-33.2 kya). Noting the limitation that modern sampling may miss lineages that have gone extinct, the observed widespread pause in phylogenetic diversification appears to be characteristic of low population size and persistent population structure during much of the Upper Pleistocene, in agreement with maternal lineage patterns (Pedro et al. 2020), the relatively small numbers of Late Pleistocene archaeological sites identified (Williams 2003; Summerhayes et al. 2017) and the strong territoriality and geographical stability seen in modern ethnographic studies of New Guinea and Indigenous Australian societies (Muke 1993; Sutton 2003). In notable contrast, haplogroup O-M175 on Mainland Asia, which is now the dominating lineage in Eastern Eurasians and comprises more than a quarter of all male lineages in the world (Yan et al. 2014), continued to radiate throughout the Pleistocene. About 33 kya (95% CI: 30.6-36.6

kya), O-M175 began to diversify rapidly, and during the next ten thousand years, six major lineages arose (Figure 2, Figure S4), which are now spread across East Asia and the Asian mainland. Some of these sub-lineages later swept into Island Southeast Asia and out into the Pacific (Kayser et al. 2006; Karafet et al. 2010). Therefore a distinction exists between Mainland Asia, with extensive population expansions and mobility throughout the Upper Pleistocene, and Island Southeast Asia, New Guinea and Australia, with phylogenies reflecting much smaller population sizes and more restricted male mobility, which may have been enhanced both by geography and local cultural practices.

A second bout of diversification 25-15 kya included a split between New Guinean and Australian lineages

After 25 kya, major periods of lineage radiation are observed across ISEA, New Guinea and Australia. These lineages stem from different chrY sub-branches arising from northern Sahul diversity (Figure 3, Figure S3) and they dispersed at different times. The most numerous Australian lineage in this sample, C4-M347 (carried by 6 men), diversified between 25 and 17 kya, coinciding with a major split ~26 kya (95% CI: 23.2-28.9 kya) in a second Australian lineage S7'8-FGC38729 present in three men (Figures S2 and S3). About 23 kya (95% CI: 20.3-25.5 kya), C2-M38 diverged into two lineages, one now found from Wallacea (eastern Indonesia) to New Britain, with the other largely radiating within mainland New Guinea itself (Figure S2). M1-M106 split ~21 kya (95% CI: 19.5-23.9 kya) into M1a'g-Page1 and M1h-MY6 (42 and 4 men, respectively), and is found today in New Guinea and New Britain, but also radiated westward (Figure 3, Figure S3). The most numerous sublineage in New Guinea, S9-P202, split ~19 kya (95% CI: 17.1-20.5 kya), with subsequent expansion ~15-13 kya creating off-shoots of different sublineages across the Wallacean Islands of Alor, Lembata, Sulawesi and Borneo (Figure 3, Figure S3). Over the same timeframe, C7-B67 split ~17 kya (95% CI: 14.6-19.6 kya) and spread between Borneo, Flores and Lembata (Figure S2), and S10-SK1807 gave rise to descendant lineages in New Guinea, New Ireland and Eastern Indonesia around the same time (95% CI: 15.9-19.6 kya) (Figure 3, Figure S3). Lineages with New Guinean origins are mostly found on the Wallacean Islands, with a few in other regions as well.

The prevailing view of regional history is of two major human movements – the initial settlement (Clarkson et al. 2017; O'Connell et al. 2018; Summerhayes et al. 2010; Leavesley et al. 2004) and the arrival of Austronesian speakers 3-3.5 kya (Bellwood 2004; Bellwood 2006), implicitly suggesting that little occurred in this region during the Upper Pleistocene. However, the more resolved chrY tree with denser sampling uncovers a far more complex paternal demographic history, consistent with some of the newer studies in genetics (Gomes et al. 2015; Pedro et al. 2020; Purnomo et al. 2021), linguistics (Schapper 2017) and archaeology (Summerhayes et al. 2007, 2017; Bellwood 2019). In particular, the increase in genetic lineage diversification correlates well with increasing population interactions and population sizes postulated from the archaeological record around 25-20 kya, based on animal, plant and object (e.g., obsidian) translocation between Northern Sahul regions, changes in the subsistence economy, increasing occupation of Highland New Guinea, and settlement of distant islands (e.g., Manus), all of which suggest population size increases and connection of local dynamics to larger regional networks (Leavesley et al. 2004; Specht 2005; Summerhayes et al. 2007, 2017).

We see these connections within a number of haplogroups. In S-B254, the split within the S4'6-PR495.2 lineages of New Britain occurred ~30 kya (95% CI: 28.0-33.2 kya), whereas S5'6-FT17887 diversified first ~23 kya (95% CI: 20.9-25.3 kya) and then ~16 kya (95% CI: 14.4-17.7 kya) (Figure 3, Figure S3). Its sister lineage, S4-P315, appears to have been isolated for some time, only diverging ~18 kya (95% CI: 15.9-20.3 kya) when it split between the New Britain and Queensland (Australia) groups. A much more recent split ~3.6 kya (95% CI: 2.7-4.7 kya) is also seen between men from these two locations (Figure 3, Figure S3). Hence, despite perceptions of long-term isolation between New Guinea and Australia (Bergström et al. 2016; Nagle et al. 2016), joint analyses of chrY from Papua New Guinea (Vernot et al. 2016) and Australia (Bergström et al. 2016) reveals phylogenetic relationships between lineages from New Britain and Australia that are much older than previously thought. This phylogenetic connection is beyond simply the Torres Strait region and is similar to that seen in autosomal data (Brucato et al. 2021). The coalescence date suggests an upper estimate of the Late Pleistocene for possible contact between these regions (Figure 3, Figure S2).

While no archaeological evidence supports direct contact between New Britain and Australia, the Torres Strait Islands (between New Guinea and Australia) may have served as a bridge between these two regions. Indeed trading and intermarriage has been reported between Torres Strait Islanders and Indigenous Australians (Beckett 1987), and some New Guinean lineages (mtDNA Q1, P2 and P3, and chrY M1c6-Z42300) are shared by men with Torres Strait ancestry (Figure S3) (Nagle et al. 2017; Pedro et al. 2020). This genetic affinity between Bismarck Islanders and Indigenous Australians may also result from genetic exchanges pre-dating the divergence between these populations during the postglacial period (18-10 kya) (Figure 3, Figure S3) (Nagle et al. 2017; Pedro et al. 2020; Brucato et al. 2021). More recent contact was also possible during the rapid spread of the Lapita Cultural Complex, which originated in the Bismarck Archipelago 3.3 kya (Denham et al. 2012), spread to the south Papuan Coast 2.9 kya (David et al. 2019), and may have subsequently influenced islands off the Queensland coast by 2.5 kya (Lilley 2019).

Also during this time period (25-15 kya), the sub-lineages of haplogroup O-M175 continued to radiate (Figure 4, Figure S4). Sub-lineages of O3-F525, today widely distributed on Mainland Asia and ISEA split ~25 kya (95% CI: 18.9-28.3 kya), including the formation of the major branch O3a'j-P164, which includes the later expansion of the two most numerous paternal lineages found in Han Chinese (Yan et al. 2014). Between ~18-15 kya, expansions and splits occurred within several lineages in parallel, notably O1-M119, O4-Page59, O6-L127.1, O3a'b-F450, O3f'h-M2055 and O3i'j-N6 (Figure 4, Figure S4, Table S5).

An explosion of activity occurred at the Pleistocene-Holocene boundary 15-10 kya

The sea level began rising ~14 kya with flooding bursts over the next five millennia leading to the current island landscape (Hanebuth et al. 2000; Voris 2000). During this time of major geologic changes, the autosomal genetic data indicates a marked increase of expansion and diversification in the Philippines (Larena et al. 2021), gene flows between islands in the region (Brucato et al. 2021), and the maternal mtDNA data from Sahul and Wallacea shows population expansion (Pedro et al. 2020). Similarly, archaeological data suggest a strong

population increase from the Holocene transition onwards, as part of a longer process that started in the late Pleistocene (25-20 kya) and culminated with the Austronesian dispersal across the Northern Sahul region (Williams 2003; Specht 2005; Bellwood 2006; Donohue and Denham 2010; Summerhayes et al. 2007, 2017). We also see a marked increase in the radiation of chrY lineages, coupled with extensive mobility across the region. Haplogroup S-B254 shows splits in New Britain between S4b-MY103 and S4c-Z41921, with the lineage later shared between men from New Britain and Queensland (Figure 3, Figure S3). Haplogroup M-P256 lineages in New Guinea and New Britain also diversified during this time. Between ~14 and 9 kya, the most numerous and widely spread sub-branch M1a'g-Page1 expanded rapidly within New Guinea (M1b-MY39, M1c-B256, M1e'f-Z43413) and New Britain (M1g-P87), with some lineages spreading west to eastern Indonesia (M1e-MY19, M1a-MY26) (Figure 3, Figure S3). Within haplogroup C-M130, ~13-10 kya, a series of divergences within C2a-M208 in New Guinea gave first an offshoot to Wallacea and then formed the lineages that, much later, spread out into the Pacific (Figure S2) (Cox et al. 2007). Within the few Australian lineages, a split in C4-MY1088 in Queensland occurred ~11 kya (95% CI: 9.2-13.0 kya) (Figure S2).

During this time, haplogroup O-M175 diversification patterns are similar to those of other chrY haplogroups in the region. Between 14-10 kya, major splits occurred in O1a-F31, O2-PK4 and many other O-M175 lineages, some of which are shared between Mainland and Island Southeast Asia. Of particular note, O3f-M7 diversified ~15 kya (95% CI: 11.2-17.4 kya), giving rise to branches found on Mainland Asia, Taiwan and the Philippines, with the latter two lineages diverging from each other ~10 kya (95% CI: 7.6-13.0 kya) within O3f-Y26403. O3f-M7 occurs widely across western Indonesia at low frequency today, and had previously been thought to reflect connections with mainland China during the historic era (Karafet et al. 2010). Although sampling remains low, the current O3f-M7 phylogeny suggests a potentially much earlier upper limit to the movement of people carrying this lineage, either in the Late Pleistocene or starting ~15 kya (Figure 3, Figure S4, Table S5). Older connections between Mainland Asia and western Indonesia were also recently shown by high levels of pre-Neolithic East Asian ancestry in autosomal genetic data of the young female buried ~7.3 kya at Leang Panninge cave in Sulawesi (Carlhoff et al. 2021).

Holocene dispersals carried few external lineages from Mainland East Asia

The sea had risen to near its current level by 8-7 kya, leading to the creation of rich coastal floodplains (Terrell et al. 2011), and rising temperatures gradually led to the higher altitudes of New Guinea being more suitable for permanent human occupation (Summerhays et al. 2017). Indigenous food plant cultivation in the highlands of New Guinea began ~10 kya (Hope and Haberle 2005; Denham and Haberle 2015), coinciding with the expansion seen in M1-Page1 and various S-B254 sub-lineages (Figure 3, Table S4). However, the Holocene story is in many ways dominated by haplogroup O-M175. From 10-2 kya, many expansion events occurred within hg O-M175 (Figure 4, Figure S4), including the rise of the most numerous paternal lineages in the world today (Yan et al. 2014). Within Island Southeast Asia, O3i-B451 and O2a2-F789 stem from this time of major haplogroup O-M175 expansion. However, the lineage with the greatest impact was O1-M119, which arose ~33 kya (95% CI: 24.5-36.9 kya), diverged into sublineages ~17 kya (95% CI: 12.3-19.3 kya), but underwent a striking period of radiation during the mid-Holocene (Figure 4, Figure S4, Table S5).

O1-M119 contains three main sublineages: O1a-F31, found in 35 men from across Mainland Asia, ISEA and in New Guinea; O1b-F819 found in nine men from the Philippines, Sunda and the Wallacean islands; and O1c-M110 found in 16 men from the Philippines, western Indonesia and New Guinea. O1a-F31 radiated rapidly between ~8-4 kya, with a sub-branch in Taiwan and Malaysia separating ~4 kya (95% CI: 2.7-4.9 kya). O1b-F819 lineages in the Philippines and Indonesia split ~9 kya (95% CI: 7.0-11.0 kya), but radiated rapidly ~4 kya (95% CI: 3.1-5.0 kya). Today, O1c-M110 is found only within Island Southeast Asia and New Guinea, where it shows evidence of rapid expansion 6-5 kya (Figure 4, Figure S4, Table S5). We note the limitation that all molecular dates have some level of uncertainty (Tables S3-S5), and not all known O1-M119 diversity is included here (Sun et al. 2021). Nevertheless, the striking radiations observed are characteristic of a diverse population of Mainland Asian origin moving into Island Southeast Asia and expanding rapidly. Due to their timing and current geographical distribution, these genetic lineages are typically associated with Austronesian-speaking peoples and Neolithic practices, which spread over a period of time, starting around 6 kya in Taiwan and reaching eastern Indonesia and New Guinea ~3.5

kya, before subsequently moving out into Oceania (Bellwood 2006; Karafet et al. 2010; Xu et al. 2012; Kirch 2017; Deng et al. 2020).

Holocene expansions in indigenous Y chromosome lineages

While arrivals from Mainland Asia have been the traditional focus of mid-Holocene studies in Island Southeast Asia, we identify a series of simultaneous radiations in lineages with long-standing local connections in C-M130, M-P256 and S-B254. In addition, C8-F725 seeded many new lineages in a short time period between ~6.5-5 kya, including radiating into the Philippines, western Indonesia (Borneo and Mentawai) and the Wallacean islands of eastern Indonesia (Figure S2). A second wave of expansions occurred around ~4-3 kya when lineages in Borneo split from their neighbors in Malaysia or Sulawesi. Local inter-island radiations on Flores and Mentawai also began ~2.5-1.5 kya (Figure S2). In haplogroup M1a'g-Page1, multiple lineages experienced radiations between 7.5-2 kya (downstream of M1c-B256, M1b-MY39, M1a-MY26 and M1g-P87), mostly within New Guinea, Wallacea and the Torres Islands (Figure 3, Figure S3). In haplogroup S-B254, the previously mentioned split within S4c-ZZ41912 also appeared between Queensland (Australia) and New Britain ~3.6 kya (95% CI: 2730-4719 kya) (Figure 3, Figure S3). Many of these expansions and population movements that occurred after ~4 kya were likely triggered by the introduction of agriculture from Austronesian-speaking cultural communities (Deng et al. 2020; Alam et al. 2021), perhaps coupled in some places with the growing influence of local Papuan agricultural practices that had commenced as early as ~10 kya (Denham and Haberle 2015).

Population mobility and lineage expansion has continued to the present

Of course, Y chromosome lineage evolution continues, with local expansions observed across nearby islands within the last two millennia. On Flores, sub-lineages that shared a common ancestor as far back as 49 kya expanded rapidly and locally within the past few thousand years (such as C8-F725, C7-B67 and M1c6-Z42300). Similar patterns of recent expansion are seen in western Indonesia (C8-F725, O3i-B450), eastern Indonesia (M1a3-MY33, O3i-B452) and New Guinea (C2a-M208), including the concomitant expansion of this same C2 lineage into and within Polynesia. Many of the fundamental driving forces observed during the

earlier periods of Island Southeast Asia, New Guinea and Australian history have therefore continued right up to the present.

Conclusions

Phylogenetic relationships of the major paternal genetic lineages of ISEA and Near Oceania are among the last in the world to be resolved. Analyzing ~9.7 Mb of Y chromosome sequence from a geographically and culturally diverse set of men has enabled us to resolve and date the phylogeny from this region. Two well-known population expansions are confirmed, with multiple rapid diversifications between ~50-40 kya reflecting rapid early settlement, and lineages expanding ~6-5 kya indicating extensive movements and interactions. Interestingly, in the period ~40-25 kya, between these major bursts of Y chromosome diversity, we observe almost no branching events, with old lineages persisting but not diversifying. A similar pause in diversification is apparent in maternal mtDNA lineages from the region (Pedro et al. 2020). Starting from the last glacial maximum and intensifying in the subsequent warming period, multiple diversification events follow the flooding of the Sunda and Sahul continents as the insular landscape we see today formed (Voris 2000) and local populations dispersed.

The improved Y chromosome topology also highlights more localized events with important historical implications. O3-M7 lineages from Taiwan and the Philippines diverged from mainland Asian groups ~15 kya, earlier than previously thought (Karafet et al. 2010). This coalescence date provides an upper limit to the movements of people, possibly supporting the growing view from other evidence of early movements between MSEA and ISEA well predating the Neolithic period (Karafet et al. 2010; Trejaut et al. 2014; Vallée, Luciani and Cox 2016). Despite the scarcity of chrY sequence data from Australia (Bergström et al. 2016), we find that some Australian lineages are nested within New Britain diversity, indicating that Australia experienced at least some interaction with the Papuan world. These and other contacts emphasize the potential of less common chrY lineages to illuminate the subtleties of regional contacts. The resolved phylogenetic framework of the Y chromosome

presented here will thus enable future exploration of isolation, interaction and change in one of the world's least understood regions.

Materials and Methods

Human subjects

To develop a fully resolved picture of Y chromosome diversity in this region, we sequenced 14 new male samples (Figure 1, Table S1) on the Illumina platform, and combined these with 26 chrY sequences from Papua New Guinea (Brucato et al. 2021) and 112 chrY sequences from Island Southeast Asia (Jacobs et al. 2019) that were not previously analyzed for their Y chromosome diversity. We then combined this new dataset with all available chrY sequences from the region, supplemented with samples from a broader geographic and phylogenetic context taken from previously published full Y chromosome studies on neighboring East Asia, Mainland Southeast Asia and other world regions (1000 Genomes Project Consortium et al. 2010; Drmanac et al. 2010; PGP 2014; Carmi et al. 2014; Karmin et al. 2015; Bergström et al. 2016; Ilumäe et al. 2016; Malaspinas et al. 2016; Mallick et al. 2016; Poznik et al. 2016; Vernot et al. 2016; Zheng-Bradley et al. 2017; Hudjashov et al. 2018; Bergström et al. 2020). The final dataset comprised 795 samples (Table S1).

Ethics

The donors of all samples provided written informed consent and all experiments were performed in accordance with the relevant guidelines and regulations of collaborating institutions. Collections in Indonesia followed protocols for the protection of human subjects established by the Eijkman Institute Research Ethics Commission (EIREC#90), Nanyang Technological University (IRB-2014-12-011) and the University of Arizona. Permission to conduct research in Indonesia was granted by the State Ministry of Research and Technology (RISTEK). Collections in Papua New Guinea were approved by the Medical Research Advisory Committee of Papua New Guinea (National Department of Health) under research ethics clearance MRAC 16.21 and by the French Ethics Committees (Committees of Protection of Persons CPP 25/21_3, n°SI:21.01.21.42754). Permission to conduct research in

Papua New Guinea was granted by the National Research Institute of Papua New Guinea (permit 99902292358).

Data availability

The complete Y chromosome sequences analyzed here for the first time have been deposited in the European Genome-Phenome Archive under accession numbers EGAS00001006028 and EGAS00001006025.

Sequencing, mapping and genotyping

All newly generated chrY sequences were generated using the Illumina technology (Illumina, San Diego, CA, USA) on the HiSeq instrument (PCR-free protocol) with 30x genome-wide coverage. We employed a series of filters and sequence masks to limit the dataset to high quality variant calls, resulting in 9.7 MB of complete chrY sequence per sample. We used the same processing pipeline for all Illumina data. FASTQ files were mapped with BWA-MEM v0.7.12 (Li and Durbin 2009) to the human reference hs37d5 (http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence). Read duplicates were removed with Picard v2.12.0 (http://broadinstitute.github.io/picard/) and the remaining unique reads were realigned around known indels, followed by base quality score recalibration (BQSR) using GATK v3.8 (Poplin et al. 2018). Variant calling was performed with GATK HaplotypeCaller in haploid mode. All-sites VCF files were filtered with bcftools v1.9 (Li 2011). The Illumina data and previously filtered data from the Complete Genomics technology (Table S1) were merged with GATK CombineVariants v3.8 (Poplin et al. 2018). We extracted the effective overlap between the two datasets by masking out all positions with a 5% or higher proportion of missing genotypes in either the Illumina or Complete Genomics datasets. We additionally excluded regions with poor mappability as described previously (Karmin et al. 2015; Poznik et al. 2016), resulting in a total of 9.7 Mb of analyzed chrY sequence. Variant positions used for phylogenetic reconstruction in each haplogroup are listed in Tables S6-S8.

Phylogeny reconstruction and dating

For the complete dataset of 795 samples, a maximum likelihood (ML) tree was constructed with GTRCAT using 200 rapid bootstrap inferences followed by a thorough ML search executed by RaXML v8.0.0 (Stamatakis 2014) (Figure 1 and S1). All identified variants for shared nodes were annotated to the ML trees and curated manually (Tables S6-S8). We then used the software package BEAST v1.7.5 (Drummond et al. 2012) to simultaneously reconstruct phylogenies and estimate coalescence times. We ran three separate lineage-based analyses for haplogroups C-M130, MS-PR2099 and O-M175. For each haplogroup we used the same parameters: a Bayesian skyline coalescent tree prior, the general time reversible (GTR) substitution model with gamma-distributed rates, a relaxed lognormal clock and the piecewise-constant coalescent model with the number of groups set to 10. For each haplogroup, four parallel runs with different random number seeds were performed. Each run had at least sixty million chains logged every 3000 steps. The results were visualized and checked for effective sample size above 200 using Tracer v1.4. Four parallel runs were combined with LogCombiner and the initial 25% discarded as burn-in. Coalescence time estimates were computed with normally distributed age priors from a previously published phylogeny that had used a mutation rate of 0.74×10^{-9} (95% CI: $0.63\text{-}0.95\times10^{-9}$) per base per year (Karmin et al. 2015). The standard deviation (SD) was set to cover the published confidence intervals of the calibration nodes (Table S2). For haplogroups MS-PR2099 and O-M175, we used an extended set of samples to provide the broader global structure of the phylogeny and the same two samples from haplogroup C-M130 as the outgroup (Figure S2 and S4). The BEAST analyses of MS-PR2099 had 154 individuals and 17621 variants (Figure S3, Table S7), whereas O-M175 had 259 individuals and 19743 variants (Figure S4, Table S8). For both phylogenies, the calibration point for coalescence age estimates was set for the GT-F1329 node with an age of 51557 years (95% confidence interval [CI]: 50206-52917) (Karmin et al. 2015) and a SD of ±1355 in the BEAST analyses. The phylogeny of haplogroup C-M130 was rooted with two samples from J-M304 (Figure S2) and consisted of 131 individuals with 9977 variant positions (Table S6). The calibration point was set to the root of haplogroup C-M130 with an age of 50865 years (95% CI: 49191-52699) (Karmin et al. 2015), with SD set to ±1835 in BEAST. We used Newick Utilities (Junier and Zdobnov 2010) for tree processing, statistical analyses and plotting was conducted in R4.1.2 (R Core

Team, 2014) with the package ggplot2 (Wickham, 2009). All trees were visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Acknowledgements

This work was supported by the French National Research Agency (ANR) (grant PAPUAEVOL n° ANR-20-CE12-0003-01), the French Ministry of Foreign and European Affairs (French Prehistoric Mission in Papua New Guinea), and the French Embassy in Papua New Guinea. We thank Georgia Kaipu (National Research Institute, Papua New Guinea), and Roxanne Tsang, Jason Kariwiga, Teppsey Beni and Kylie Sesuki from the University of Papua New Guinea (Papua New Guinea) for their help during sampling. Data analysis was performed on the High-Performance Computing Center cluster at the University of Tartu, Estonia.

Author Contributions: M.K. and M.P.C designed the project; M.K., R.F., L.S. and G.H. analyzed data; N.B., C.C.-D., S.L., H.S., M.G.L. and F.-X.R. conducted field work, performed community engagement, and provided samples; N.B., M.L., P.E., M.J., M.G.L., M.M. and F.-X.R. contributed interpretations; M.K. and M.P.C wrote the paper.

Competing Interest Statement: The authors declare no conflicts of interest

Supplemental Information

Supplemental Table	S1	Samples in the study
	S2	Calibration points
	S3	Haplogroup C-M130 coalescence ages
	S4	Haplogroup MS-PR2099 coalescence ages
	S5	Haplogroup O-M175 coalescence ages
	S6	Variants in haplogroup C-M130
	S7	Variants in haplogroup MS-PR2099
	S8	Variants in haplogroup O-M175

S9 Method details from mapping to masking

Supplemental Figures S1-S3

Figure

S1. Distribution of the number of mutations from terminal tips to the common BT node on the ML tree of 795 individuals

S2. Dated phylogenetic tree of haplogroup C-M130

S3. Dated phylogenetic tree of haplogroup MS-PR2099

S4. Dated phylogenetic tree of haplogroup O-M175

Supplemental Files S1-S4

Supplemental File S1. Phylogeny of haplogroup C-M130 in FigTree format

S2. Phylogeny of haplogroup MS-PR2099 in FigTree format

S3. Phylogeny of haplogroup O-M175 in FigTree format

S4. Whitelist of chrY regions in bed format

583 **Figure Legends**

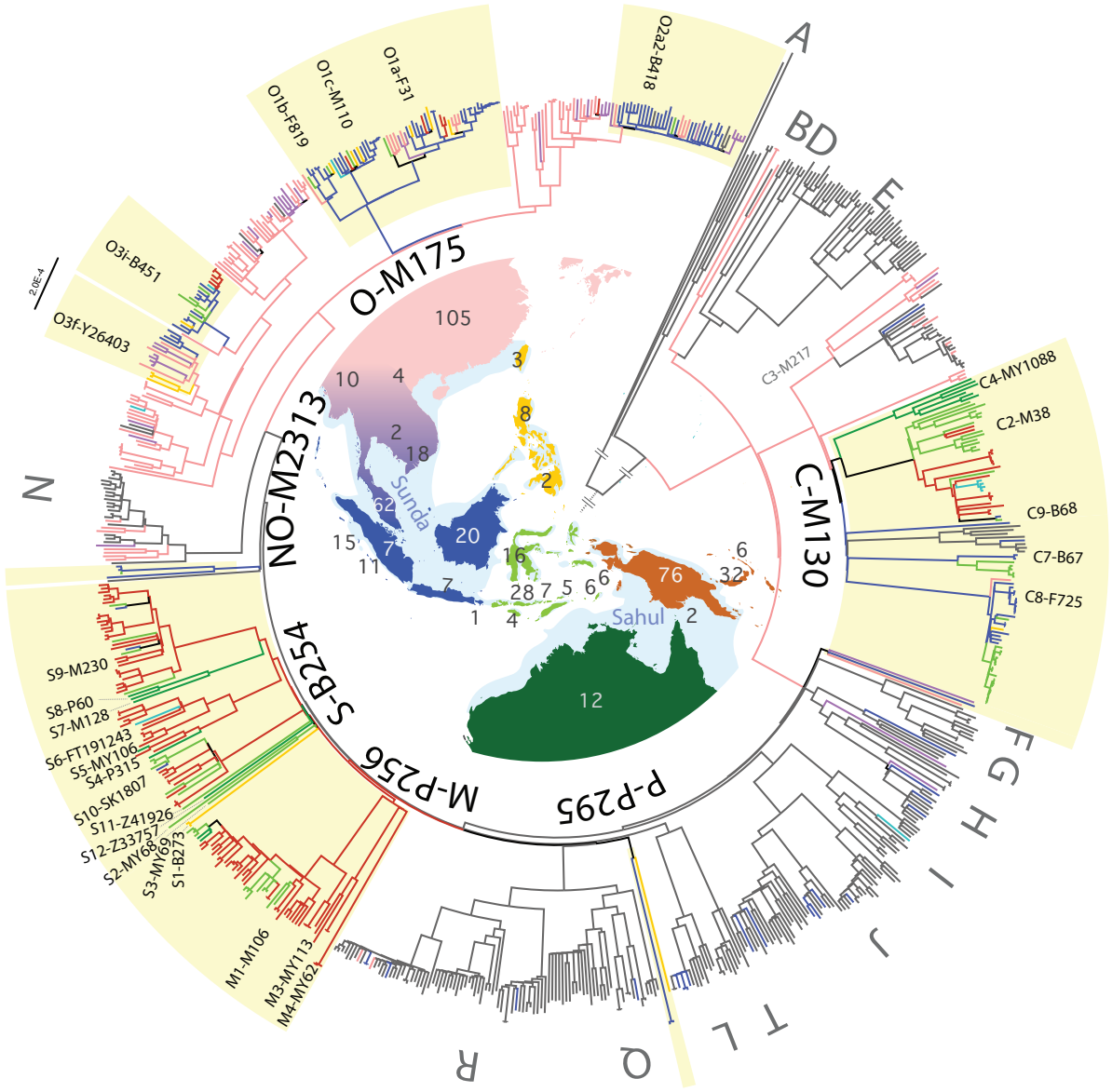
584 **Figure 1. Paternal lineages from Island Southeast Asia and Near Oceania in the context**
585 **of the global Y chromosome phylogeny.** Maximum Likelihood tree of 795 Y chromosome
586 sequences. Colors in the target region correspond to the geographic origin of the samples
587 (numbers shown on the map): dark blue – western Indonesia and the Malay peninsula
588 (Sunda); light green – Eastern Indonesia (Wallacean Islands), yellow – Philippines and
589 Taiwan; brick orange – New Guinea and the Bismarck Archipelago (Northern Sahul); dark
590 green – Australia (Southern Sahul); violet – Mainland Southeast Asia; pink – East Asia;
591 bright blue – Pacific islands; gray – other world regions. The approximate maximum extent
592 of the Sunda and Sahul land masses are shown on the map in light blue. Yellow boxes
593 highlight lineages from the studied region. The root and the deepest splits have been
594 shortened for better fit on the figure.
595

596 **Figure 2. Region-specific lineages of C-M130, M-P256 and S-B254 diversified rapidly**
597 **and early in contrast to later-arriving O-M175.** Dated schematic Y chromosomal
598 phylogenies of haplogroups C-M130 (N=129), F-MY1500 (N=3), M-P256 (N=52), NO2-
599 MY1503 (N=5), O-M175 (N=230), P-B252 (N=3) and S-B254 (N=73). Black – samples
600 from Island Southeast Asia (ISEA) and Near Oceania; gray – Mainland Asia and Eurasia.
601 Lineages with coalescence dates more recent than ~25 kya have been collapsed to emphasize
602 the early branching events and the long pause ~40-30 kya in many of the local lineages
603 within haplogroups M-P256, S-B254 and C-M130, in contrast to the mainland Asian lineages
604 of haplogroups O-M175 and C3-M217. The rare lineages of F-MY1500, P2-B252 and NO2-
605 MY1503 all show early branching events, although their rarity prevents further conclusions
606 about their later diversification history. Triangles are proportional to sample size, and
607 background colors denote time windows as framed in the text. All full or target-captured
608 chrY sequences available to us from ISEA, Oceania and Australia are included. Haplogroups
609 G-M201, P1-M45, N-M231 are represented by only a few lineages; the main structures
610 within C3-M217 and some O-M175 subgroups from other well-sampled global regions are
611 represented, but thinned to a smaller number of individuals.
612

Figure 3. Contacts between Northern Sahul and neighboring areas within the MS-PR2099 lineages. The basic structure of haplogroup MS-PR2099 lineages is shown in the upper inset; circles denote sub-lineages that show evidence of contact between Northern Sahul and other regions, and are illustrated in more detail in the main figure. Colors on the tree and map denote the geographic origin of individuals. Light green boxes show westward paternal gene flow from Northern Sahul mostly to the Wallacean Islands, but also beyond; darker green boxes denote contact between New Britain and Australia. Light brick brown boxes show likely back migration to New Guinea from Wallacea. Light teal represents lineages from Pacific islands. For more details, see Figure S3 and Table S1.

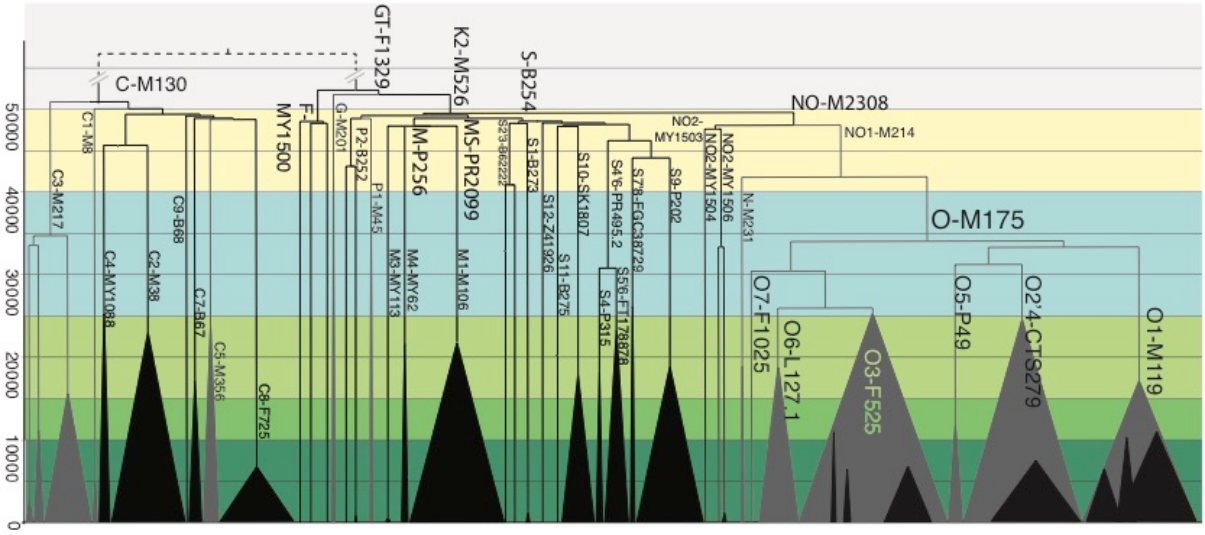
Figure 4. Haplogroup O-M175 sub-lineages in Island Southeast Asia, New Guinea and Oceania. A number of distinct sublineages (black) of O-M175 have penetrated from the Asian mainland into ISEA and beyond, and are likely associated with the expansion of Austronesian speakers. Most of these lineages coalesce ~6-5 kya within subsets of the wider Asian diversity, except for O1b-F819 and O3f-M7 sublineages, which have earlier diversification times (details in Figure S4 and Table S1). In contrast, lineage NO2-MY1506, found today on Sunda and Sulawesi, split very early from its South Asian sister lineage.

632 Figure 1

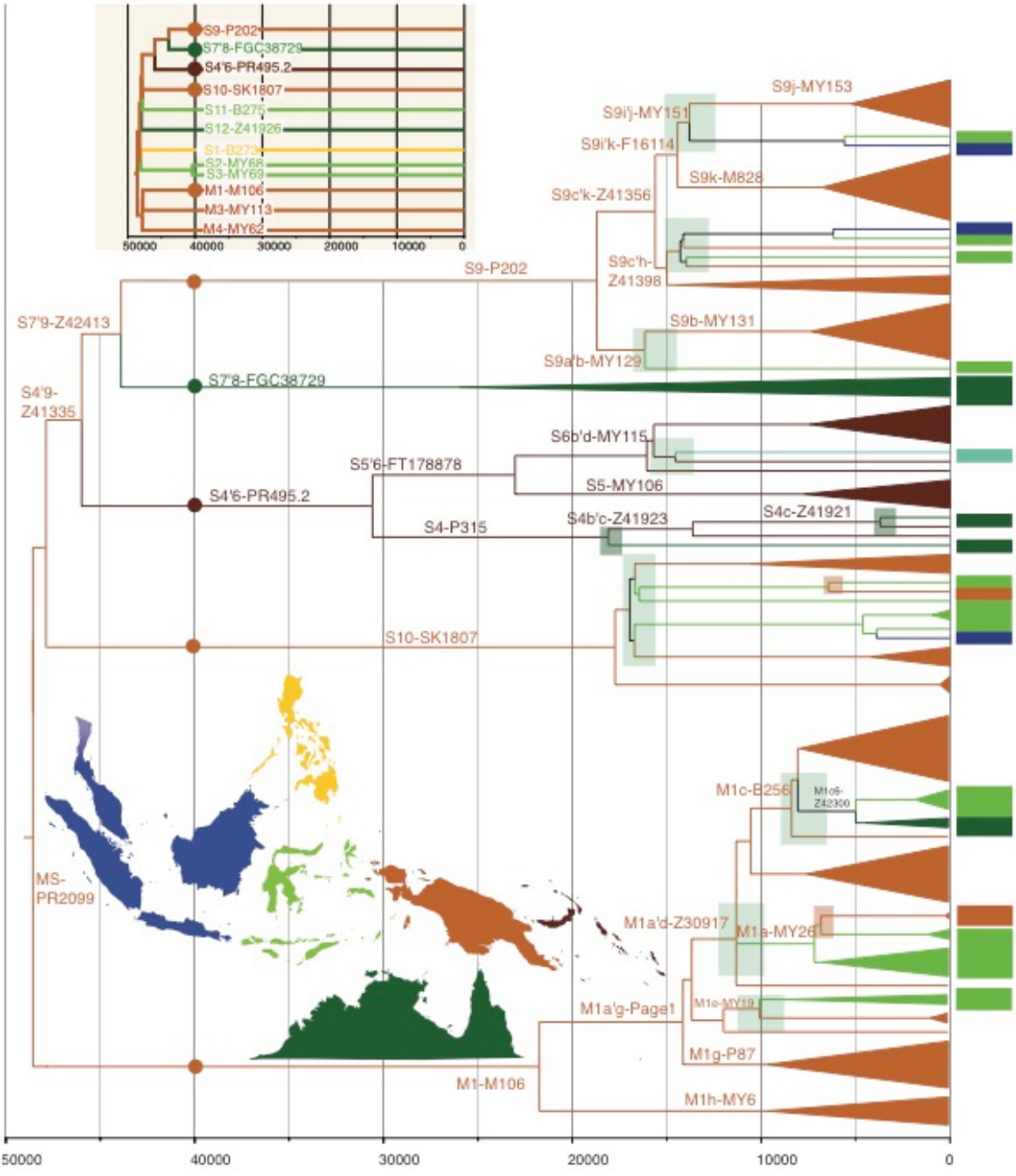


633

Figure 2

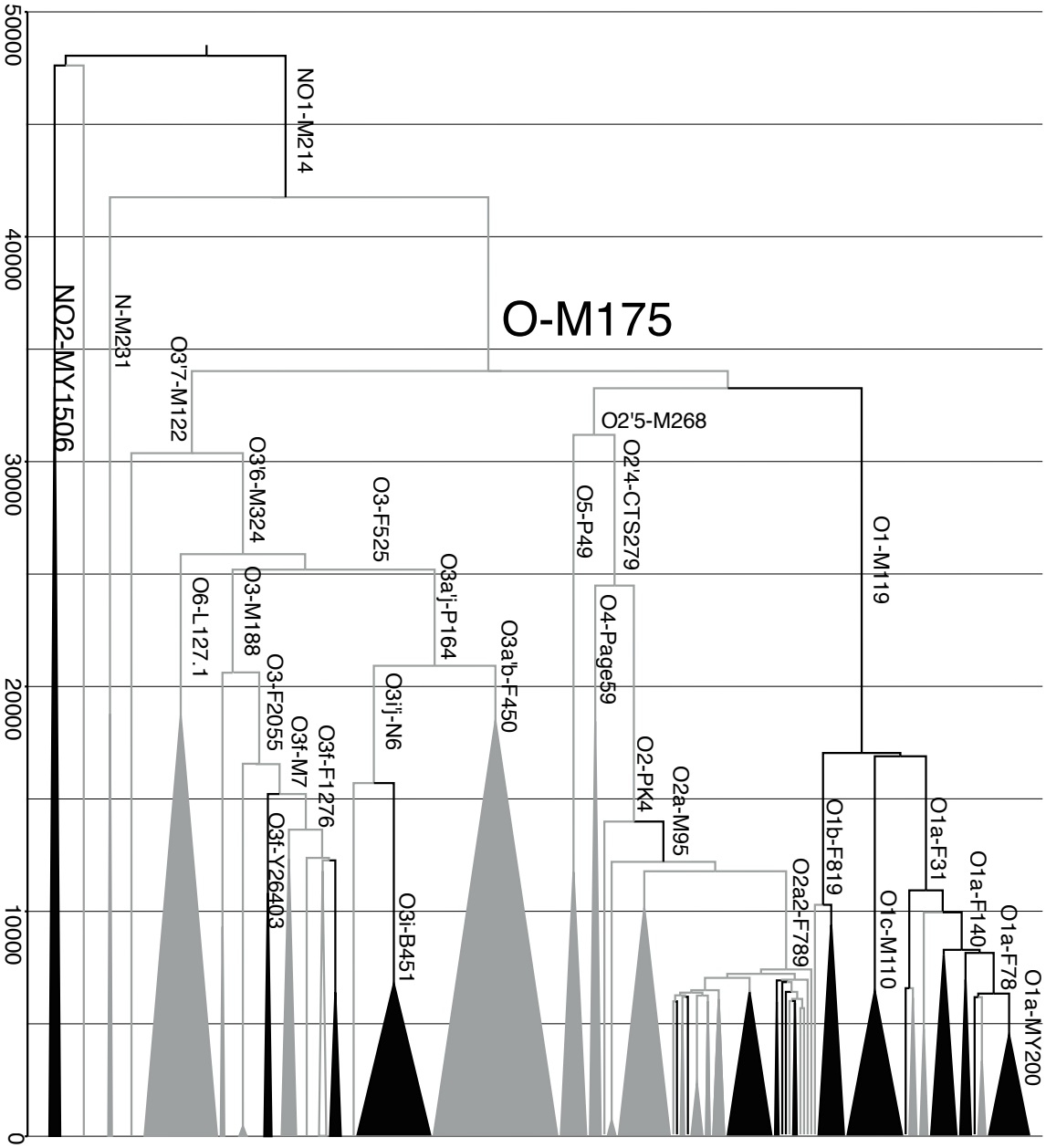

C-M130
GT-F1329
K2-M526
S-B254
NO-M2308
NO1-M214
O-M175
O7-F1025
O6-L127.1
O3-F525
O5-P49
O2'4-CTS279
O1-M119
C1-M8
C3-M217
C4-MY1088
C2-M38
C9-B68
C7-B67
C5-M356
C8-F725
G-M201
P2-B252
P1-M45
M4-MY62
M3-MY113
M-P256
MS-PR2099
M1-M106
S23-B62222
S1-B273
S12-Z41926
S11-B275
S10-SK1807
S4-P315
S56-FT178878
S46-PR495.2
S78-FGC38729
S9-P202
NO2-MY1503
NO2-MY1504
NO2-MY1506
N-M231
50000
40000
30000
20000
10000
0

637 Figure 3
638



639

Figure 4


50000
40000
30000
20000
10000
0
NO1-M214
NO2-MY1506
N-M231
O-M175
O3'7-M122
O3'6-M324
O3-F525
O3a'j-P164
O3a'b-F450
O3i'j-N6
O3i-B451
O3-M188
O3-F2055
O3f-M7
O3f-F1276
O3f-Y26403
O6-L127.1
O2'5-M268
O2'4-CTS279
O5-P49
O4-Page59
O2-PK4
O2a-M95
O2a2-F789
O1-M119
O1b-F819
O1c-M110
O1a-F31
O1a-F140
O1a-F78
O1a-MY200

643 **References**

644 1000 Genomes Project Consortium, Abecasis GR, Altshuler DL, Bentley DR,
645 Chakravarti A, Clark AG, Collins FS, De La Vega FM, Donnelly P, Egholm M, et
646 al. 2010. A map of human genome variation from population-scale sequencing.
647 *Nature* 467:1061–1073.
648 Alam O, Gutaker RM, Wu C-C, Hicks KA, Bocinsky K, Castillo CC, Acabado S, Fuller
649 D, d'Alpoim Guedes JA, Hsing Y-I, et al. 2021. Genome analysis traces regional
650 dispersal of rice in Taiwan and Southeast Asia. *Mol. Biol. Evol.* 38:4832–4846.
651 Beckett, J. 1987. Torres Strait Islanders: Custom and Colonialism. Cambridge
652 University Press: Cambridge.
653 Bellwood P. 2004. First Farmers: The Origins of Agricultural Societies. Wiley: New
654 York.
655 Bellwood P. 2006. Austronesian Prehistory in Southeast Asia: Homeland, Expansion
656 and Transformation. In: Bellwood P, Fox JJ, Tryon D, editors. *The Austronesians:*
657 *Historical and Comparative Perspectives*. Australian National University Press:
658 Canberra. p. 103–118.
659 Bellwood P. 2019. The Indonesian–Australian Archaeological Research Project in the
660 Northern Moluccas. *Terra Australia* 50:1–13.
661 Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, Chen Y, Felkel
662 S, Hallast P, Kamm J, et al. 2020. Insights into human genetic variation and
663 population history from 929 diverse genomes. *Science* 367:eaay5012.
664 Bergström A, Nagle N, Chen Y, McCarthy S, Pollard MO, Ayub Q, Wilcox S, Wilcox
665 L, Van Oorschot RAH, McAllister P, et al. 2016. Deep roots for Aboriginal
666 Australian Y chromosomes. *Curr. Biol.* 26:809–813.
667 Bergström A, Oppenheimer SJ, Mentzer AJ, Auckland K, Robson K, Attenborough R,
668 Alpers MP, Koki G, Pomat W, Siba P, et al. 2017. A Neolithic expansion, but
669 strong genetic structure, in the independent history of New Guinea. *Science*
670 357:1160–1163.
671 Bowler JM, Johnston H, Olley JM, Prescott JR et al. 2003. New ages for human
672 occupation and climatic change at Lake Mungo, Australia. *Nature* 421: 837–840.

673 Brucato N, Kusuma P, Cox MP, Pierron D, Purnomo GA, et al. 2016. Malagasy genetic
674 ancestry comes from an historical Malay trading post in southeast Borneo. *Mol.*
675 *Biol. Evol.* 33:2396-2400.
676 Brucato N, André M, Tsang R, Saag L, Kariwiga J, Sesuki K, Beni T, Pomat W, Muke
677 J, Meyer V, et al. 2021. Papua New Guinean genomes reveal the complex
678 settlement of North Sahul. *Mol. Biol. Evol.* 38:5107–5121.
679 Carlhoff S, Duli A, Nägele K, Nur M, Skov L, Sumantri I, Oktaviana AA, Hakim B,
680 Burhan B, Syahdar FA, et al. 2021. Genome of a middle Holocene hunter-gatherer
681 from Wallacea. *Nature* 596:543–547.
682 Carmi S, Hui KY, Kochav E, Liu X, Xue J, Grady F, Guha S, Upadhyay K, Ben-
683 Avraham D, Mukherjee S, et al. 2014. Sequencing an Ashkenazi reference panel
684 supports population-targeted personal genomics and illuminates Jewish and
685 European origins. *Nat. Commun.* 5:4835.
686 Clarkson C, Jacobs Z, Marwick B, Fullagar R, Wallis L, Smith M, Roberts RG, Hayes
687 E, Lowe K, Carah X, et al. 2017. Human occupation of northern Australia by
688 65,000 years ago. *Nature* 547:306–310.
689 Cox MP, Redd AJ, Karafet TM, Ponder CA, Lansing JS, Sudoyo H et al. 2007. A
690 Polynesian motif on the Y chromosome: Population structure in remote Oceania.
691 *Human Biology* 79:525-535.
692 David BO, Aplin K, Peck H, Skelly R, Leavesley M, Mialanes J, Szabo K, Koppel B,
693 Petchey F, Richards T, Ulm S, McNiven I, Rowe C, Aird SJ, Faulkner P, Ford, A.
694 2019. Moiapu 3: Settlement on Moiapu Hill at the very end of Lapita, Caution Bay
695 hinterland. In S. Bedford & M. Spriggs (eds.), *Debating Lapita: Distribution,*
696 *Chronology, Society and Subsistence*. Australian National University Press:
697 Canberra. pp 61–88.
698 Deng Z, Hung H, Carson MT, Oktaviana AA, Hakim B, Simanjuntak T. 2020.
699 Validating earliest rice farming in the Indonesian Archipelago. *Sci. Reports* 10:1–9.
700 Denham T, Bronk Ramsey C, Specht J. 2012. Dating the appearance of Lapita pottery in
701 the Bismarck Archipelago and its dispersal to Remote Oceania. *Archaeology in*
702 *Oceania* 47:39–46.

703 Denham T, Haberle S. 2015. Agricultural emergence and transformation in the Upper
704 Wahgi Valley, Papua New Guinea during the Holocene: Theory, method and
705 practice. *Holocene* 18:481–496.
706 Donohue M, Tim Denham. 2010. Farming and language in island Southeast Asia:
707 Reframing Austronesian history. *Curr Anthropol.* 51:223.
708 Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P,
709 Nazarenko I, Nilsen GB, Yeung G, et al. 2010. Human genome sequencing using
710 unchained base reads on self-assembling DNA nanoarrays. *Science* 327:78–81.
711 Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with
712 BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29:1969–1973.
713 Godelier M. 1982. La production des grands hommes: pouvoir et domination masculine
714 chez les Baruya de Nouvelle-Guinée. Paris: Fayard.
715 Gomes SM, Bodner M, Souto L, Zimmermann B, Huber G, Strobl C, Röck AW, Achilli
716 A, Olivieri A, Torroni A, et al. 2015. Human settlement history between Sunda and
717 Sahul: A focus on East Timor (Timor-Leste) and the Pleistocenic mtDNA diversity.
718 *BMC Genomics* 16:70.
719 Guillot EG, Tumonggor MK, Lansing JS, Sudoyo H, Cox MP. 2013. Climate change
720 influenced female population sizes through time across the Indonesian Archipelago.
721 *Hum. Biol.* 85:135–152.
722 Groube, L., 1986. Waisted axes of Asia, Melanesia and Australia. In: Ward, G. (ed.),
723 *Archaeology in ANZAAS*. Canberra Archaeological Society: Canberra. pp. 168-177.
724 Hallast P, Batini C, Zadik D, Delser PM, Jon H, Arroyo-pardo E, Cavalleri GL, Knijff P
725 De, Bisol D, Dupuy BM, et al. 2015. The Y-chromosome tree bursts into leaf:
726 13,000 high-confidence SNPs covering the majority of known clades. *Mol. Biol.*
727 *Evol.* 32:661–673.
728 Hammer M. 2002. A nomenclature system for the tree of human Y-chromosomal binary
729 haplogroups. *Genome Res.* 12:339–348.
730 Hanebuth T, Stattegger K, Grootes PM. 2000. Rapid flooding of the Sunda Shelf: A late-
731 glacial sea-level record. *Science.* 288:1033–1035.
732 Hope GS, Haberle SG. 2005. The history of the human landscapes of New Guinea. In:
733 Pawley Andrew, Attenborough R, Golson J, Hide R, editors. *Papuan Pasts:*

734 *Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples*. Pacific
735 Linguistics: Canberra. pp 541–554.
736 Hudjashov G, Endicott P, Post H, Nagle N, Ho SYW, Lawson DJ, Reidla M, Karmin M,
737 Rootsi S, Metspalu E, et al. 2018. Investigating the origins of eastern Polynesians
738 using genome-wide data from the Leeward Society Isles. *Sci. Rep.* 8:1823.
739 Hudjashov G, Karafet TM, Lawson DJ, Downey S, Savina O, Sudoyo H, Lansing JS,
740 Hammer MF, Cox MP. 2017. Complex patterns of admixture across the Indonesian
741 archipelago. *Mol. Biol. Evol.* 34:2439–2452.
742 Ilumäe A-M, Reidla M, Chukhryaeva M, Järve M, Post H, Karmin M, Saag L,
743 Agdzhoyan A, Kushniarevich A, Litvinov S, et al. 2016. Human Y chromosome
744 haplogroup N: A non-trivial time-resolved phylogeography that cuts across
745 language families. *Am. J. Hum. Genet.* 99:163–173.
746 Jacobs GS, Hudjashov G, Saag L, Kusuma P, Darusallam CC, Lawson DJ, Mondal M,
747 Pagani L, Ricaut FX, Stoneking M, et al. 2019. Multiple deeply divergent
748 Denisovan ancestries in Papuans. *Cell* 177:1010-1021.
749 Junier T, Zdobnov EM. 2010. The Newick Utilities: High-throughput Phylogenetic tree
750 Processing in the UNIX Shell. Bioinformatics,26:1669-70.
751 Karafet TM, Hallmark B, Cox MP, Sudoyo H, Downey S, Lansing JS, Hammer MF.
752 2010. Major east-west division underlies Y chromosome stratification across
753 Indonesia. *Mol. Biol. Evol.* 27:1833–1844.
754 Karmin M, Saag L, Vicente M, Wilson Sayres MA, Järve M, Talas UG, Rootsi S,
755 Ilumäe A-M, Mägi R, Mitt M, et al. 2015. A recent bottleneck of Y chromosome
756 diversity coincides with a global change in culture. *Genome Res.* 25:459–466.
757 Kayser M, Brauer S, Cordaux R, Casto A, Lao O, Zhivotovsky LA, Moyse-Faurie C,
758 Rutledge RB, Schiefenhövel W, Gil D, et al. 2006. Melanesian and Asian origins of
759 Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Mol. Biol.*
760 *Evol.* 23:2234–2244.
761 Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill PA, Stoneking M. 2001.
762 Independent histories of human Y chromosomes from Melanesia and Australia.
763 *Am. J. Hum. Genet.* 68:173–190.

764 Kirch PV. 2017. *On the Road of the Winds: An Archaeological History of the Pacific*
765 *Islands before European Contact.* University of California Press: Los Angeles.
766 Kusuma P, Cox MP, Pierron D, Razafindrazaka H, Brucato N, Tonasso L et al. 2015.
767 Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar
768 by Indonesian sea nomad populations. *BMC Genomics* 16:191.
769 Lansing JS, Watkins JC, Hallmark B, Cox MP, Karafet TM, Sudoyo H, Hammer MF.
770 2008. Male dominance rarely skews the frequency distribution of Y chromosome
771 haplotypes in human populations. *Proc. Natl. Acad. Sci. USA* 105:11645–11650.
772 Larena M, Sanchez-Quinto F, Sjödin P, McKenna J, Ebeo C, Reyes R, Casel O, Huang
773 J-YY, Hagada KP, Guilay D, et al. 2021. Multiple migrations to the Philippines
774 during the last 50,000 years. *Proc. Natl. Acad. Sci. USA* 118:e2026132118 .
775 Leavesley M., Chappell J. 2004. Buang Merabak: additional early radiocarbon evidence
776 of the colonisation of the Bismarck Archipelago, Papua New Guinea. *Antiquity*
777 78:301.
778 Li H. 2011. A statistical framework for SNP calling, mutation discovery, association
779 mapping and population genetical parameter estimation from sequencing data.
780 *Bioinformatics* 27:2987–2993.
781 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
782 transform. *Bioinformatics* 25:1754–1760.
783 Lilley I. 2019. Lapita: The Australian Connection. In S. Bedford & M. Spriggs (eds.),
784 *Debating Lapita: Distribution, Chronology, Society and Subsistence*. Australian
785 National University Press: Canberra. pp 105–114.
786 Malaspinas A-S, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, Bergström A,
787 Athanasiadis G, Cheng JY, Crawford JE, et al. 2016. A genomic history of
788 Aboriginal Australia. *Nature* 538:207–214.
789 Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N,
790 Nordenfelt S, Tandon A, et al. 2016. The Simons Genome Diversity Project: 300
791 genomes from 142 diverse populations. *Nature* 538:201–206.
792 Mona S, Tommaseo-Ponzetta M, Brauer S, Sudoyo H, Marzuki S, Kayser M. 2007.
793 Patterns of Y-chromosome diversity intersect with the trans-New Guinea
794 hypothesis. *Mol. Biol. Evol.* 24:2546–2555.

795 Muke J. 1993. The Wahgi opo kumbo: an account of warfare in the Central Highlands of
796 New Guinea. PhD thesis, University of Cambridge, UK.
797 Nagle N, Ballantyne KN, Oven M van, Tyler-Smith C, Xue Y, Taylor D, Wilcox S,
798 Wilcox L, Turkalov R, van Oorschot RAH et al. 2016. Antiquity and diversity of
799 aboriginal Australian Y-chromosomes. *Am. J. Phys. Anthropol.* 159:367–381.
800 O'Connell JF, Allen J. 2015. The process, biotic impact, and global implications of the
801 human colonization of Sahul about 47,000 years ago. *J. Archaeol. Sci.* 56:73–84.
802 O'Connell JF, Allen J, Williams MAJ, Williams AN, Turney CSM, Spooner NA,
803 Kamminga J, Brown G, Cooper A. 2018. When did *Homo sapiens* first reach
804 Southeast Asia and Sahul? *Proc. Natl. Acad. Sci. USA* 115:8482–8490.
805 Pavlides C, Gosden C. 1994. 35,000-year-old sites in the rainforests of West New
806 Britain, Papua New Guinea. *Antiquity* 68:604–610.
807 Pedro N, Brucato N, Fernandes V, André M, Saag L, Pomat W, Besse C, Boland A,
808 Deleuze JF, Clarkson C, et al. 2020. Papuan mitochondrial genomes and the
809 settlement of Sahul. *J. Hum. Genet.* 65:875–887.
810 Harvard Personal Genome Project. 2014. http://www. personalgenomes.org
811 Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Auwera GA Van
812 der, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. 2018. Scaling
813 accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 201178.
814 Available from: https://www.biorxiv.org/content/10.1101/201178v3
815 Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, Wilson Sayres MA, Ayub Q,
816 McCarthy SA, Narechania A, Kashin S, et al. 2016. Punctuated bursts in human
817 male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat.*
818 *Genet.* 48:593–599.
819 Purnomo GA, Mitchell KJ, O'Connor S, Kealy S, Taufik L, Schiller S, Rohrlach A,
820 Cooper A, Llamas B, Sudoyo H, et al. 2021. Mitogenomes reveal two major
821 influxes of Papuan ancestry across Wallacea following the Last Glacial Maximum
822 and Austronesian contact. *Genes* 12:965.
823 R Core Team. 2014. R: A language and environment for statistical computing. R
824 Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-
825 project.org/

826 Schapper A. 2017. Chapter 7. Farming and the Trans-New Guinea family. In *Language*
827 *Dispersal Beyond Farming*. John Benjamins Publishing Company: New York. pp
828 155–181.
829 Specht J. 2005. Revisiting the Bismarcks: some alternative views. In: Pawley Andrew,
830 Attenborough R, Golson J, Hide R, editors. *Papuan Pasts: Cultural, Linguistic and*
831 *Biological Histories of Papuan-Speaking Peoples*. Canberra: Pacific Linguistics. pp
832 235–288.
833 Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-
834 analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
835 Summerhayes, GR. 2007. Island Melanesian Pasts: A view from archaeology. In
836 Friedlaender JS (ed). *Genes, Language, and Culture History in the Southwest*
837 *Pacific*. Oxford University Press: New York. pp 10-35.
838 Summerhayes GR, Leavesley, M, Fairbairn, A, Mandui, H, Field, J, Ford, A, Fullagar,
839 R. 2010. Human adaptation and use of plants in highland New Guinea 49,000-
840 44,000 years ago. *Science* 330:78-81.
841 Summerhayes GR, Field JH, Shaw B, Gaffney D. 2017. The archaeology of forest
842 exploitation and change in the tropics during the Pleistocene: the case of Northern
843 Sahul (Pleistocene New Guinea). *Quat Int.* 448:14-30.
844 Sun J, Li YX, Ma PC, Yan S, Cheng HZ, Fan ZQ, Deng XH, Ru K, Wang CC, Chen G,
845 et al. 2021. Shared paternal ancestry of Han, Tai-Kadai-speaking, and
846 Austronesian-speaking populations as revealed by the high resolution phylogeny of
847 O1a-M119 and distribution of its sub-lineages within China. *Am. J. Phys.*
848 *Anthropol.* 174:686–700.
849 Sutton, P. 2003. Native title in Australia: An ethnographic perspective. Cambridge
850 University Press: Cambridge.
851 Terrell JE, Pope KO, Goff JR. 2011. Chapter 3: Context and relevance. *Fieldiana*
852 *Anthropol.* 42:21–28.
853 Torrence R, Neall V, Doelman T, Rhodes E, McKee C, Davies H, Bonetti R,
854 Guglielmetti A, Manzoni A, Oddone M, et al. 2004. Pleistocene colonisation of the
855 Bismarck Archipelago: New evidence from West New Britain. *Archaeol. Ocean.*
856 39:101–130.

857 Trejaut JA, Poloni ES, Yen J-C, Lai Y-H, Loo J-H, Lee C-L, He C-L, Lin M. 2014.
858 Taiwan Y-chromosomal DNA variation and its relationship with Island Southeast
859 Asia. *BMC Genet.* 15:1–23.
860 Tumonggor MK, Karafet TM, Downey S, Stephen Lansing J, Norquest P, Sudoyo H,
861 Hammer MF, Cox MP. 2014. Isolation, contact and social behavior shaped genetic
862 diversity in West Timor. *J. Hum. Genet.* 59:494–503.
863 Vallée F, Luciani A, Cox MP. 2016. Reconstructing demography and social behavior
864 during the Neolithic expansion from genomic diversity across Island Southeast
865 Asia. *Genetics* 204:1495–1506.
866 Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM, Dannemann M,
867 Grote S, McCoy RC, Norton H et al. 2016. Excavating Neandertal and Denisovan
868 DNA from the genomes of Melanesian individuals. *Science* 352:235-239.
869 Voris HK. 2000. Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river
870 systems and time durations. *J. Biogeogr.* 27:1153–1167.
871 Wickham H. 2009. ggplot2: elegant graphics for data analysis. Springer New York.
872 Williams, A, 2013. A new population curve for prehistoric Australia. *Proceedings of the*
873 *Royal Society Series B* 280:20130486.
874 Xu S, Pugach I, Stoneking M, Kayser M, Jin L, Consortium THP-AS. 2012. Genetic
875 dating indicates that the Asian–Papuan admixture through Eastern Indonesia
876 corresponds to the Austronesian expansion. *Proc. Natl. Acad. Sci. USA* 109:4574-
877 4579.
878 Yan S, Wang C-CC, Zheng H-XX, Wang W, Qin Z-DD, Wei L-HH, Wang Y, Pan X-
879 DD, Fu W-QQ, He Y-GG et al. 2014. Y chromosomes of 40% Chinese descend
880 from three Neolithic super-grandfathers. *PLoS One* 9:e105691.
881 Zheng-Bradley X, Streeter I, Fairley S, Richardson D, Clarke L, Consortium 1000
882 Genomes Project. 2017. Alignment of 1000 Genomes Project reads to reference
883 assembly GRCh38tle. *Gigascience* 6:1–8.
884