



**HAL**  
open science

## **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.

1  
2  
3  
4  
5  
6  
7 **Episodes of diversification and isolation in Island Southeast Asian and Near Oceanian**  
8 **male lineages**  
9

10 Monika Karmin<sup>\*1,2</sup>, Rodrigo J. Flores<sup>2,3</sup>, Lauri Saag<sup>2</sup>, Georgi Hudjashov<sup>1,2</sup>, Nicolas Brucato<sup>4</sup>,  
11 Chelzie Crenna-Darusallam<sup>5</sup>, Maximilian Larena<sup>6</sup>, Phillip L. Endicott<sup>2,7</sup>, Mattias Jakobsson<sup>6</sup>,  
12 J. Stephen Lansing<sup>8</sup>, Herawati Sudoyo<sup>5,9,10</sup>, Matthew Leavesley<sup>9,10</sup>, Mait Metspalu<sup>2</sup>, François-  
13 Xavier Ricaut<sup>4</sup> and Murray P. Cox<sup>\*1</sup>  
14

15 <sup>1</sup> School of Natural Sciences, Massey University, Palmerston North 4442, New Zealand

16 <sup>2</sup> Institute of Genomics, University of Tartu, Tartu 51010, Estonia

17 <sup>3</sup> Institute of Computer Science, University of Tartu, Tartu 51009, Estonia

18 <sup>4</sup> Laboratoire Evolution et Diversité Biologique (EDB UMR 5174), Université de Toulouse  
19 Midi-Pyrénées, CNRS, IRD, UPS, Toulouse,

20 <sup>5</sup> Genome Diversity and Disease Laboratory, Eijkman Institute for Molecular Biology,  
21 Jakarta 10430, Indonesia

22 <sup>6</sup> Department of Organismal Biology, University of Uppsala, Uppsala 75236, Sweden

23 <sup>7</sup> Department Hommes Natures Societies, Musée de l'Homme, 75016, Paris, Ile de France,  
24 France

25 <sup>8</sup> Complexity Science Hub Vienna, Vienna 1080, Austria

26 <sup>9</sup> School of Humanities and Social Sciences, University of Papua New Guinea, National  
27 Capital District, Papua New Guinea

28 <sup>10</sup> CABAH and College of Arts, Society and Education, James Cook University, Cairns, QLD  
29 4870, Australia  
30

31 **\*Corresponding authors:** m.p.cox@massey.ac.nz, monika.karmin@ut.ee  
32  
33  
34  
35

36 **Abstract**

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

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

55

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

58

59

## 60 **Introduction**

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

69

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

75 them, lakes and river systems likely served as potential pathways for communication and  
76 movement, as they still do today. The Sunda and Sahul continents reached their maximum  
77 extent when sea levels were lowest during the Last Glacial Maximum (LGM) ~25 kya. Water  
78 levels subsequently rose with bursts of flooding taking place ~14 kya, and the land links  
79 between modern islands and mainland Asia were largely drowned by ~9 kya (Hanebuth et al.  
80 2000; Voris 2000).

81

82 Archaeology shows that modern humans reached Australia ~65-52 kya (O'Connell and Allen  
83 2015; Clarkson et al. 2017), mainland New Guinea ~49-40 kya (Summerhayes et al. 2010;  
84 Groube et al. 1986), with the easternmost fringes of the Bismarck Archipelago being settled  
85 soon after, ~45-35 kya (Leavesley et al. 2004; Torrence et al. 2004). Much later during the  
86 Holocene (<10 kya), the advent and widespread adoption of farming in mainland Asia  
87 triggered influential movements of people. This led to the spread of Austronesian languages  
88 and Neolithic practices through ISEA and out into Oceania, starting around 6 kya in Taiwan  
89 and reaching eastern Indonesia by ~3.5 kya (Bellwood 2006; Xu et al. 2012; Kirch 2017;  
90 Deng et al. 2020; Kusuma et al. 2015; Brucato et al. 2016). Movements within the region, in  
91 part stimulated by independent instances of plant cultivation (Denham and Haberle 2015),  
92 were also extensive (Hudjashov et al. 2017). The result is that Indonesia and Papua New  
93 Guinea today host the richest linguistic diversity in the world, with over 700 and 800  
94 languages spoken in those two modern nations. Coupled with this linguistic richness is an  
95 astonishing diversity of socio-cultural practices – post marital residence patterns, kinship  
96 systems, social norms, and community structures, many of which are only now transitioning  
97 away from traditional practices (Godelier 1982; Lansing et al. 2008; Guillot et al. 2013;  
98 Tumonggor et al. 2014).

99

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

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

112

## 113 **Results and Discussion**

114

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

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

128

### 129 **A fully resolved Y chromosome phylogeny with robust dates**

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

138 classified only as paragroup K\*-M526 (Kayser et al. 2006; Bergström et al. 2016; Nagle et al.  
139 2016; Bergström et al. 2017). We also provide additional resolution to the rare lineage NO-  
140 M2313, previously classified as K2a1\*-M2313 (Poznik et al. 2016). The result is a fully  
141 resolved phylogeny of regional Y chromosome diversity, with all individuals uniquely  
142 assigned and no paragroups (Figure 1, Figures S1-S3). Sequence data allows dating of  
143 diversifications and expansions. Coalescence times were inferred using the Bayesian  
144 algorithms of BEAST (Drummond et al. 2012) and dates were calibrated using previously  
145 published split times and clock rates (Table S2) (Karmin et al. 2015).

146

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

159

### 160 **Major radiations around 50 kya reflect first settlement in the region**

161 In concordance with previous findings from global chrY phylogenies (Karmin et al. 2015;  
162 Poznik et al. 2016), rapid diversification of lineages C-M130, M-P256 and S-B254 indicate  
163 human settlement of ISEA, New Guinea and Australia over a very short time frame close to  
164 50 kya. These major haplogroups arose in the region with few distinguishing mutations,  
165 indicating rapid diversification (Figure 2, Figures S2-S4). Even today many of these branches  
166 have limited geographic distributions. Indeed, some remain rare single lineages with no  
167 further branching events, reflecting long-term isolation and small population size subsequent  
168 to this initial expansion event. Examples include S12-Z41926 in Australia, C9-B68, S2-

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

172

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

174 A small number of C-M130, M-P256 and S-B254 lineages from the early dispersal event  
175 dominate the eastern regions of ISEA and Near Oceania, while some are geographically  
176 widespread today. Within our extended dataset, eight C-M130 sub-haplogroups arose  
177 between 50-40 kya, of which five are found today in ISEA, New Guinea and/or Australia  
178 (Figure 1, Figure S2). With 32 individuals carrying C2-M38, this is the lineage with the  
179 widest geographical distribution, reaching from the Wallacean Islands to as far east as  
180 Polynesia (Figure 1, Figure S2). Lineage C4-MY1088 has been sequenced in six Australian  
181 men and is restricted to Australia since its separation from C2-M38 ~45 kya (95% CI: 41.6-  
182 49.6 kya) (Figure S2, File S1) (Bergström et al. 2016). Sequenced in seven and 35 men, C7-  
183 B67 and C8-F725 stem from the early dispersal and became distributed more recently across  
184 both Sunda and the Wallacean Islands in modern western and eastern Indonesia, respectively  
185 (Figure S2).

186

187 Before splitting ~49 kya (95% CI: 45.9-51.3 kya), haplogroups M-P256 and S-B254 (Figure  
188 2, Figure S3) shared five common mutations (Table S7). Subsequently, both lineages  
189 accumulated only a few haplogroup-defining variants, instead rapidly diversifying further.  
190 Before ~40 kya, M-P256 split into three lineages and S-B254 into eight (Figure 2, Figure S3).  
191 All lineages that we have sampled from this early diversification further split much later,  
192 some even after tens of thousands of years. Today, all sampled M-P256 lineages are present  
193 in New Guinea and the Bismarck Archipelago, with two lineages restricted to that area. The  
194 most numerous lineage, M1-M106 with 46 individuals, also has the widest geographic  
195 distribution, with individuals from the Wallacean Islands of eastern Indonesia and the Torres  
196 Strait nested within New Guinean diversity (Figure 3, Figure S3, File S2).

197

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

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

212

### 213 **A long pause followed initial settlement**

214 Regional chrY lineages show a striking lack of new diversity between 40 and 25 kya (Figure  
215 2, Figures S2 and S3, Tables S3 and S4). Existing lineages in C-M130, M-P256 and S-B254  
216 survived through this gap, but most show no evidence of further radiation during this time.  
217 This striking lack of new diversity between 40 and 25 kya is also observed in the regional  
218 mitochondrial DNA (mtDNA) lineages (Pedro et al. 2020). ChrY haplogroup S4'6-PR495.2,  
219 which is mostly found in New Britain (eastern Papua New Guinea), is the only lineage that  
220 split during the hiatus, ~30 kya (95% CI: 28.0-33.2 kya). Noting the limitation that modern  
221 sampling may miss lineages that have gone extinct, the observed widespread pause in  
222 phylogenetic diversification appears to be characteristic of low population size and persistent  
223 population structure during much of the Upper Pleistocene, in agreement with maternal  
224 lineage patterns (Pedro et al. 2020), the relatively small numbers of Late Pleistocene  
225 archaeological sites identified (Williams 2003; Summerhayes et al. 2017) and the strong  
226 territoriality and geographical stability seen in modern ethnographic studies of New Guinea  
227 and Indigenous Australian societies (Muke 1993; Sutton 2003). In notable contrast,  
228 haplogroup O-M175 on Mainland Asia, which is now the dominating lineage in Eastern  
229 Eurasians and comprises more than a quarter of all male lineages in the world (Yan et al.  
230 2014), continued to radiate throughout the Pleistocene. About 33 kya (95% CI: 30.6-36.6



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

239

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

242 After 25 kya, major periods of lineage radiation are observed across ISEA, New Guinea and  
243 Australia. These lineages stem from different chrY sub-branches arising from northern Sahul  
244 diversity (Figure 3, Figure S3) and they dispersed at different times. The most numerous  
245 Australian lineage in this sample, C4-M347 (carried by 6 men), diversified between 25 and  
246 17 kya, coinciding with a major split ~26 kya (95% CI: 23.2-28.9 kya) in a second Australian  
247 lineage S7'8-FGC38729 present in three men (Figures S2 and S3). About 23 kya (95% CI:  
248 20.3-25.5 kya), C2-M38 diverged into two lineages, one now found from Wallacea (eastern  
249 Indonesia) to New Britain, with the other largely radiating within mainland New Guinea itself  
250 (Figure S2). M1-M106 split ~21 kya (95% CI: 19.5-23.9 kya) into M1a'g-Page1 and M1h-  
251 MY6 (42 and 4 men, respectively), and is found today in New Guinea and New Britain, but  
252 also radiated westward (Figure 3, Figure S3). The most numerous sublineage in New Guinea,  
253 S9-P202, split ~19 kya (95% CI: 17.1-20.5 kya), with subsequent expansion ~15-13 kya  
254 creating off-shoots of different sublineages across the Wallacean Islands of Alor, Lembata,  
255 Sulawesi and Borneo (Figure 3, Figure S3). Over the same timeframe, C7-B67 split ~17 kya  
256 (95% CI: 14.6-19.6 kya) and spread between Borneo, Flores and Lembata (Figure S2), and  
257 S10-SK1807 gave rise to descendant lineages in New Guinea, New Ireland and Eastern  
258 Indonesia around the same time (95% CI: 15.9-19.6 kya) (Figure 3, Figure S3). Lineages with  
259 New Guinean origins are mostly found on the Wallacean Islands, with a few in other regions  
260 as well.

261

262 The prevailing view of regional history is of two major human movements – the initial  
263 settlement (Clarkson et al. 2017; O’Connell et al. 2018; Summerhayes et al. 2010; Leavesley  
264 et al. 2004) and the arrival of Austronesian speakers 3-3.5 kya (Bellwood 2004; Bellwood  
265 2006), implicitly suggesting that little occurred in this region during the Upper Pleistocene.  
266 However, the more resolved chrY tree with denser sampling uncovers a far more complex  
267 paternal demographic history, consistent with some of the newer studies in genetics (Gomes  
268 et al. 2015; Pedro et al. 2020; Purnomo et al. 2021), linguistics (Schapper 2017) and  
269 archaeology (Summerhayes et al. 2007, 2017; Bellwood 2019). In particular, the increase in  
270 genetic lineage diversification correlates well with increasing population interactions and  
271 population sizes postulated from the archaeological record around 25-20 kya, based on  
272 animal, plant and object (e.g., obsidian) translocation between Northern Sahul regions,  
273 changes in the subsistence economy, increasing occupation of Highland New Guinea, and  
274 settlement of distant islands (e.g., Manus), all of which suggest population size increases and  
275 connection of local dynamics to larger regional networks (Leavesley et al. 2004; Specht  
276 2005; Summerhayes et al. 2007, 2017).

277  
278 We see these connections within a number of haplogroups. In S-B254, the split within the  
279 S4’6-PR495.2 lineages of New Britain occurred ~30 kya (95% CI: 28.0-33.2 kya), whereas  
280 S5’6-FT17887 diversified first ~23 kya (95% CI: 20.9-25.3 kya) and then ~16 kya (95% CI:  
281 14.4-17.7 kya) (Figure 3, Figure S3). Its sister lineage, S4-P315, appears to have been  
282 isolated for some time, only diverging ~18 kya (95% CI: 15.9-20.3 kya) when it split between  
283 the New Britain and Queensland (Australia) groups. A much more recent split ~3.6 kya (95%  
284 CI: 2.7-4.7 kya) is also seen between men from these two locations (Figure 3, Figure S3).  
285 Hence, despite perceptions of long-term isolation between New Guinea and Australia  
286 (Bergström et al. 2016; Nagle et al. 2016), joint analyses of chrY from Papua New Guinea  
287 (Vernot et al. 2016) and Australia (Bergström et al. 2016) reveals phylogenetic relationships  
288 between lineages from New Britain and Australia that are much older than previously  
289 thought. This phylogenetic connection is beyond simply the Torres Strait region and is  
290 similar to that seen in autosomal data (Brucato et al. 2021). The coalescence date suggests an  
291 upper estimate of the Late Pleistocene for possible contact between these regions (Figure 3,  
292 Figure S2).

293

294 While no archaeological evidence supports direct contact between New Britain and Australia,  
295 the Torres Strait Islands (between New Guinea and Australia) may have served as a bridge  
296 between these two regions. Indeed trading and intermarriage has been reported between  
297 Torres Strait Islanders and Indigenous Australians (Beckett 1987), and some New Guinean  
298 lineages (mtDNA Q1, P2 and P3, and chrY M1c6-Z42300) are shared by men with Torres  
299 Strait ancestry (Figure S3) (Nagle et al. 2017; Pedro et al. 2020). This genetic affinity  
300 between Bismarck Islanders and Indigenous Australians may also result from genetic  
301 exchanges pre-dating the divergence between these populations during the postglacial period  
302 (18-10 kya) (Figure 3, Figure S3) (Nagle et al. 2017; Pedro et al. 2020; Brucato et al. 2021).  
303 More recent contact was also possible during the rapid spread of the Lapita Cultural  
304 Complex, which originated in the Bismarck Archipelago 3.3 kya (Denham et al. 2012),  
305 spread to the south Papuan Coast 2.9 kya (David et al. 2019), and may have subsequently  
306 influenced islands off the Queensland coast by 2.5 kya (Lilley 2019).

307

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

315

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

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

323 population increase from the Holocene transition onwards, as part of a longer process that  
324 started in the late Pleistocene (25-20 kya) and culminated with the Austronesian dispersal  
325 across the Northern Sahul region (Williams 2003; Specht 2005; Bellwood 2006; Donohue  
326 and Denham 2010; Summerhayes et al. 2007, 2017). We also see a marked increase in the  
327 radiation of chrY lineages, coupled with extensive mobility across the region. Haplogroup S-  
328 B254 shows splits in New Britain between S4b-MY103 and S4c-Z41921, with the lineage  
329 later shared between men from New Britain and Queensland (Figure 3, Figure S3).

330 Haplogroup M-P256 lineages in New Guinea and New Britain also diversified during this  
331 time. Between ~14 and 9 kya, the most numerous and widely spread sub-branch M1a'g-  
332 Page1 expanded rapidly within New Guinea (M1b-MY39, M1c-B256, M1e'f-Z43413) and  
333 New Britain (M1g-P87), with some lineages spreading west to eastern Indonesia (M1e-  
334 MY19, M1a-MY26) (Figure 3, Figure S3). Within haplogroup C-M130, ~13-10 kya, a series  
335 of divergences within C2a-M208 in New Guinea gave first an offshoot to Wallacea and then  
336 formed the lineages that, much later, spread out into the Pacific (Figure S2) (Cox et al. 2007).  
337 Within the few Australian lineages, a split in C4-MY1088 in Queensland occurred ~11 kya  
338 (95% CI: 9.2-13.0 kya) (Figure S2).

339

340 During this time, haplogroup O-M175 diversification patterns are similar to those of other  
341 chrY haplogroups in the region. Between 14-10 kya, major splits occurred in O1a-F31, O2-  
342 PK4 and many other O-M175 lineages, some of which are shared between Mainland and  
343 Island Southeast Asia. Of particular note, O3f-M7 diversified ~15 kya (95% CI: 11.2-17.4  
344 kya), giving rise to branches found on Mainland Asia, Taiwan and the Philippines, with the  
345 latter two lineages diverging from each other ~10 kya (95% CI: 7.6-13.0 kya) within O3f-  
346 Y26403. O3f-M7 occurs widely across western Indonesia at low frequency today, and had  
347 previously been thought to reflect connections with mainland China during the historic era  
348 (Karafet et al. 2010). Although sampling remains low, the current O3f-M7 phylogeny  
349 suggests a potentially much earlier upper limit to the movement of people carrying this  
350 lineage, either in the Late Pleistocene or starting ~15 kya (Figure 3, Figure S4, Table S5).

351 Older connections between Mainland Asia and western Indonesia were also recently shown  
352 by high levels of pre-Neolithic East Asian ancestry in autosomal genetic data of the young  
353 female buried ~7.3 kya at Leang Panninge cave in Sulawesi (Carlhoff et al. 2021).

354

355 **Holocene dispersals carried few external lineages from Mainland East Asia**

356 The sea had risen to near its current level by 8-7 kya, leading to the creation of rich coastal  
357 floodplains (Terrell et al. 2011), and rising temperatures gradually led to the higher altitudes  
358 of New Guinea being more suitable for permanent human occupation (Summerhays et al.  
359 2017). Indigenous food plant cultivation in the highlands of New Guinea began ~10 kya  
360 (Hope and Haberle 2005; Denham and Haberle 2015), coinciding with the expansion seen in  
361 M1-Page1 and various S-B254 sub-lineages (Figure 3, Table S4). However, the Holocene  
362 story is in many ways dominated by haplogroup O-M175. From 10-2 kya, many expansion  
363 events occurred within hg O-M175 (Figure 4, Figure S4), including the rise of the most  
364 numerous paternal lineages in the world today (Yan et al. 2014). Within Island Southeast  
365 Asia, O3i-B451 and O2a2-F789 stem from this time of major haplogroup O-M175 expansion.  
366 However, the lineage with the greatest impact was O1-M119, which arose ~33 kya (95% CI:  
367 24.5-36.9 kya), diverged into sublineages ~17 kya (95% CI: 12.3-19.3 kya), but underwent a  
368 striking period of radiation during the mid-Holocene (Figure 4, Figure S4, Table S5).

369

370 O1-M119 contains three main sublineages: O1a-F31, found in 35 men from across Mainland  
371 Asia, ISEA and in New Guinea; O1b-F819 found in nine men from the Philippines, Sunda  
372 and the Wallacean islands; and O1c-M110 found in 16 men from the Philippines, western  
373 Indonesia and New Guinea. O1a-F31 radiated rapidly between ~8-4 kya, with a sub-branch in  
374 Taiwan and Malaysia separating ~4 kya (95% CI: 2.7-4.9 kya). O1b-F819 lineages in the  
375 Philippines and Indonesia split ~9 kya (95% CI: 7.0-11.0 kya), but radiated rapidly ~4 kya  
376 (95% CI: 3.1-5.0 kya). Today, O1c-M110 is found only within Island Southeast Asia and  
377 New Guinea, where it shows evidence of rapid expansion 6-5 kya (Figure 4, Figure S4, Table  
378 S5). We note the limitation that all molecular dates have some level of uncertainty (Tables  
379 S3-S5), and not all known O1-M119 diversity is included here (Sun et al. 2021).

380 Nevertheless, the striking radiations observed are characteristic of a diverse population of  
381 Mainland Asian origin moving into Island Southeast Asia and expanding rapidly. Due to their  
382 timing and current geographical distribution, these genetic lineages are typically associated  
383 with Austronesian-speaking peoples and Neolithic practices, which spread over a period of  
384 time, starting around 6 kya in Taiwan and reaching eastern Indonesia and New Guinea ~3.5

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

387

### 388 **Holocene expansions in indigenous Y chromosome lineages**

389 While arrivals from Mainland Asia have been the traditional focus of mid-Holocene studies  
390 in Island Southeast Asia, we identify a series of simultaneous radiations in lineages with  
391 long-standing local connections in C-M130, M-P256 and S-B254. In addition, C8-F725  
392 seeded many new lineages in a short time period between ~6.5-5 kya, including radiating into  
393 the Philippines, western Indonesia (Borneo and Mentawai) and the Wallacean islands of  
394 eastern Indonesia (Figure S2). A second wave of expansions occurred around ~4-3 kya when  
395 lineages in Borneo split from their neighbors in Malaysia or Sulawesi. Local inter-island  
396 radiations on Flores and Mentawai also began ~2.5-1.5 kya (Figure S2). In haplogroup  
397 M1a'g-Page1, multiple lineages experienced radiations between 7.5-2 kya (downstream of  
398 M1c-B256, M1b-MY39, M1a-MY26 and M1g-P87), mostly within New Guinea, Wallacea  
399 and the Torres Islands (Figure 3, Figure S3). In haplogroup S-B254, the previously  
400 mentioned split within S4c-ZZ41912 also appeared between Queensland (Australia) and New  
401 Britain ~3.6 kya (95% CI: 2730-4719 kya) (Figure 3, Figure S3). Many of these expansions  
402 and population movements that occurred after ~4 kya were likely triggered by the  
403 introduction of agriculture from Austronesian-speaking cultural communities (Deng et al.  
404 2020; Alam et al. 2021), perhaps coupled in some places with the growing influence of local  
405 Papuan agricultural practices that had commenced as early as ~10 kya (Denham and Haberle  
406 2015).

407

### 408 **Population mobility and lineage expansion has continued to the present**

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

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

418

## 419 **Conclusions**

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

433

434 The improved Y chromosome topology also highlights more localized events with important  
435 historical implications. O3-M7 lineages from Taiwan and the Philippines diverged from  
436 mainland Asian groups ~15 kya, earlier than previously thought (Karafet et al. 2010). This  
437 coalescence date provides an upper limit to the movements of people, possibly supporting the  
438 growing view from other evidence of early movements between MSEA and ISEA well  
439 predating the Neolithic period (Karafet et al. 2010; Trejaut et al. 2014; Vallée, Luciani and  
440 Cox 2016). Despite the scarcity of chrY sequence data from Australia (Bergström et al.  
441 2016), we find that some Australian lineages are nested within New Britain diversity,  
442 indicating that Australia experienced at least some interaction with the Papuan world. These  
443 and other contacts emphasize the potential of less common chrY lineages to illuminate the  
444 subtleties of regional contacts. The resolved phylogenetic framework of the Y chromosome

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

447

448

## 449 **Materials and Methods**

### 450 **Human subjects**

451 To develop a fully resolved picture of Y chromosome diversity in this region, we sequenced  
452 14 new male samples (Figure 1, Table S1) on the Illumina platform, and combined these with  
453 26 chrY sequences from Papua New Guinea (Brucato et al. 2021) and 112 chrY sequences  
454 from Island Southeast Asia (Jacobs et al. 2019) that were not previously analyzed for their Y  
455 chromosome diversity. We then combined this new dataset with all available chrY sequences  
456 from the region, supplemented with samples from a broader geographic and phylogenetic  
457 context taken from previously published full Y chromosome studies on neighboring East  
458 Asia, Mainland Southeast Asia and other world regions (1000 Genomes Project Consortium  
459 et al. 2010; Drmanac et al. 2010; PGP 2014; Carmi et al. 2014; Karmin et al. 2015;  
460 Bergström et al. 2016; Ilumäe et al. 2016; Malaspinas et al. 2016; Mallick et al. 2016; Poznik  
461 et al. 2016; Vernot et al. 2016; Zheng-Bradley et al. 2017; Hudjashov et al. 2018; Bergström  
462 et al. 2020). The final dataset comprised 795 samples (Table S1).

463

### 464 **Ethics**

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



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

477

#### 478 **Data availability**

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

482

#### 483 **Sequencing, mapping and genotyping**

484 All newly generated chrY sequences were generated using the Illumina technology (Illumina,  
485 San Diego, CA, USA) on the HiSeq instrument (PCR-free protocol) with 30x genome-wide  
486 coverage. We employed a series of filters and sequence masks to limit the dataset to high  
487 quality variant calls, resulting in 9.7 MB of complete chrY sequence per sample. We used the  
488 same processing pipeline for all Illumina data. FASTQ files were mapped with BWA-MEM  
489 v0.7.12 (Li and Durbin 2009) to the human reference hs37d5  
490 ([http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2\\_reference\\_assembly\\_s](http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence)  
491 [equence](http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence)). Read duplicates were removed with Picard v2.12.0  
492 (<http://broadinstitute.github.io/picard/>) and the remaining unique reads were realigned around  
493 known indels, followed by base quality score recalibration (BQSR) using GATK v3.8 (Poplin  
494 et al. 2018). Variant calling was performed with GATK HaplotypeCaller in haploid mode.  
495 All-sites VCF files were filtered with bcftools v1.9 (Li 2011). The Illumina data and  
496 previously filtered data from the Complete Genomics technology (Table S1) were merged  
497 with GATK CombineVariants v3.8 (Poplin et al. 2018). We extracted the effective overlap  
498 between the two datasets by masking out all positions with a 5% or higher proportion of  
499 missing genotypes in either the Illumina or Complete Genomics datasets. We additionally  
500 excluded regions with poor mappability as described previously (Karmin et al. 2015; Poznik  
501 et al. 2016), resulting in a total of 9.7 Mb of analyzed chrY sequence. Variant positions used  
502 for phylogenetic reconstruction in each haplogroup are listed in Tables S6-S8.

503

#### 504 **Phylogeny reconstruction and dating**

505 For the complete dataset of 795 samples, a maximum likelihood (ML) tree was constructed  
506 with GTRCAT using 200 rapid bootstrap inferences followed by a thorough ML search  
507 executed by RaXML v8.0.0 (Stamatakis 2014) (Figure 1 and S1). All identified variants for  
508 shared nodes were annotated to the ML trees and curated manually (Tables S6-S8). We then  
509 used the software package BEAST v1.7.5 (Drummond et al. 2012) to simultaneously  
510 reconstruct phylogenies and estimate coalescence times. We ran three separate lineage-based  
511 analyses for haplogroups C-M130, MS-PR2099 and O-M175. For each haplogroup we used  
512 the same parameters: a Bayesian skyline coalescent tree prior, the general time reversible  
513 (GTR) substitution model with gamma-distributed rates, a relaxed lognormal clock and the  
514 piecewise-constant coalescent model with the number of groups set to 10. For each  
515 haplogroup, four parallel runs with different random number seeds were performed. Each run  
516 had at least sixty million chains logged every 3000 steps. The results were visualized and  
517 checked for effective sample size above 200 using Tracer v1.4. Four parallel runs were  
518 combined with LogCombiner and the initial 25% discarded as burn-in. Coalescence time  
519 estimates were computed with normally distributed age priors from a previously published  
520 phylogeny that had used a mutation rate of  $0.74 \times 10^{-9}$  (95% CI:  $0.63-0.95 \times 10^{-9}$ ) per base per  
521 year (Karmin et al. 2015). The standard deviation (SD) was set to cover the published  
522 confidence intervals of the calibration nodes (Table S2). For haplogroups MS-PR2099 and O-  
523 M175, we used an extended set of samples to provide the broader global structure of the  
524 phylogeny and the same two samples from haplogroup C-M130 as the outgroup (Figure S2  
525 and S4). The BEAST analyses of MS-PR2099 had 154 individuals and 17621 variants  
526 (Figure S3, Table S7), whereas O-M175 had 259 individuals and 19743 variants (Figure S4,  
527 Table S8). For both phylogenies, the calibration point for coalescence age estimates was set  
528 for the GT-F1329 node with an age of 51557 years (95% confidence interval [CI]: 50206-  
529 52917) (Karmin et al. 2015) and a SD of  $\pm 1355$  in the BEAST analyses. The phylogeny of  
530 haplogroup C-M130 was rooted with two samples from J-M304 (Figure S2) and consisted of  
531 131 individuals with 9977 variant positions (Table S6). The calibration point was set to the  
532 root of haplogroup C-M130 with an age of 50865 years (95% CI: 49191-52699) (Karmin et  
533 al. 2015), with SD set to  $\pm 1835$  in BEAST. We used Newick Utilities (Junier and Zdobnov  
534 2010) for tree processing, statistical analyses and plotting was conducted in R4.1.2 (R Core

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

537

### 538 **Acknowledgements**

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

547

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

552

553

554 **Competing Interest Statement:** The authors declare no conflicts of interest

555

### 556 **Supplemental Information**

557

558	Supplemental Table	S1	Samples in the study
559		S2	Calibration points
560		S3	Haplogroup C-M130 coalescence ages
561		S4	Haplogroup MS-PR2099 coalescence ages
562		S5	Haplogroup O-M175 coalescence ages
563		S6	Variants in haplogroup C-M130
564		S7	Variants in haplogroup MS-PR2099
565		S8	Variants in haplogroup O-M175

566 S9 Method details from mapping to masking

567

568

569 **Supplemental Figures S1-S3**

570 Figure

571 S1. Distribution of the number of mutations from terminal tips to the common BT

572 node on the ML tree of 795 individuals

573 S2. Dated phylogenetic tree of haplogroup C-M130

574 S3. Dated phylogenetic tree of haplogroup MS-PR2099

575 S4. Dated phylogenetic tree of haplogroup O-M175

576

577 **Supplemental Files S1-S4**

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

579 S2. Phylogeny of haplogroup MS-PR2099 in FigTree format

580 S3. Phylogeny of haplogroup O-M175 in FigTree format

581 S4. Whitelist of chrY regions in bed format

582

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

613 **Figure 3. Contacts between Northern Sahul and neighboring areas within the MS-**  
614 **PR2099 lineages.** The basic structure of haplogroup MS-PR2099 lineages is shown in the  
615 upper inset; circles denote sub-lineages that show evidence of contact between Northern  
616 Sahul and other regions, and are illustrated in more detail in the main figure. Colors on the  
617 tree and map denote the geographic origin of individuals. Light green boxes show westward  
618 paternal gene flow from Northern Sahul mostly to the Wallacean Islands, but also beyond;  
619 darker green boxes denote contact between New Britain and Australia. Light brick brown  
620 boxes show likely back migration to New Guinea from Wallacea. Light teal represents  
621 lineages from Pacific islands. For more details, see Figure S3 and Table S1.  
622

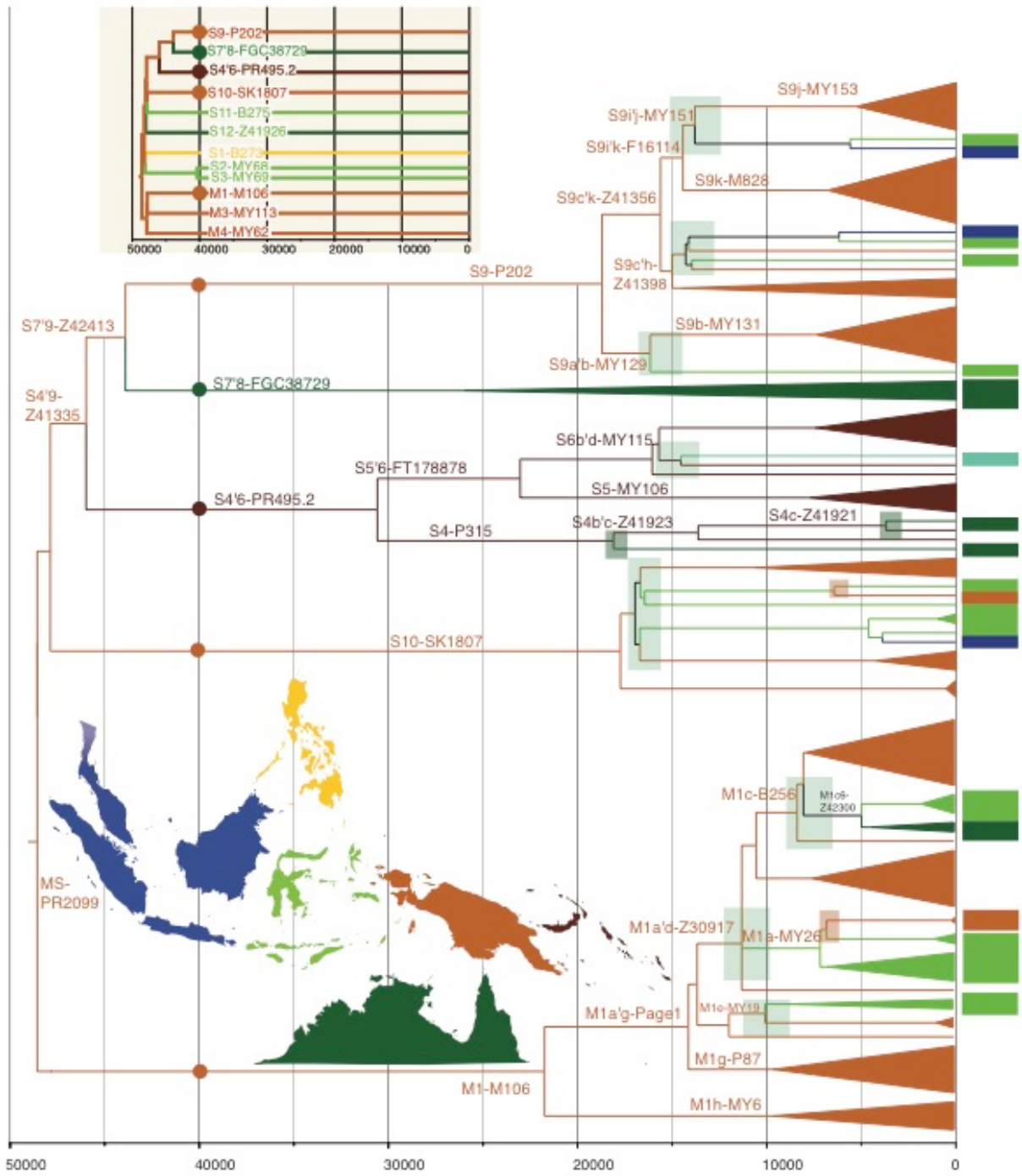
623 **Figure 4. Haplogroup O-M175 sub-lineages in Island Southeast Asia, New Guinea and**  
624 **Oceania.** A number of distinct sublineages (black) of O-M175 have penetrated from the  
625 Asian mainland into ISEA and beyond, and are likely associated with the expansion of  
626 Austronesian speakers. Most of these lineages coalesce ~6-5 kya within subsets of the wider  
627 Asian diversity, except for O1b-F819 and O3f-M7 sublineages, which have earlier  
628 diversification times (details in Figure S4 and Table S1). In contrast, lineage NO2-MY1506,  
629 found today on Sunda and Sulawesi, split very early from its South Asian sister lineage.  
630  
631





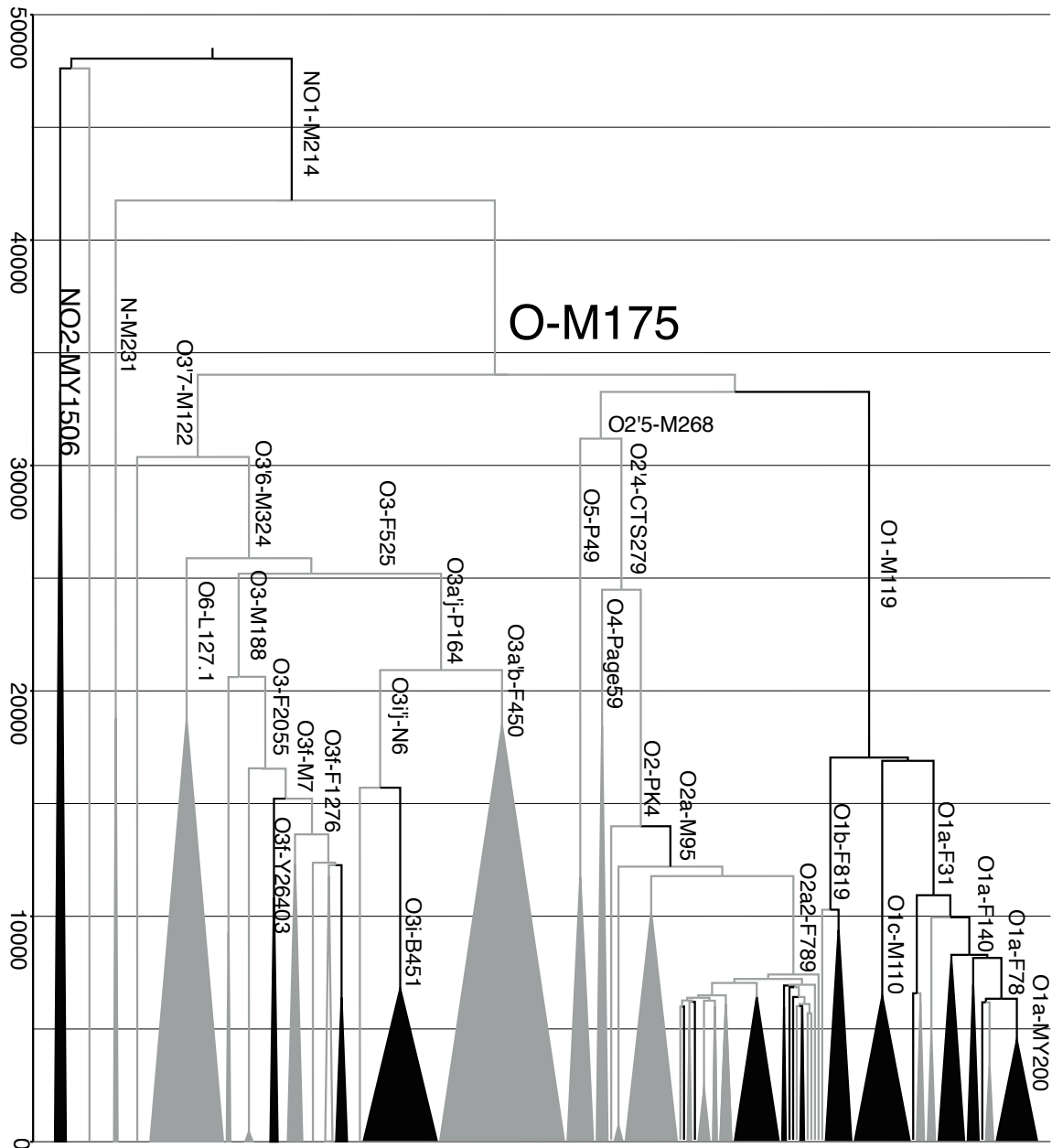


637 Figure 3  
 638



639

640 Figure 4  
641



642

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, Stattenger 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, Moyses-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](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-](http://www.R-project.org/)  
825 [project.org/](http://www.R-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 GRCh38t1e. *Gigascience* 6:1–8.

884