



HAL
open science

Papua New Guinean genomes reveal the complex settlement of north Sahul

Nicolas Brucato, Mathilde André, Roxanne Tsang, Lauri Saag, Jason Kariwiga, Kylie Sesuki, Teppsy Beni, William Pomat, John Muke, Vincent Meyer, et al.

► **To cite this version:**

Nicolas Brucato, Mathilde André, Roxanne Tsang, Lauri Saag, Jason Kariwiga, et al.. Papua New Guinean genomes reveal the complex settlement of north Sahul. *Molecular Biology and Evolution*, 2021, 10.1093/molbev/msab238 . hal-03320801

HAL Id: hal-03320801

<https://hal.science/hal-03320801>

Submitted on 16 Aug 2021

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 Papua New Guinean genomes reveal the complex settlement of north Sahul

2
3 Nicolas Brucato¹, Mathilde Andr ^{1,2}, Roxanne Tsang^{3,4}, Lauri Saag², Jason Kariwiga^{4,5}, Kylie
4 Sesuki⁴, Teppsy Beni⁴, William Pomat⁶, John Muke⁷, Vincent Meyer⁸, Anne Boland⁸, Jean-
5 Fran ois Deleuze⁸, Herawati Sudoyo⁹, Mayukh Mondal², Luca Pagani^{2,10}, Irene Gallego
6 Romero¹¹, Mait Metspalu², Murray P. Cox¹², Matthew Leavesley^{4,13,14}, Fran ois-Xavier Ricaut¹

7
8 ¹Laboratoire  volution and Diversit  Biologique (EDB UMR 5174), Universit  de Toulouse
9 Midi-Pyr n es, CNRS, IRD, UPS. 118 route de Narbonne, Bat 4R1, 31062 Toulouse cedex 9,
10 France

11 ²Institute of Genomics, University of Tartu, Tartu, Tartumaa 51010, Estonia

12 ³School of Humanities, Languages and Social Science and Place, Evolution and Rock Art
13 Heritage Unit, Griffith University Centre for Social and Cultural Research, Griffith University,
14 Australia

15 ⁴Strand of Anthropology, Sociology and Archaeology, School of Humanities and Social
16 Sciences, University of Papua New Guinea, PO Box 320, University 134, National Capital
17 District, Papua New Guinea

18 ⁵School of Social Science, University of Queensland, Australia, St Lucia, QLD 4072, Australia

19 ⁶Papua New Guinea Institute of Medical Research, Goroka, Papua New Guinea

20 ⁷Social Research Institute, Papua New Guinea

21 ⁸Universit  Paris-Saclay, CEA, Centre National de Recherche en G nomique Humaine, 91057
22 Evry, France

23 ⁹Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular Biology, Jakarta
24 10430, Indonesia

25 ¹⁰ Department of Biology, University of Padua, Italy

26 ¹¹ Melbourne Integrative Genomics, University of Melbourne, Australia

27 ¹² Statistics and Bioinformatics Group, School of Fundamental Sciences, Massey University,
28 Palmerston North 4442, New Zealand

29 ¹³ College of Arts, Society and Education, James Cook University, P.O.
30 Box 6811, Cairns, Queensland, 4870, Australia

31 ¹⁴ ARC Centre of Excellence for Australian Biodiversity and Heritage,
32 University of Wollongong, Wollongong, New south Wales, 2522, Australia

33

34 Corresponding authors:

35 Nicolas Brucato: nicolasbrucato@gmail.com

36 François-Xavier Ricaut: francois-xavier.ricaut@univ-tlse3.fr

37

38 Keywords:

39 Papuan; Human genome; Demographic history; Sahul; Oceania

40

41 **Abstract**

42 The settlement of Sahul, the lost continent of Oceania, remains one of the most ancient and
43 debated human migrations. Modern New Guineans inherited a unique genetic diversity
44 tracing back 50,000 years, and yet there is currently no model reconstructing their past
45 population dynamics. We generated 58 new whole genome sequences from Papua New
46 Guinea, filling geographical gaps in previous sampling, specifically to address alternative
47 scenarios of the initial migration to Sahul and the settlement of New Guinea. Here, we
48 present the first genomic models for the settlement of northeast Sahul considering one or
49 two migrations from Wallacea. Both models fit our dataset, reinforcing the idea that
50 ancestral groups to New Guinean and Indigenous Australians split early, potentially during
51 their migration in Wallacea where the northern route could have been favored. The earliest
52 period of human presence in Sahul was an era of interactions and gene flow between
53 related but already differentiated groups, from whom all modern New Guineans, Bismarck
54 islanders and Indigenous Australians descend. The settlement of New Guinea was probably
55 initiated from its southeast region, where the oldest archaeological sites have been found.
56 This was followed by two migrations into the south and north lowlands that ultimately
57 reached the west and east highlands. We also identify ancient gene flows between
58 populations in New Guinea, Australia, East Indonesia and the Bismarck Archipelago,
59 emphasizing the fact that the anthropological landscape during the early period of Sahul
60 settlement was highly dynamic rather than the traditional view of extensive isolation.

61

62 New Guineans are descendants of one of the most exceptional and ancient migrations in
63 human history, but the nature of their initial settlement of New Guinea is still veiled in
64 mystery (Allen and O'Connell 2020). When modern human first arrived in Southeast Asia 50-
65 70 thousand years ago (kya) (Westaway, et al. 2017), the lower sea level during a colder
66 climate caused two gigantic ancient continents to emerge: Sunda joined many current
67 southeast Asian islands with mainland Asia, while Australia and New Guinea formed a single
68 landmass called Sahul. These vast territories were separated by a substantial water crossing,
69 Wallacea, containing a scattering of small islands and creating the major ecological barrier
70 known as the Wallace line (Wallace 1869).

71 The first anatomically modern humans to cross from Sunda to Sahul numbered probably no
72 more than a few hundred, but migration through the Wallacean islands was nonetheless
73 likely a deliberate effort (Bird, et al. 2019; Bradshaw, et al. 2019). Several initial settlement
74 paths are possible based on paleogeological reconstructions: a northern route through what
75 today are the Maluku islands (Kealy, et al. 2018; Norman, et al. 2018), and a southern route
76 through Flores and Timor (Birdsell 1977; Norman, et al. 2018). Although not mutually
77 exclusive, for decades the consensus has been that a single migration from Sunda to Sahul
78 created the entire biological diversity of both New Guineans and Indigenous Australians
79 today (Malaspina, et al. 2016; O'Connell, et al. 2018). Both populations have unique
80 genomic diversities marked by their high percentage of genetic inheritance coming from
81 Denisovans (3-3.5%) (Jacobs, et al. 2019; Reich, et al. 2011; Sankararaman, et al. 2016;
82 Vernot, et al. 2016). However, recent studies now suggest that the arrival of just one human
83 group is unlikely to be sufficient to explain the deep genetic diversity still observed in this
84 region (Allen and O'Connell 2020; Pedro, et al. 2020). In part, this argument draws from the
85 uniparental genetic markers found in New Guineans and Indigenous Australians. While both
86 populations undoubtedly share a common maternal genetic ancestry, marked by the
87 presence of haplogroups M*, N* and R*, the diversity of their respective mitochondrial DNA
88 (mtDNA) sub-haplogroups indicate stark genetic differences (Redd and Stoneking 1999).
89 Most mtDNA sub-haplogroups are exclusive to either New Guinea (e.g., P1, P2, P4a, M27,
90 M28, Q*) (Pedro, et al. 2020) or Australia (e.g., P8, P4b, M42, S*) (Tobler, et al. 2017), in
91 agreement with local evolution of some of these region-specific haplogroups. Coalescence
92 dates between Australian and New Guinean maternal lineages suggest a split at the time of
93 the initial settlement of Sahul (Tobler, et al. 2017), or potentially before, as shown by the

94 coalescence dates of P2'P8'P10 (49kya), P4 (50kya) (Pedro, et al. 2020). Similar patterns are
95 observed for paternal genetic lineages on the Y chromosome (Bergstrom, et al. 2016; Nagle,
96 et al. 2016). Such different uniparental genetic patterns suggest that either New Guineans
97 and Indigenous Australians separated very early after the initial migration to Sahul and that
98 uniparental lineages segregated accordingly, or that both populations descend from two
99 initially separate migrations (Redd and Stoneking 1999), perhaps at different times as
100 suggested by some archaeological records (Clarkson, et al. 2017).

101 Once in Sahul, modern human dispersals within what is now the island of New Guinea began
102 by at least 49kya (Summerhayes, et al. 2010). Given the paucity of archaeological records,
103 how New Guinea was settled remains poorly resolved. Of particular note, the initial entry
104 point into New Guinea has proven particularly difficult to locate, leading to alternative
105 hypotheses of migration via the New Guinean north coast, through the mountainous
106 highlands, along the southern foothills, or even further south via a vast plain that is now
107 submerged and forms the Arafura Sea (Summerhayes, et al. 2017). Archaeological records
108 perhaps favor the Arafura plain pathway, since the oldest known archaeological sites in New
109 Guinea are located in its southeast region. The suggestion is that even earlier sites, marking
110 those first movements, now lie flooded under the Arafura Sea (Summerhayes, et al. 2017).
111 Nevertheless, the alternative hypotheses remain feasible. Although the highlands were not
112 hospitable during the Upper Pleistocene, long-term human presence is attested there from
113 at least 25kya (Summerhayes, et al. 2017). Major rivers carved into the New Guinean geology
114 may have formed early natural corridors during initial settlement, allowing human mobility
115 as they still do along the Markham and Ramu Rivers in the northeast, the Fly River in the
116 south and the Sepik River in the north (Pawley 2005). The Sepik route has been supported by
117 close genetic relationships between modern New Guinea highlanders and lowland groups
118 from the Sepik region, but this association may date to a later time period (Bergstrom, et al.
119 2017).

120 After the initial settlement of New Guinea, independent development of agricultural
121 practices in New Guinea between 10-7 kya led to a unique Neolithic revolution (Denham, et
122 al. 2003; Golson, et al. 2017), driving demographic expansion in the highlands that heavily
123 impacted the genetic and linguistic diversity of these populations (Bergstrom, et al. 2017;
124 Greenhill 2015; Reesink, et al. 2009). By the time Austronesian settlers arrived around 3kya

125 (Bellwood 2007; Lipson, et al. 2014), New Guineans had already developed a complex
126 network of diverse cultures (Pawley 2005).

127 Given its extraordinary diversity today, comparable to large continental areas elsewhere, the
128 genetics of New Guinean populations are poorly explored, and scenarios of initial and post-
129 settlement movements are lacking. We have therefore generated 58 high quality whole
130 genome sequences (WGS) from Papua New Guinean individuals (Supplementary table 1).
131 Coupling these with a dataset of 905 genomes from diverse human groups, but particularly
132 individuals from neighboring Indonesia and Australia (Supplementary table 2), we
133 reconstruct the key genetic events during the Upper Pleistocene that led to the settlement
134 of New Guinea. Our two primary questions are: was the migration to Sahul conducted by
135 one or several human groups, and what migratory routes did they take? And how did New
136 Guineans subsequently settle the diverse environments of the modern island, and what
137 were the dynamics of this settlement?

138

139 **Results**

140 *Samples and diversity*

141 Fifty-eight Papua New Guinean participants were included in this study to increase the
142 geographical coverage of New Guinea, especially in the southern regions underrepresented
143 in published whole genome datasets (Supplementary figure 1 and supplementary table 1).
144 Whole genome sequencing data were obtained at an average depth of 36x, characterized for
145 41,250,479 SNPs. Uniparental haplogroups were obtained from mitochondrial DNA and Y
146 chromosome data (Supplementary table 1). New Guinean mitochondrial DNA diversity was
147 previously described (Pedro, et al. 2020), showing mostly haplogroups specific to the region
148 (Q*:38%; P*:29%) as well as the presence of haplogroups frequent in Austronesian
149 populations (B4a1a1:20%). New Guinean Y chromosome diversity is dominated by M1 (39%)
150 and S* (18%) haplogroups frequent in the region, and haplogroups of Austronesian origin
151 (O1:14%) as previously described (Karafet, et al. 2010; Mona, et al. 2007). Autosomal data
152 were combined with published genomes, notably from Indonesia including the western half
153 of New Guinea Island)(Jacobs, et al. 2019), northeast Australia (Mallick, et al. 2016) and the
154 Bismarck Archipelago (Vernot, et al. 2016), and with genotyping data from North Maluku
155 (Kusuma, et al. 2017) and imputed genotyping data from Papua New Guinea (Bergstrom, et
156 al. 2017), resulting in a dataset of 966 individuals (including genomes from Denisovan,

157 Neandertal and Chimpanzee, Supplementary table 2). By increasing the number of New
158 Guinean whole genome sequencing data as well as the geographic resolution of New
159 Guinean genetic diversity, we created a high quality dataset (e.g., with good phasing
160 (Delaneau, et al. 2012) and imputation (Howie, et al. 2009)), thus allowing us to perform
161 fine-scale demographic analyses.

162

163 *Characterization of genetic structure*

164 We analyzed genetic patterns in individuals from New Guinea, northeast Australia, the
165 Bismarck Archipelago and Wallacea to identify key regional genetic groups. A coancestry
166 matrix was built by scanning each pair of genomes to detect Identity-By-Descent (IBD)
167 fragments using fineSTRUCTURE (Lawson, et al. 2012)(Supplementary Fig. 2). Most genomes
168 clustered according to their respective geographical locations, defining 15 regional groups,
169 including seven regional groups from the Wallacea/Sahul region: New Guinean southeast
170 lowlands, New Guinean north and south lowlands, New Guinean east highlands, New
171 Guinean west highlands, northeast Australia, the Bismarck Archipelago, and Wallacea. Of
172 these seven regional groups of interest, we also characterized 58 subgroups based on the
173 fineSTRUCTURE dendrogram (Fig. 1A and Supplementary table 2). Each of these 58
174 subgroups defines a ‘population’ based on genetic similarity (Lawson and Falush 2012), a
175 definition that can differ from a linguistic or anthropological perspective but which is more
176 accurate regarding the analyses performed in our work.

177 As previously described (Bergstrom, et al. 2017; Hudjashov, et al. 2017), recent gene flow
178 from Asian groups, mostly linked to the Austronesian dispersal ~3kya, has substantially
179 impacted regional genetic diversity. Since our study focused on Upper Pleistocene
180 movements, we decided to mask this later Asian genetic ancestry using haplotypic
181 information with PCAdmix (Brisbin, et al. 2012). Principal Component Analysis (PCA) shows
182 that this procedure was efficient; prior to masking the Asian genetic ancestry, all genomes
183 from New Guinea, Wallacea, northeast Australia and the Bismarck Archipelago were spread
184 along the first principal component towards Asian genomes, while after masking, these
185 genomes instead clustered together (Supplementary Fig. 3). The dataset was also analyzed
186 using ADMIXTURE prior and after the masking of Asian tracks (Supplementary figure 4).
187 Asian genetic ancestry is present at an average of 22.7% in New Guinean lowlands, the
188 Bismarck Archipelago, Wallacea and northeast Australia. It ranges between 0% and 62.1% in

189 individuals from New Guinean lowlands, but absent in New Guinean highlanders, as
190 previously reported (Bergstrom, et al. 2017), and between 0% and 49.2% in individuals from
191 the Bismarck Archipelago, between 38.8% and 63.7% in individuals from Wallacea and
192 between 8.1% and 12.1% in individuals northeast Australia, as previously reported
193 (Hudjashov, et al. 2017; Mallick, et al. 2016) (Supplementary figure 4). After the masking, the
194 average Asian genetic ancestry is greatly reduced (Supplementary figure 4), which allows us
195 to infer demographic scenarios that are not strongly biased by recent Asian genetic
196 admixture.

197 We first described the masked dataset using a PCA focused on individuals in the 7 regional
198 groups of interest (Fig. 1B). All genomes from the same geographical regions are grouped
199 together in distinctive clusters. Clusters of New Guinean lowlanders, Wallacean islanders and
200 northeast Indigenous Australians fall relatively close together. New Guinean highlanders,
201 especially individuals from the New Guinean east highlands, are separated on the second
202 principal component. Individuals from the Bismarck Archipelago form two clusters separated
203 by both main components. This apparent genetic structure coincides with F_{ST} genetic
204 distances (Supplementary Fig. 5). The genetic diversity of the Bismarck Archipelago is
205 particularly differentiated from other New Guineans, Wallaceans and northeast Indigenous
206 Australians ($0.18 < F_{ST} < 0.29$; Supplementary Fig. 5). Within New Guinea, larger genetic
207 distances are found between New Guinean lowlanders and highlanders ($0.05 < F_{ST} < 0.18$;
208 Supplementary Fig. 5) than between any other New Guinean sub-groups, as previously
209 reported (Bergstrom, et al. 2017). Wallaceans are genetically closer to New Guineans
210 ($0.04 < F_{ST} < 0.26$) than to Bismarck Islanders or northeast Indigenous Australians
211 ($0.06 < F_{ST} < 0.318$). We further confirm this geographical structure using EEMS (Petkova, et al.
212 2016), which places the biggest genetic barriers between the Bismarck Archipelago and New
213 Guinea, between the New Guinean lowlands and highlands, as previously noted (Bergstrom,
214 et al. 2017), and between northeast Australia and New Guinea (Supplementary Fig. 6). We
215 do not detect any strong genetic barriers in Wallacea.

216 All of these analyses clearly show that there is strong geographical structure of human
217 genetic diversity in Sahul today. However, surprisingly, it also reveals a relative genetic
218 continuity between New Guinea and Wallacea, suggesting that Wallacea might have
219 represented less of a barrier to humans than other floral and faunal species.

220

221 *Identification of gene flow*

222 We analyzed the genetic diversity of our dataset to identify possible connections between
223 each individual in the seven regional groups from New Guinea, northeast Australia, the
224 Bismarck Archipelago and Wallacea using ADMIXTURE(Alexander, et al. 2009). The genetic
225 diversity in our dataset is best deconstructed into eight genetic components (Supplementary
226 Fig. 7-9), four of which are dominated by New Guinean, northeast Indigenous Australian,
227 Bismarck Archipelago and Wallacean groups (Fig.1C and Supplementary Fig.9). This strong
228 genetic structure is confirmed by the genetic ancestries defined in New Guinean east and
229 west highlanders as well as in the Bismarck Archipelago (red, orange and purple components
230 respectively). The fourth genetic component (yellow) is more widespread, as it is present in
231 individuals from Wallacea, the New Guinean lowlands, northeast Australia and the Bismarck
232 Archipelago. It is the main component in Wallacea and New Guinea, and reaches its highest
233 value in the southeast region (31-63%; Fig. 2A, Supplementary Fig.7 and 9).

234 To determine the history of regional mixing, population splits and admixture events were
235 estimated without *a priori* assumptions with TREEMIX (Pickrell and Pritchard 2012)(Fig. 1D
236 and Supplementary Fig. 10). An optimal model is obtained including five gene flows, one of
237 which is in the Wallacea/Sahul region: from ancestors of northeast Indigenous Australians to
238 a group ancestral to most Bismarck Archipelago islanders (Fig. 1D and Supplementary Fig.
239 10). This may partly explain the shared genetic ancestry observed in the ADMIXTURE plot
240 (purple, Fig. 1C).

241 The limited number of admixture events is confirmed by f_3 -statistics performed between all
242 pairs of the 58 sub-groups (Patterson, et al. 2012). Only seven genetic admixture events are
243 detected (Supplementary table 3), four of which show multiples waves of admixture. For all
244 of these gene flows, we estimated the dates of the admixture events using MALDER (Loh, et
245 al. 2013). Gene flows are detected between sub-groups of New Guinean west highlands,
246 mostly between 9.9-11.9kya (Supplementary table 3). Multiple gene flow events between
247 New Guinean east highlanders and New Guinean sub-groups from the south lowlands
248 occurred between 25.4-30.2kya and 2.6-4.1kya. During the same two time frames, we also
249 observe gene flows between a New Guinean sub-group from the south lowlands and a
250 Wallacean sub-group from the Lesser Sunda Islands (Supplementary table 3). Although
251 limited in number, these gene flows occurred at different time frames spanning several

252 millennia, indicative of a population dynamics model involving multiple periods of isolation
253 and contact.

254

255 *Modeling the genetic scenario of the migration in Wallacea*

256 Our analyses show that New Guinean groups appear to be more closely related to
257 populations in Wallacea than to groups in northeast Australia and the Bismarck Archipelago
258 (Fig.1D and Supplementary Fig.5). Since human populations settled Sahul from Sunda
259 through Wallacea (Allen and O'Connell 2020), we expected that genetic diversity from New
260 Guinea, northeast Australia and the Bismarck Archipelago would branch from the diversity of
261 Wallacean sub-groups. This highest genetic affinity between New Guineans and Wallaceans
262 could be due to ancient gene flow between these populations (Supplementary table 3), or to
263 a genetic divergence between northeast Indigenous Australians and Wallaceans prior to the
264 genetic divergence between New Guineans and Wallaceans (Fig. 1D). We therefore explored
265 different scenarios that could have led to the observed genetic patterns.

266 We first sought to identify the most likely route taken by the first settlers of Sahul. We used
267 an f_3 -outgroup approach to compare the genetic diversity present in New Guinean,
268 northeast Indigenous Australian and Bismarck Archipelago islander sub-groups to each
269 Wallacean sub-group, using African Yoruba as an outgroup (Supplementary Fig. 11). All sub-
270 groups from New Guinea, northeast Australia and the Bismarck Archipelago have f_3 -
271 statistics values equally distant to all Wallacean sub-groups located on the possible southern
272 route to Sahul (Lesser Sunda Islands and Flores Island)(Supplementary Fig. 11A-C). However,
273 when we included a sub-group located on the northern route (North Maluku islands, Kei
274 Island), we detected a stronger genetic link with all studied sub-groups (Supplementary Fig.
275 11D-F). Most populations from New Guinea and the Bismarck Archipelago are significantly
276 genetically closer to the population from North Maluku using f_4 -statistics
277 ($f_4(\text{Yoruba}, X; \text{Lesser Sunda Islands}, \text{North Maluku}) > 0.001$, $Z\text{-score} > 3$, Supplementary Fig. 12).
278 We found higher f_3 -statistics and more significant f_4 -statistics values for New Guinean sub-
279 groups with North Maluku than for sub-groups from the Bismarck Archipelago and northeast
280 Australia (Fig. 2A and Supplementary Fig. 12). This suggests that the ancestors of New
281 Guineans, Bismarck Archipelago islanders and northeast Indigenous Australians are all more
282 closely related to North Maluku islanders than to groups from Lesser Sunda islands or Flores
283 Island, and that this link is probably more important in New Guineans. This analysis supports

284 two non-exclusive hypotheses: (i) the northern route as the most likely path for the initial
285 settlement of Sahul, (ii) secondary gene flows between groups of North Maluku and New
286 Guinea, the Bismarck Archipelago and northeast Australia.

287 We modeled these hypotheses using qpGraph (Patterson, et al. 2012). For completeness, we
288 tested several models that consider migrations from Sunda to Sahul along either the
289 southern route (represented by the Lesser Sunda Islands) or the northern route
290 (represented by the North Maluku Islands) (Supplementary Fig. 13), as well as different
291 possibilities of admixture between Wallacean and New Guinean groups as suggested by our
292 other analyses (Fig. 1C, Supplementary Fig. 6, Supplementary table 3). Since our analyses
293 suggested that Wallaceans share a closer genetic link with New Guineans than northeast
294 Indigenous Australians (Fig. 1D and 2A, Supplementary Fig. 6), we decided to model
295 demographic scenarios with one or two migrations from Sunda to Sahul. Considering one
296 migration from Sunda to Sahul, meaning that both New Guineans and Indigenous Australians
297 share a common ancestor more recently than with Wallaceans, we obtained one model
298 fitting the data ($Z=-2.558$, Figure 2B and Supplementary Fig. 13). It describes a common
299 ancestry to both New Guineans and Indigenous Australians, after the split from Wallaceans,
300 followed by major secondary gene flows from New Guinea to Wallacea (between 61-79%,
301 Figure 2B and Supplementary Fig. 13). In this model no preferable route in Wallacea can be
302 distinguished. We then tested models considering two migrations from Sunda to Sahul, in
303 which northeast Indigenous Australians and New Guineans descended separately from
304 groups moving along the northern or southern routes (Supplementary Fig. 14). We obtained
305 one significant model: two migrations to Sahul, including one via the northern route, an
306 admixture event between the ancestors of New Guineans and northeast Indigenous
307 Australians, and a subsequent gene flow from ancestral groups from New Guinea to North
308 Maluku (52%, $Z = 2.337$, Supplementary Fig. 14). In this model the ancestral group to New
309 Guineans is related to the northern route, via North Maluku, but no route can be proposed
310 for the ancestral Indigenous Australian group. Our qpGraph analyses thus support both
311 models of migrations from Sunda to Sahul, with one or two migrations, followed by high
312 secondary gene flows from New Guinea to Wallacea.

313

314 *Modeling the genetic scenario of the settlement of Sahul*

315 We refined our models to include a group from the Bismarck Archipelago (represented by
316 New Britain), considering it either to be a sister group to New Guineans or the result of a
317 separate admixture event with ancestral northeast Indigenous Australians (Supplementary
318 Fig. 15), as suggested by other analyses (Fig. 1D). For both models, with one or two
319 migrations to Sahul, we obtained models fitting the data when the Bismarck Archipelago
320 group results from a separate admixture event between ancestral New Guineans and
321 northeast Indigenous Australians (One migration: $Z = -2.570$, Two migrations: $Z = 2.301$, Fig.
322 2B and Supplementary Fig. 15). This corresponds to the gene flow observed between an
323 ancestral Indigenous Australian group and ancestral Bismarck Archipelago group detected by
324 TREEMIX (Fig. 1D).

325 We added further detail to this scenario by reconstructing the effective population sizes of
326 these groups, and estimating divergence dates between all studied groups using
327 MSMC2 (Schiffels and Durbin 2014). Since archaic admixture can be a confounding factor in
328 these analyses (Mallick, et al. 2016), we masked both Neandertal and Denisovan genomic
329 tracks using ChromoPainter (Lawson, et al. 2012). The archaic genetic ancestry in the newly
330 sequenced genomes ranges between 1.8% and 2.8% of Denisovan ancestry and ranges
331 between 2.3% and 2.5% of Neandertal ancestry, corresponding to previous estimations
332 (Jacobs, et al. 2019; Vernot, et al. 2016). All estimated effective population size curves show
333 a bottleneck around 60-70 kya, as found for European and Asian populations
334 (Supplementary Fig. 16). Around 50kya, estimated effective population sizes for populations
335 from Wallacea and Sahul are between 1671-2100 individuals, and between 2540-2999
336 individuals in Eurasia and 9506 individuals in Africa. All divergence dates for Wallacea/Sahul
337 groups from Africa, represented by Yoruba genomes, show similar results (74 ± 4.2 kya,
338 Supplementary Fig. 17). We note here that this date is older than that obtained between
339 Eurasian and African genomes (66.5 ± 3.7 kya), a result previously reported and interpreted as
340 a potential methodological bias or the signal of an even earlier human migration from Africa
341 (Bergstrom, et al. 2017; Malaspinas, et al. 2016; Mallick, et al. 2016; Pagani, et al. 2016). We
342 cannot confirm any of these interpretations but the similar dates of cross-coalescence with
343 African genomes and the similar profiles of effective population size curves clearly show that
344 all genomes from Wallacea, New Guinea, northeast Australia and the Bismarck Archipelago
345 descend from a common ancestor, probably located in Sunda, as represented by our
346 TREEMIX analysis (Fig. 1D).

347 We then estimated divergence dates between each of the studied groups (Fig. 2C). Genomes
348 from New Guinean and Wallacean groups diverged between 23.7-26.7kya. Groups from New
349 Guinea and the Bismarck Archipelago show genetic divergence between 17.5-19.8kya.
350 Genomes of New Guineans diverged from the genomes of northeast Indigenous Australians
351 between 13.5-15.3kya. We found that genomes from northeast Indigenous Australians
352 diverged from the genomes of Wallaceans 13.3-15.2kya, and from genomes of Bismarck
353 Archipelago islanders between 15.2-17.3kya. We caution that these analyses are sensitive to
354 recent admixture events, and the masking of Asian genomic tracks reduces the amount of
355 data to accurately estimate divergence dates. It likely affected results on Wallacean
356 genomes for which the masked Asian genetic data were large (>50%). Moreover, the limited
357 number of Indigenous Australian genomes likely resulted in an insufficient phasing quality
358 for the two available genomes which could strongly affect MSMC2 analyses, as previously
359 noted (Malaspina, et al. 2016). A previous study on Indigenous Australians genomes
360 (Malaspina, et al. 2016)(not available for this study) found that all Indigenous Australians
361 groups diverged from each other around 25kya, according to MSMC2 analyses, and from
362 New Guinean highlanders around 30-40kya. Our estimation of the divergence between
363 Indigenous Australians and New Guinean is even more recent, potentially caused by the
364 identified gene flow between groups of northeast Australia and the Bismarck Archipelago
365 (Fig. 1D). Although these dates cannot be interpreted as being part of the initial migration to
366 Sahul, they strongly indicate long-term contacts between all of these groups, as seen in our
367 models. The population dynamics between ancestral groups was probably important, over
368 long distances and periods of time, leading to the emergence of modern populations like
369 New Guineans.

370

371 *Modeling the genetic scenario of the settlement of New Guinea*

372 Our TREEMIX analysis show that New Guinean sub-groups branch together. The root of the
373 tree lies in the genetic diversity present today in New Guinean southeast lowlanders (Fig.
374 1D). This result reflects the ADMIXTURE analysis, which shows that the main genetic
375 component in Wallacea is also the main component in southeast New Guinea, reinforcing
376 the hypothesis of an ancient link (yellow component, Fig. 1C and Supplementary Fig. 9).
377 Subsequently, sub-groups from the New Guinean north and south lowlands form two
378 different genetic clusters that are genetically drifted from the New Guinean southeast

379 lowlanders (Fig. 1D). This suggests that both clusters emerged from two separate migrations,
380 south and north, initiated from the southeast New Guinean region. We modeled the
381 settlement of New Guinea based on our results showing that most of New Guinean genetic
382 diversity could be described by genetic drift with limited occurrences of gene flows (Fig. 1D
383 and Supplementary table 3). We used qpGraph to determine the best scenario fitting data
384 from the five New Guinean groups (southeast lowlanders, south lowlanders, north
385 lowlanders, east highlanders and west highlanders) and one African outgroup (Yoruba). First
386 we modeled the most probable entry point into New Guinea, considering possible
387 migrations from the north lowlands, the south lowlands and the southeast lowlands, with or
388 without admixture events (Supplementary Fig. 18). One model is significant (Z-score = -0.06);
389 this places southeast lowlands as the most likely entry point but only if there is a secondary
390 admixture event with ancestral New Guinean north lowlanders (Supplementary Fig. 18). This
391 model corresponds with results from other analyses, which positions the New Guinean
392 southeast group at the root of the New Guinean diversity (Fig. 1D).

393 Highlanders are the last to branch off in our TREEMIX analysis (Fig. 1D). Our analyses showed
394 that New Guinean west and east highlanders form two clearly separated clusters (Fig. 1A and
395 1B, Supplementary Fig. 2 and 5), and that both New Guinean highlander groups drifted from
396 New Guinean lowlanders located in the southern region (Fig. 1D). We first added a New
397 Guinean east highlander group to model dispersal in the highlands from the north, south or
398 southeast lowlands; using qpGraph (Supplementary fig. 19). The model best fitting the data
399 connects New Guinean east highlanders to a branch leading to the New Guinean north
400 lowlanders (Z-score=-0.068, Supplementary Fig. 19). Finally we added a New Guinean west
401 highlander group, defined as a sister group to New Guinean east highlanders or coming from
402 a second migration from the lowlands (Supplementary Fig. 20). The best model places New
403 Guinean west highlanders with New Guinean south lowlanders (Z-score=2.759,
404 Supplementary Fig. 20). This branching describes two separate migrations leading to the
405 genetic diversity observed in New Guinean east and west highlanders, which would explain
406 the stark genetic difference observed in other analyses (Fig. 1B and 1C). However we note
407 here that it could appear to be in contrast to the TREEMIX analysis which showed the New
408 Guinean east highlanders branched with some PNG South lowlands groups (Fig.1D). In
409 particular these South lowlands groups (PNG South lowlands subgroups 39 and 40, Fig.1D)
410 are geographically located close to North lowlands groups. We computed an *f3*-outgroup

411 analysis to further explore the connection between PNG highlanders and lowlanders. It
412 confirmed the closer link between PNG West highlanders and PNG South lowlanders but also
413 showed that PNG East highlanders are almost genetically equidistant to both PNG North and
414 South lowlanders, although still more closely related to PNG South highlanders
415 (Supplementary figure 21). Based on our results, we postulate that PNG east highlanders
416 came from a place intermediate to both South and North lowlanders, potentially in the
417 current Morobe province. The qpGraph model also supports a secondary admixture event
418 between New Guinean west highlanders and north lowlanders (Fig. 3A and Supplementary
419 figure 20), which could explain the presence of the main New Guinean west highlanders
420 genetic component in many lowlanders (Fig. 1C) as well as previous reports on the link
421 between PNG north lowlanders and PNG highlanders (Bergstrom, et al. 2017). This final
422 model identifies the southeast lowlands as the entry point into New Guinea, followed by two
423 migrations to the south and north lowlands, which in turn led to the independent
424 settlements of the east and west highlands.

425 Given this scenario, we estimated divergence dates between each New Guinean group of
426 interest using MSMC2 (Fig. 3B). Divergence dates between New Guinean lowlanders and
427 New Guinean highlanders ranged between 8.8-12.8kya, which broadly correspond to
428 previous estimations (12-16kya)(Bergstrom, et al. 2017). West and east New Guinean
429 highlander groups diverged between 7.6-8.8kya, similarly to previously reported (6-
430 12kya)(Bergstrom, et al. 2017). These rather recent dates of genetic divergence probably
431 reflect long-term contacts between New Guinean groups following the initial settlement of
432 the island, initiated as early as 49kya based on archaeological records (Summerhayes, et al.
433 2010). They thus probably illustrate the gene flows between highlander groups, and
434 between highlanders and lowlanders, which are estimated to have occurred at the same
435 time (9.9-11.9kya, Supplementary table 3). Effective population sizes estimated with MSMC2
436 for each New Guinean group do not show any marked differences (Supplementary fig.16). It
437 was previously interpreted as a bias due to phasing quality (Bergstrom, et al. 2017), but it
438 could also mean that contacts between New Guinean highlanders and lowlanders were
439 frequent enough so that populations grew similarly.

440

441 **Discussion**

442 Our study provides the first models of the settlement of north Sahul and New Guinea based
443 on whole genome data (Fig. 4). Our models support an early split of the ancestral groups of
444 New Guineans from northeast Indigenous Australians, potentially in Wallacea. Among the
445 possible migratory routes, our model brings new evidence for the use of the northern route,
446 via the North Maluku islands. Our genetic data describe a settlement of New Guinea
447 initiated from what is now its southeast lowland via the Arafura Plain, and long-term
448 population contact between ancestral populations of New Guinea, northeast Australia,
449 Wallacea and the Bismarck Archipelago.

450

451 *Multiple migrations between Wallacea and Sahul*

452 To reach Sahul, the first human settlers had to cross Wallacea, a region always composed of
453 small islands, large water crossings and strong ecological barriers (Wallace 1869). These
454 restrictions on movement are reflected in the substantial genetic structuring of current
455 populations (Fig. 1 and Supplementary Fig. 5). But our results also show relative genetic
456 continuity between groups from Wallacea to New Guinea, with stronger genetic barriers
457 between New Guinea and Australia, as well as around the Bismarck Archipelago
458 (Supplementary Fig. 6), as described in a previous report (Bergstrom, et al. 2017). Wallacea
459 was probably less a strong geographical frontier than a place of possible migrations to Sahul
460 (Allen and O'Connell 2020; Birdsell 1977; Kealy, et al. 2017). The ancestral population to New
461 Guineans, Bismarck Archipelago Islanders and northeast Indigenous Australians might have
462 been composed of around 1300 individuals based on ecological information (Bradshaw, et al.
463 2019) and 1671-2100 individuals based on genetic data (Supplementary Fig.16).

464 Our models, like those of others (Malaspinas, et al. 2016; Teixeira and Cooper 2019), show
465 that this ancestral population split into different groups early during this migration, upon the
466 arrival in Sahul 50kya (model A, fig.3B and 4), or even before in Wallacea, from which
467 multiple migrations could have been conducted within a small time frame (i.e. in less than
468 900 years (Bradshaw, et al. 2019)) (model B, Fig.3B and 4). A model with a single migration is
469 still valid given our genetic data but there is growing non-genetic evidence that the initial
470 settlement of Sahul might have been made by multiple migrations (Bird, et al. 2019;
471 Bradshaw, et al. 2021; Bradshaw, et al. 2019; Clarkson, et al. 2017; Crabtree, et al. 2021).
472 This hypothesis has never been clearly explored using genomic data. Our study offers the
473 first opportunity to test this model, and we found that our genomic data support a model

474 with two migrations to Sahul to explain the current genetic diversity in New Guineans,
475 Bismarck Archipelago Islanders and Indigenous Australians (model B, Fig.3B and 4). Among
476 the possible routes from Wallacea to Sahul, this model describes the northern route through
477 North Maluku as a preferable path taken by the ancestors of New Guineans (Fig. 2A). It is
478 consistent with studies on paleogeological reconstructions of coastlines (Fig. 4)(Bird, et al.
479 2019; Kealy, et al. 2017) as well as studies using ecological models (Bradshaw, et al. 2021).
480 We cannot propose a location for the second route given the limited number of available
481 Indigenous Australian genomes, but it may have been different, and potentially used at
482 different times during the Upper Pleistocene, as was recently described (i.e. possible entry
483 though the northwest Sahul Shelf around 75kya (Bradshaw, et al. 2021)). Our genomic model
484 B also agrees with previous studies on uniparental markers that propose stark differences in
485 genetic diversity between New Guineans and northeast Indigenous Australians as a sign of
486 multiple migrations (Pedro, et al. 2020; Redd and Stoneking 1999). The genomic model B
487 thus brings together a consensus view that New Guineans and northeast Indigenous
488 Australians share a common ancestor, in Sunda, (Malaspinas, et al. 2016; Teixeira and
489 Cooper 2019), therefore integrating recent perspectives on multiple migrations to Sahul
490 (Allen and O'Connell 2020; Redd and Stoneking 1999). Future work on modern and ancient
491 DNA in Wallacea, New Guinea and Australia will surely help to disentangle the two models
492 presented here.

493 Wallacea was clearly not a major barrier to human migrations (Supplementary Fig. 6) but it
494 was also not necessarily a simple highway to Sahul (Fig. 4). We hypothesize that the
495 multitude of Wallacean islands might have driven genetic differentiation between related
496 groups. These human groups would have progressively migrated east, in a planned effort
497 spanning over a millennium or more (Bradshaw, et al. 2019). This time was probably
498 necessary for the populations to grow large enough before migrating further east
499 (Bradshaw, et al. 2021), leading to the emergence of genetic differentiation. Several human
500 populations could have crossed Wallacea to land on Sahul, using different routes (Bradshaw,
501 et al. 2021).

502 During this very early period of the settlement of Sahul, it is likely that the demographic
503 successes of these groups were different, as some archaeological sites might indicate
504 (Clarkson, et al. 2017). Admixture between ancestral populations, which countered the
505 potential effects of inbreeding and drift in isolated groups, might have been an important

506 factor for the dispersal success of certain human groups in Sahul. A previous study found
507 relatively late divergence dates between New Guinean highlanders and Indigenous
508 Australians (30-40kya), as well as between Indigenous Australians groups (25kya)
509 (Malaspinas, et al. 2016). It also found signals of post-divergence contacts between ancestral
510 groups of New Guineans and Indigenous Australians in the north of Australia (Malaspinas, et
511 al. 2016), which agrees with the late genomic divergence dates that we obtain between
512 southeast New Guinean lowlanders and northeast Indigenous Australians (13.5-15.3kya, Fig.
513 2C). These dates need to be interpreted with caution, because of technical issues raised
514 above and by others (Malaspinas, et al. 2016), but they do show that Sahul, and especially
515 the north region, was probably a place of interactions between human groups with related,
516 but already differentiated, genomes. These interactions probably involved few individuals
517 given the number of region-specific mitochondrial DNA (Pedro, et al. 2020) and Y
518 chromosome lineages (Bergstrom, et al. 2016) shared between New Guineans and
519 Indigenous Australians. This led to the emergence of the unique genetic diversities passed
520 on to the modern populations of northeast Australia, the Bismarck Archipelago and New
521 Guinea.

522

523 *The complex settlement of New Guinea*

524 New Guinea was settled rapidly after the initial migrations to Sahul as revealed by
525 archaeological records (Summerhayes, et al. 2010). This fast dispersal has made it
526 challenging to determine the route taken to reach this territory. Our study strongly supports
527 an entry point located in what is now the New Guinean southeast region (Fig. 4). The genetic
528 distances between current New Guinean southeast lowlanders and Wallaceans are smaller
529 than with any other New Guineans (Supplementary Fig. 5). The main genetic component in
530 Wallaceans is also mostly present in groups from the New Guinean southeast lowlands (Fig.
531 1C and Supplementary Fig. 9). Among the different models tested, the only statistically
532 significant scenario places New Guinean southeast lowlanders at the root of New Guinean
533 genetic diversity, even considering secondary gene flows (Fig. 3A). This remarkably
534 corresponds to the location of the oldest known archaeological sites, in the Ivané valley,
535 dated at 45-49kya (Summerhayes, et al. 2010). Our genetic model located the, apparently
536 ancient, genetic background to the far east of New Guinea bringing new support for an initial
537 migration via the ancient plain currently under the Arafura Sea (Allen and O'Connell 2020).

538 Recent studies using ecological modeling identified the most likely route to New Guinea in
539 the ancient plain currently under the Arafura Sea along the southeast coast of New Guinea
540 (Bradshaw, et al. 2021; Crabtree, et al. 2021), converging with our genetic model. Other
541 hypothetical routes through the New Guinean north coast, the highlands or the south
542 foothills could have been taken by other groups of settlers (Summerhayes, et al. 2017) but
543 their demographic success, based on genetic data, must have been limited. During the Upper
544 Pleistocene, the Arafura Plain south of modern New Guinea was dominated by savannah and
545 lakes (Summerhayes, et al. 2017). It would have been a favorable ecosystem for human
546 dispersal from the west coast to the east coast of Sahul, before reaching what is now the
547 southeast coast of the island of New Guinea (Fig. 4).

548 Settlement within New Guinea was initiated from the lowlands. We found that the genetic
549 diversity of groups from the New Guinean north and south lowlands derived separately from
550 groups in the southeast region (Fig. 1D and 3A). During the first stages of the settlement of
551 New Guinea, the lowlands were probably the most hospitable places due to their access to
552 littoral resources (Summerhayes, et al. 2017). However, the dry conditions and steep
553 geography might have created locations unfavorable to permanent human settlements,
554 potentially numerous in the highlands and the north lowlands (Terrell 2004). Around 35kya,
555 these adverse conditions appeared to have significantly reduced human mobility in New
556 Guinea (Allen and O'Connell 2020). This would explain the relative slow population growth
557 we observe at this period for all New Guinean groups, in comparison to European or African
558 groups (Supplementary Fig. 16). A similar explanation was put forward to interpret the late
559 coalescence dates of many New Guinean maternal lineages, postdating 30kya (Pedro, et al.
560 2020). This lack of resources could have induced New Guinean populations to explore other
561 locations, such as the highlands (Summerhayes, et al. 2017).

562 Our analyses show that the New Guinean highlands were settled by two separate migratory
563 waves: one migration from the east foothills to the east highlands, and another migration
564 from the south lowlands to the west highlands (Fig. 1D, 3A and 4). These two migrations
565 could have used two major geological corridors defined by the Markham and Ramu Rivers in
566 the east, and the Fly River in the south (Summerhayes, et al. 2017). Although we do not find
567 support for the Sepik River as a major migratory route for the initial settlement of the
568 highlands (Bergstrom, et al. 2017), all these natural corridors were likely a major path for
569 secondary gene flows between New Guinean highlanders and New Guinean lowlanders, as

570 indicated by our analyses (Fig. 3A). Many gene flows appear to have occurred during the
571 early period of the settlement of the highlands, around 25.4-30.2kya, suggesting continuous
572 contacts between highlanders and lowlanders (Fig. 3A and Supplementary table 3). This is
573 consistent with archaeological data arguing for long-term, but not permanent, settlements
574 in the highlands around 25kya (Summerhayes, et al. 2017). New Guinean populations
575 probably lived in semi-permanent settlements, gathering resources in the highlands and the
576 foothills, but keeping contacts with lowlander groups. These populations would have
577 progressively fixed their settlement in the highlands, reducing contacts with the lowlanders,
578 since no gene flow was detected between 20kya and 3kya, and increasing the genetic
579 differentiation between highlanders and lowlanders (Fig. 1 and 3, Supplementary Fig. 5), as
580 previously reported (Bergstrom, et al. 2017).

581 However an intense period of population dynamics occurred locally in the New Guinean
582 highlands around 10kya. We detected several admixture events between west highlander
583 groups at this time (Supplementary table 3). This period was marked by the development of
584 agricultural practices and the emergence of a Neolithic culture, especially in the Waghi valley
585 of the west highlands (Denham, et al. 2003; Shaw, et al. 2020). As previously reported
586 (Bergstrom, et al. 2017), this was a major cultural change that likely influenced gene flows in
587 the highlands and the dispersal of cultural traits, such as the Trans-New Guinean linguistic
588 family (Reesink, et al. 2009). Lowlanders might not have been significantly influenced by the
589 development of a Neolithic culture in the highlands (Pawley 2005), but their population size
590 was not necessarily much lower than those found in the highlands (Supplementary fig.16)
591 thanks to exchanges of goods and genes between the two regions (Supplementary table 3).
592 A second wave of gene flows between highlanders and lowlanders is only detected around
593 2.6-4.1kya (Supplementary table 3). At the same period, gene flow is also detected between
594 New Guinean lowlanders and Wallaceans (Supplementary table 3). These events likely
595 correspond to the influence of the Austronesian culture 3kya (Lipson, et al. 2014), but rather
596 than initiating new contacts between New Guinean groups, the arrival of Austronesians
597 seems to have only reignited New Guinean population dynamics that already existed around
598 25kya (Supplementary table 3). Genes, material culture and ideas were already moving
599 within New Guinea (Pawley 2005; Torrence and Swadling 2008). The last stage of the
600 settlement of New Guinea was probably another pulse in a network of population
601 interactions built up over millennia.

602

603 New Guinean genomes reveal possible scenarios describing the migrations from Sunda to
604 Sahul and a settlement of New Guinea initiated in the southeast lowlands, followed by two
605 separate dispersals to the highlands (Fig. 4). Our models will necessarily be confronted by
606 future genetic and archaeological data, particularly in the western half of New Guinea Island
607 and Australia. Paleogenomics in Oceania is an emerging field and it has already proven to be
608 highly informative about the history of populations with New Guinean genetic ancestry
609 (Choin, et al. 2021; Skoglund, et al. 2016). Future research will certainly bring key
610 information on the complex population dynamics that existed during the past 50,000 years
611 in north Sahul.

612

613 **Materials and Methods**

614 *Sampling*

615 Fifty eight individuals from Papua New Guinea participated in the study (Supplementary
616 Figure 1 and Supplementary table1). These new samples were obtained in Papua New
617 Guinea between 2016 and 2019. The study was approved by the Medical Research Advisory
618 Committee of Papua New Guinea (National Department of Health) under research ethics
619 clearance MRAC 16.21 and by the French Ethics Committees (Committees of Protection of
620 Persons IE-2015-837 (1)). Permission to conduct research in Papua New Guinea was granted
621 by the National Research Institute of Papua New Guinea (permit 99902292358), with full
622 support from the School of Humanities and Social Sciences, University of Papua New Guinea.
623 All samples were collected from healthy unrelated adult donors, all of whom provided
624 written informed consent (Supplementary text). DNA was extracted from saliva samples with
625 the Oragene sampling kit according to the manufacturer's instructions. After a full
626 presentation of the project to a wide audience, a discussion with each individual willing to
627 participate ensured that the project was fully understood. Once participants had signed
628 informed-consent forms, we interviewed them to obtain information on their date and place
629 of birth, their spoken language(s), and similar data related to their genealogy (up to 2nd or
630 3rd generation) to establish local ancestry. Based on the participants' genealogy, we chose
631 the samples to be sequenced so as to cover most of Papua New Guinean genomic diversity.

632

633 *Datasets*

634 Sequencing libraries were prepared using the TruSeq DNA PCR-Free HT kit. 150 bp paired-
635 end sequencing was performed on the Illumina HiSeq X5 sequencer. All raw data are
636 available on the European Genome-Phenome data repository (EGA;
637 <https://www.ebi.ac.uk/ega/>): EGAS00001005393. The 58 new Papuan genomes were
638 combined with raw reads from the Simons Genome Diversity Project (SGDP, n=292, EBI
639 European Nucleotide Archive: PRJEB9586 and ERP010710) (Mallick, et al. 2016); 60 Papuan
640 genomes (EGA: EGAS00001001247 dbGAP: phs001085.v1.p1) (Malaspinas, et al. 2016;
641 Vernot, et al. 2016) and 161 genomes from Indonesia (EGA: EGAS00001003054)(Jacobs, et
642 al. 2019). SNP calling was performed on the combined dataset. Reads were aligned to the
643 'decoy' version of the GRCh37 human reference sequence (hs37d5) using BWA MEM (Li
644 2014). Duplicated reads were removed with picard-tools v.2.12.0
645 (<http://broadinstitute.github.io/picard>) and local realignment around indels were performed
646 with GATK v. 3.5 (Poplin, et al. 2018). After alignment, the average sequencing depth across
647 the samples was 36x, with a minimum of 31x coverage. Base-calling and per-sample gVCF
648 files were generated using GATK HaplotypeCaller (reads mapping quality ≥ 20) (Poplin, et al.
649 2018). Multisample VCF files were obtained with CombineGVCFs and GenotypeGVCFs.
650 Genotype calls were subsequently filtered for base depth ($\geq 8x$ and $\leq 400x$) and genotyping
651 quality (≥ 30) using BCFtools v. 1.4 (Li 2014). We removed sites within segmental
652 duplications, repeats and low complexity regions based on available masks
653 (<http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/genomicSuperDups.txt.gz>;
654 [http://software.broadinstitute.org/software/genomestrip/node ReferenceMetadata.html](http://software.broadinstitute.org/software/genomestrip/node_ReferenceMetadata.html)).
655 Only biallelic sites with high quality variant calls in at least 99% of samples were retained,
656 leading to a whole genome dataset of 571 individuals typed for 41,250,479 SNPs
657 (Supplementary table 2). All whole genomes were phased using SHAPEIT v. 2.r79033
658 (Delaneau, et al. 2012) with default settings, except for iterations (50) and states (200) and
659 used the HapMap phase 2 genetic map (International HapMap, et al. 2007) without a
660 reference panel.
661 In addition to the whole genome dataset, we compiled another dataset including genotyping
662 data from Papua New Guinea (Bergstrom, et al. 2017). This genotyping data was phased and
663 imputed using all phased New Guinean whole genomes as a reference panel (N=166) with
664 IMPUTE2 v. 2.3.2 (Howie, et al. 2009). Ten Papuan WGS randomly selected were
665 downgraded to the same set of SNPs as the genotyping data to serve as an internal control.

666 All SNPs with an imputation score above 0.4 were kept, leaving 12,168,294 SNPs. The 10
667 downgraded genomes are over 98% consistent with actual genotypes from whole genomes
668 for the same individuals. We also added genotyping data from North Maluku (Kusuma, et al.
669 2017), but we did not impute missing SNPs since no whole genomes are available for this
670 group. Finally we included data from Neandertal (Prufer, et al. 2014), Denisovan (Reich, et al.
671 2010) and Chimpanzee (Mikkelsen, et al. 2005). The final dataset of 966 genomes was
672 phased with SHAPEIT v. 2.r79033 (Delaneau, et al. 2012) without a reference panel with the
673 same settings as above.

674

675 To focus on Upper Pleistocene genetic history, recent Asian ancestry was masked in admixed
676 individuals from New Guinea, Wallacea, northeast Australia and the Bismarck Archipelago
677 using PCAdmix v.1.0 (Brisbin, et al. 2012). Since both available Indigenous Australian
678 genomes showed signs of recent European admixture (Mallick, et al. 2016), we also masked
679 European genetic ancestry for all individuals. Two parental metapopulations were defined by
680 randomly choosing 100 Asian-European individuals and 100 New Guinean individuals. For
681 the latter group, only unadmixed New Guinean whole genome sequences were selected, as
682 defined by ADMIXTURE v. 1.3 (Alexander, et al. 2009). Local ancestry was estimated for all
683 individuals from New Guinea, Wallacea, northeast Australia and the Bismarck Archipelago
684 showing non-Eurasian genetic ancestry above 20%. Posterior probabilities for New Guinean
685 ancestry above 0.9 and Asian ancestry below 0.1 were criteria to define SNPs with New
686 Guinean ancestry. We applied quality controls using Plink v. 1.932 (Chang, et al. 2015) to
687 filter for and exclude (1) close relatives by using an identity-by-descent (IBD) estimation with
688 an upper threshold of 0.25 (second-degree relatives); (2) SNPs that failed the Hardy-
689 Weinberg exact test ($p < 10^{-6}$) within each group; (3) samples with a call rate < 0.99 and
690 displaying missing rates > 0.05 across all samples in each population. For analyses requiring
691 linkage disequilibrium pruning, we used Plink to filter for variants in high linkage
692 disequilibrium ($r^2 > 0.2$), leaving 74,067 SNPs. Some Wallacean groups were pooled together
693 because of the high rate of masked SNPs, according to the fineSTRUCTURE analysis and
694 geography, in order to perform population-based analyses such as TREEMIX, f3 (ex:
695 Wallacea_3-4_6-7 refers to the pool of data from the subgroups Wallacea_1, Wallacea_2,
696 Wallacea_6, Wallacea_7).

697

698 *Analyses*

699 Principal Components Analyses were run with smartpca from the EIGENSOFT v.7.2.1 package
700 (Patterson, et al. 2006). Populations were genetically defined using fineSTRUCTURE v. 2.07
701 (Lawson, et al. 2012). This package performs model-based Bayesian clustering of genotypes
702 based on the shared IBD fragments between each pair of individuals, without self-copying,
703 calculated with ChromoPainter v. 2.0 (Lawson, et al. 2012). Five million MCMC runs were
704 performed. A coancestry heatmap and a dendrogram were inferred to visualize the number
705 of statistically defined clusters that describe the data (Supplementary Fig. 2). 15 genetic
706 groups were defined, and subdivided into 106 sub-groups, including 58 sub-groups from
707 Wallacea and Sahul (Supplementary Fig. 2 and Supplementary table 2). Genotyping data
708 from North Maluku were defined as an additional subgroup (Wallacea_11) since the SNP
709 density is much lower than whole genome sequencing data and no imputation could be
710 performed.

711 Genetic barriers and corridors were defined using EEMS v. 1 (Petkova, et al. 2016). We used
712 approximate geographic coordinates, a genetic dissimilarity matrix between populations and
713 a map of New Guinea and the surrounding areas defining a grid of 700 modeled demes.
714 Depending on their location, several sub-groups can be included in one deme, whose size
715 increased accordingly. We ran 5 million MCMC iterations before checking for convergence of
716 the MCMC chain (Supplementary Fig. 6). Plots were generated in R v. 3.4.3 according to
717 instructions in the EEMS manual. F_{ST} genetic distances were obtained with the EIGENSOFT
718 v.7.2.1 package (Patterson, et al. 2006).

719 Genetic ancestries present in the dataset were estimated by ADMIXTURE v. 1.3 (Alexander,
720 et al. 2009), with default settings, for components $K = 2$ to $K = 15$. Ten iterations with
721 randomized seeds were run and compiled with CLUMPAK v. 1 (Kopelman, et al. 2015). We
722 used the minimum average cross-validation value to define the most informative K
723 components (here, $K = 8$), and the major modes defined by CLUMPAK v. 1 were reported.
724 Geographical distributions of the main genetic components found in Wallacea and Sahul
725 were obtained using inverse distance weighting interpolation between data points with QGIS
726 v. 3.16.1 (GDAL-SOFTWARE-SUITE 2013).

727 f_3 -statistics and f_3 -outgroup were computed with qp3pop from the ADMIXTOOLS v.6.0
728 package (Patterson, et al. 2012) between all trios of populations. f_4 -statistics was performed
729 with qp4pop from the ADMIXTOOLS v.6.0 package (Patterson, et al. 2012). Admixture dates

730 were estimated with MALDER v. 1.0 for all negatively significant scenarios from the f_3 -
731 statistics analysis (Loh, et al. 2013). All dates mentioned in the manuscript are converted into
732 years based on 30 years per generation. TREEMIX v. 1.12 analysis (Pickrell and Pritchard
733 2012) was performed by setting blocks to 500 SNPs, accounting for linkage disequilibrium,
734 and migration edges were added sequentially until the model explained 99% of the variance,
735 rooting the tree with Chimpanzee (panTro5).

736 Scenarios of the settlement of Sahul were tested for the full dataset using qpgraph from the
737 ADMIXTOOLS v.6.0 package (Patterson, et al. 2012). One subgroup was chosen as
738 representative population for a given regional cluster (as defined in Supplementary table 2),
739 based on the number of individuals, the number of SNPs after masking for Asian SNPs and
740 their location. Populations were added one-by-one, testing different scenarios with or
741 without admixture events, following recommendations in (Lipson and Reich 2017). Models
742 with a $|Z\text{-score}| < 3$ were defined as fitting the data. Over one hundred different models
743 were tested. Only models fitting the data and scenarios based on major hypotheses were
744 represented. Results were visualized with GraphViz v. 2.46.0 (GraphViz). Population effective
745 sizes and population divergence dates were estimated with MSMC2 v. 2.1.2 (Schiffels and
746 Durbin 2014). Two individuals from each population were randomly chosen. Sample-specific
747 masks were generated from the corresponding BAM files using bamCaller.py from MSMC-
748 Tools packages (Schiffels and Wang 2020). Individual masks for Asian admixture were
749 generated from previously mentioned analyses with PCAdmix v.1.0 (Brisbin, et al. 2012). For
750 all analyzed individual, the mask of Asian genetic tracks ranged between 0% and 49% of the
751 total genome. The individual archaic masks were generated with ChromoPainter v.2.0
752 (Lawson, et al. 2012). Two metapopulations were defined: one metapopulation from 100
753 randomly selected Sub-Saharan Africans; and one metapopulation including Neandertal,
754 Denisovan and Chimpanzee genomes. Only stretches of more than 4 SNPs 'painted' by the
755 Sub-Saharan metapopulations were kept. A fourth individual mask was added based on
756 human genome mappability hs37d5 (Schiffels and Durbin 2014). MSMC2 was run with a
757 pattern of fixed time segments (1*5+40*1+5*3+3*5) and 25 iterations. Ambiguously phased
758 SNPs were skipped using the $-s$ flag. We used $1.25e-8$ as the mutation rate.

759

760 **Supplementary Materials**

761 Supplementary data are available at Molecular Biology and Evolution online.

762

763 **Acknowledgments**

764 We thank Georgia Kaipu (National Research Institute, Papua New Guinea), Christopher Kinipi
765 (University of Papua New Guinea Clinic), Alois Kuaso and Kenneth Miamba (National
766 Museum and Art Gallery, Papua New Guinea) for their help during the sampling campaigns.
767 We thank Chris Clarkson for his comments on the manuscript. We thank J. Stephen Lansing
768 (Santa Fe Institute) for financially supporting some of the sequencing. We thank Monika
769 Karmin for her help on the analysis of Y chromosome data. We thank Opale Coutant
770 (University of Toulouse) and Adeline Morez (Liverpool University) for their help in data
771 analysis. We acknowledge support from the GenoToul bioinformatics facility of Genopole
772 Toulouse Midi-Pyrénées, France; the LabEx TULIP, France; and the Centre National de
773 Recherche en Génomique Humaine (CNRGH), Evry, France. We especially thank all of our
774 study participants. We acknowledge support from the National Geographic Society (Grant
775 HJ-156R-17) and the Leakey Foundation. This research was supported by the European Union
776 through Horizon 2020 research and innovation programme under grant no 810645 and through the
777 European Regional Development Fund project no. MOBEC008. This work was supported by the
778 French Ministry of Research grant Agence Nationale de la Recherche (ANR PAPUAEVOL), the
779 French Ministry of Foreign and European Affairs (French Prehistoric Mission in Papua New
780 Guinea), and the French Embassy in Papua New Guinea, as well as a fellowship from the
781 Alexander von Humboldt Foundation.

782

783 **Author contributions**

784 NB, WP, JM, ML, FXR organised the sampling campaigns. NB, RT, JK, KS, TB, ML, FXR
785 collected biological samples. CB, AB, JFD generated whole genome sequences. LS, NB
786 processed raw whole genome sequences. NB, MM, MPC, ML, FXR designed the study. NB
787 performed the analyses and wrote the paper based on the input from all the other authors.
788 All authors participated in the interpretation of the results and approved the submission.

789

790 **Competing interest Declaration**

791 The authors declare no competing interests.

792

793 **Data and Code Availability Statement**

794 All raw data are available on the European Genome-Phenome data repository:
795 EGAS00001005393.

796

797 **References**

- 798 Alexander DH, Novembre J, Lange K 2009. Fast model-based estimation of ancestry in
799 unrelated individuals. *Genome Research* 19: 1655-1664. doi: 10.1101/gr.094052.109
- 800 Allen J, O'Connell JF 2020. A different paradigm for the initial colonisation of Sahul.
801 *Archaeology in Oceania* 55: 1-14.
- 802 Bellwood P. 2007. *Prehistory of the Indo-Malaysian Archipelago*. Canberra, Australia: ANU E
803 Press.
- 804 Bergstrom A, Nagle N, Chen Y, McCarthy S, Pollard MO, Ayub Q, Wilcox S, Wilcox L, van
805 Oorschot RA, McAllister P, Williams L, Xue Y, Mitchell RJ, Tyler-Smith C 2016. Deep Roots
806 for Aboriginal Australian Y Chromosomes. *Curr Biol* 26: 809-813. doi:
807 10.1016/j.cub.2016.01.028
- 808 Bergstrom A, Oppenheimer SJ, Mentzer AJ, Auckland K, Robson K, Attenborough R, Alpers
809 MP, Koki G, Pomat W, Siba P, Xue Y, Sandhu MS, Tyler-Smith C 2017. A Neolithic
810 expansion, but strong genetic structure, in the independent history of New Guinea.
811 *Science* 357: 1160-1163. doi: 10.1126/science.aan3842
- 812 Bird MI, Condie SA, O'Connor S, O'Grady D, Reepmeyer C, Ulm S, Zega M, Saltre F, Bradshaw
813 CJA 2019. Early human settlement of Sahul was not an accident. *Sci Rep* 9: 8220. doi:
814 10.1038/s41598-019-42946-9
- 815 Birdsell JB 1977. *The recalibration of a paradigm for the first peopling of Greater Australia.*
816 *Sunda and Sahul: Prehistoric Studies in Southeast Asia, Melanesia, and Australia.:* 113-
817 167.
- 818 Bradshaw CJ, Norman K, Ulm S, Williams AN, Clarkson C, Chadœuf J, Lin SC, Jacobs Z, Roberts
819 RG, Bird MI 2021. Stochastic models support rapid peopling of Late Pleistocene Sahul.
820 *Nature Communications* 12: 1-11.
- 821 Bradshaw CJA, Ulm S, Williams AN, Bird MI, Roberts RG, Jacobs Z, Laviano F, Weyrich LS,
822 Friedrich T, Norman K, Saltre F 2019. Minimum founding populations for the first peopling
823 of Sahul. *Nat Ecol Evol* 3: 1057-1063. doi: 10.1038/s41559-019-0902-6
- 824 Brisbin A, Bryc K, Byrnes J, Zakharia F, Omberg L, Degenhardt J, Reynolds A, Ostrer H, Mezey
825 JG, Bustamante CD 2012. PCAdmix: Principal components-based assignment of ancestry

826 along each chromosome in individuals with admixed ancestry from two or more
827 populations. *Human Biology* 84: 343-364. doi: 10.3378/027.084.0401

828 Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ 2015. Second-generation
829 PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4: 7. doi:
830 10.1186/s13742-015-0047-8

831 Choin J, Mendoza-Revilla J, Arauna LR, Cuadros-Espinoza S, Cassar O, Larena M, Ko AM-S,
832 Harmant C, Laurent R, Verdu P 2021. Genomic insights into population history and
833 biological adaptation in Oceania. *Nature*: 1-7.

834 Clarkson C, Jacobs Z, Marwick B, Fullagar R, Wallis L, Smith M, Roberts RG, Hayes E, Lowe K,
835 Carah X, Florin SA, McNeil J, Cox D, Arnold LJ, Hua Q, Huntley J, Brand HEA, Manne T,
836 Fairbairn A, Shulmeister J, Lyle L, Salinas M, Page M, Connell K, Park G, Norman K, Murphy
837 T, Pardoe C 2017. Human occupation of northern Australia by 65,000 years ago. *Nature*
838 547: 306-310. doi: 10.1038/nature22968

839 Collier M 2009. SahulTime: rethinking archaeological representation in the digital age.
840 *Archaeologies* 5: 110-123.

841 Crabtree SA, White DA, Bradshaw CJ, Saltr e F, Williams AN, Beaman RJ, Bird MI, Ulm S 2021.
842 Landscape rules predict optimal superhighways for the first peopling of Sahul. *Nature*
843 *Human Behaviour*: 1-11.

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

846 Denham TP, Haberle SG, Lentfer C, Fullagar R, Field J, Therin M, Porch N, Winsborough B
847 2003. Origins of agriculture at Kuk Swamp in the highlands of New Guinea. *Science* 301:
848 189-193. doi: 10.1126/science.1085255

849 GDAL-SOFTWARE-SUITE. 2013. QGIS.

850 Golson J, Denham T, Hughes P, Swadling P, Muke J. 2017. Ten thousand years of cultivation
851 at Kuk Swamp in the highlands of Papua New Guinea.

852 GraphViz. <http://graphviz.org>.

853 Greenhill SJ 2015. TransNewGuinea.org: An Online Database of New Guinea Languages. *PLoS*
854 *One* 10: e0141563. doi: 10.1371/journal.pone.0141563

855 Howie BN, Donnelly P, Marchini J 2009. A flexible and accurate genotype imputation method
856 for the next generation of genome-wide association studies. *PLoS Genet* 5: e1000529.
857 doi: 10.1371/journal.pgen.1000529

858 Hudjashov G, Karafet TM, Lawson DJ, Downey S, Savina O, Sudoyo H, Lansing JS, Hammer
859 MF, Cox MP 2017. Complex Patterns of Admixture across the Indonesian Archipelago. *Mol*
860 *Biol Evol* 34: 2439-2452. doi: 10.1093/molbev/msx196

861 International HapMap C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA,
862 Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F,
863 Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu
864 J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo
865 A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M,
866 Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y,
867 Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK,
868 Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS,
869 Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C,
870 Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller
871 P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T,
872 Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP,
873 Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE,
874 Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll
875 S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF,
876 Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson
877 GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro
878 HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S,
879 Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G,
880 Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang
881 H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo
882 CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M,
883 Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton
884 EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM,
885 Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ,
886 Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M,
887 Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L,
888 Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li
889 Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL,

890 Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J 2007.
891 A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851-
892 861. doi: 10.1038/nature06258

893 Jacobs GS, Hudjashov G, Saag L, Kusuma P, Darusallam CC, Lawson DJ, Mondal M, Pagani L,
894 Ricaut FX, Stoneking M, Metspalu M, Sudoyo H, Lansing JS, Cox MP 2019. Multiple Deeply
895 Divergent Denisovan Ancestries in Papuans. *Cell* 177: 1010-1021 e1032. doi:
896 10.1016/j.cell.2019.02.035

897 Karafet TM, Hallmark B, Cox MP, Sudoyo H, Downey S, Lansing JS, Hammer MF 2010. Major
898 east–west division underlies Y chromosome stratification across Indonesia. *Molecular*
899 *Biology and Evolution* 27: 1833-1844. doi: 10.1093/molbev/msq063

900 Kealy S, Louys J, O'Connor S 2018. Least-cost pathway models indicate northern human
901 dispersal from Sunda to Sahul. *J Hum Evol* 125: 59-70. doi: 10.1016/j.jhevol.2018.10.003

902 Kealy S, Louys J, O'Connor S 2017. Reconstructing Palaeogeography and Inter-island Visibility
903 in the Wallacean Archipelago During the Likely Period of Sahul Colonization, 65-
904 45000Years Ago. *Archaeological Prospection* 24: 259-272. doi: 10.1002/arp.1570

905 Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I 2015. Clumpak: a program
906 for identifying clustering modes and packaging population structure inferences across K.
907 *Molecular ecology resources* 15: 1179-1191. doi: 10.1111/1755-0998.12387

908 Kusuma P, Brucato N, Cox MP, Letellier T, Manan A, Grangé P, Sudoyo H, Ricaut FX 2017. The
909 last sea nomads of the Indonesian archipelago: genomic origins and dispersal. *European*
910 *Journal of Human Genetics*. doi: doi: 10.1038/ejhg.2017.88. [Epub ahead of print]

911 Lawson DJ, Falush D 2012. Population identification using genetic data. *Annu Rev Genomics*
912 *Hum Genet* 13.

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

915 Li H 2014. Toward better understanding of artifacts in variant calling from high-coverage
916 samples. *Bioinformatics* 30: 2843-2851. doi: 10.1093/bioinformatics/btu356

917 Lipson M, Loh P-R, Patterson N, Moorjani P, Ko Y-C, Stoneking M, Berger B, Reich D 2014.
918 Reconstructing Austronesian population history in Island Southeast Asia. *bioRxiv*. doi:
919 10.1101/005603

920 Lipson M, Reich D 2017. A Working Model of the Deep Relationships of Diverse Modern
921 Human Genetic Lineages Outside of Africa. *Mol Biol Evol* 34: 889-902. doi:
922 10.1093/molbev/msw293

923 Loh PR, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, Berger B 2013. Inferring
924 admixture histories of human populations using linkage disequilibrium. *Genetics* 193:
925 1233-1254. doi: 10.1534/genetics.112.147330

926 Malaspina AS, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, Bergstrom A, Athanasiadis
927 G, Cheng JY, Crawford JE, Heupink TH, Macholdt E, Peischl S, Rasmussen S, Schiffels S,
928 Subramanian S, Wright JL, Albrechtsen A, Barbieri C, Dupanloup I, Eriksson A, Margaryan
929 A, Moltke I, Pugach I, Korneliussen TS, Levkivskyi IP, Moreno-Mayar JV, Ni S, Racimo F,
930 Sikora M, Xue Y, Aghakhanian FA, Brucato N, Brunak S, Campos PF, Clark W, Ellingvag S,
931 Fourmile G, Gerbault P, Injie D, Koki G, Leavesley M, Logan B, Lynch A, Matisoo-Smith EA,
932 McAllister PJ, Mentzer AJ, Metspalu M, Migliano AB, Murgha L, Phipps ME, Pomat W,
933 Reynolds D, Ricaut FX, Siba P, Thomas MG, Wales T, Wall CM, Oppenheimer SJ, Tyler-
934 Smith C, Durbin R, Dortch J, Manica A, Schierup MH, Foley RA, Lahr MM, Bownern C, Wall
935 JD, Mailund T, Stoneking M, Nielsen R, Sandhu MS, Excoffier L, Lambert DM, Willerslev E
936 2016. A genomic history of Aboriginal Australia. *Nature* 538: 207-214. doi:
937 10.1038/nature18299

938 Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N,
939 Nordenfelt S, Tandon A, Skoglund P, Lazaridis I, Sankararaman S, Fu Q, Rohland N, Renaud
940 G, Erlich Y, Willems T, Gallo C, Spence JP, Song YS, Poletti G, Balloux F, van Driem G, de
941 Knijff P, Romero IG, Jha AR, Behar DM, Bravi CM, Capelli C, Hervig T, Moreno-Estrada A,
942 Posukh OL, Balanovska E, Balanovsky O, Karachanak-Yankova S, Sahakyan H, Toncheva D,
943 Yepiskoposyan L, Tyler-Smith C, Xue Y, Abdullah MS, Ruiz-Linares A, Beall CM, Di Rienzo A,
944 Jeong C, Starikovskaya EB, Metspalu E, Parik J, Villems R, Henn BM, Hodoglugil U, Mahley
945 R, Sajantila A, Stamatoyannopoulos G, Wee JT, Khusainova R, Khusnutdinova E, Litvinov S,
946 Ayodo G, Comas D, Hammer MF, Kivisild T, Klitz W, Winkler CA, Labuda D, Bamshad M,
947 Jorde LB, Tishkoff SA, Watkins WS, Metspalu M, Dryomov S, Sukernik R, Singh L, Thangaraj
948 K, Paabo S, Kelso J, Patterson N, Reich D 2016. The Simons Genome Diversity Project: 300
949 genomes from 142 diverse populations. *Nature* 538: 201-206. doi: 10.1038/nature18964

950 Mikkelsen T, Hillier L, Eichler E, Zody M, Jaffe D, Yang S-P, Enard W, Hellmann I, Lindblad-Toh
951 K, Altheide T 2005. Initial sequence of the chimpanzee genome and comparison with the
952 human genome. *Nature* 437: 69-87.

953 Mona S, Tommaseo-Ponzetta M, Brauer S, Sudoyo H, Marzuki S, Kayser M 2007. Patterns of
954 Y-chromosome diversity intersect with the Trans-New Guinea hypothesis. *Molecular*
955 *Biology and Evolution* 24: 2546-2555.

956 Nagle N, Ballantyne KN, van Oven M, Tyler-Smith C, Xue Y, Taylor D, Wilcox S, Wilcox L,
957 Turkalov R, Van Oorschot RA 2016. Antiquity and diversity of aboriginal Australian
958 Y-chromosomes. *American Journal of Physical Anthropology* 159: 367-381.

959 Norman K, Inglis J, Clarkson C, Faith JT, Shulmeister J, Harris D 2018. An early colonisation
960 pathway into northwest Australia 70-60,000 years ago. *Quaternary Science Reviews* 180:
961 229-239. doi: 10.1016/j.quascirev.2017.11.023

962 O'Connell JF, Allen J, Williams MAJ, Williams AN, Turney CSM, Spooner NA, Kamminga J,
963 Brown G, Cooper A 2018. When did *Homo sapiens* first reach Southeast Asia and Sahul?
964 *Proc Natl Acad Sci U S A* 115: 8482-8490. doi: 10.1073/pnas.1808385115

965 O'Connor S 2007. New evidence from East Timor contributes to our understanding of earliest
966 modern human colonisation east of the Sunda Shelf. *Antiquity* 81.

967 Pagani L, Lawson DJ, Jagoda E, Morseburg A, Eriksson A, Mitt M, Clemente F, Hudjashov G,
968 DeGiorgio M, Saag L, Wall JD, Cardona A, Magi R, Sayres MA, Kaewert S, Inchley C, Scheib
969 CL, Jarve M, Karmin M, Jacobs GS, Antao T, Iliescu FM, Kushniarevich A, Ayub Q, Tyler-
970 Smith C, Xue Y, Yunusbayev B, Tambets K, Mallick CB, Saag L, Pocheshkhova E, Andriadze
971 G, Muller C, Westaway MC, Lambert DM, Zoraqi G, Turdikulova S, Dalimova D, Sabitov Z,
972 Sultana GN, Lachance J, Tishkoff S, Momyaliev K, Isakova J, Damba LD, Gubina M,
973 Nymadawa P, Evseeva I, Atramentova L, Utevska O, Ricaut FX, Brucato N, Sudoyo H,
974 Letellier T, Cox MP, Barashkov NA, Skaro V, Mulahasanovic L, Primorac D, Sahakyan H,
975 Mormina M, Eichstaedt CA, Lichman DV, Abdullah S, Chaubey G, Wee JT, Mihailov E,
976 Karunas A, Litvinov S, Khusainova R, Ekomasova N, Akhmetova V, Khidiyatova I,
977 Marjanovic D, Yepiskoposyan L, Behar DM, Balanovska E, Metspalu A, Derenko M,
978 Malyarchuk B, Voevoda M, Fedorova SA, Osipova LP, Lahr MM, Gerbault P, Leavesley M,
979 Migliano AB, Petraglia M, Balanovsky O, Khusnutdinova EK, Metspalu E, Thomas MG,
980 Manica A, Nielsen R, Villems R, Willerslev E, Kivisild T, Metspalu M 2016. Genomic

981 analyses inform on migration events during the peopling of Eurasia. *Nature* 538: 238-242.
982 doi: 10.1038/nature19792

983 Patterson N, Price AL, Reich D 2006. Population structure and eigenanalysis. *PLoS Genet* 2:
984 e190. doi: 10.1371/journal.pgen.0020190

985 Patterson NJ, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T,
986 Reich D 2012. Ancient admixture in human history. *Genetics* 192: 1065-1093. doi:
987 10.1534/genetics.112.145037

988 Pawley A. 2005. Papuan pasts: Cultural, linguistic and biological histories of Papuan-speaking
989 peoples: Pacific Linguistics, Research School of Pacific and Asian Studies, The

990 Pedro N, Brucato N, Fernandes V, Andre M, Saag L, Pomat W, Besse C, Boland A, Deleuze JF,
991 Clarkson C, Sudoyo H, Metspalu M, Stoneking M, Cox MP, Leavesley M, Pereira L, Ricaut
992 FX 2020. Papuan mitochondrial genomes and the settlement of Sahul. *J Hum Genet*. doi:
993 10.1038/s10038-020-0781-3

994 Petkova D, Novembre J, Stephens M 2016. Visualizing spatial population structure with
995 estimated effective migration surfaces. *Nat Genet* 48: 94-100. doi: 10.1038/ng.3464

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

998 Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling
999 DE, Gauthier LD, Levy-Moonshine A, Roazen D, Shakir K, Thibault J, Chandran S, Whelan C,
1000 Lek M, Gabriel S, Daly MJ, Neale B, MacArthur DG, Banks E 2018. Scaling accurate genetic
1001 variant discovery to tens of thousands of samples. *bioRxiv*: 201178. doi: 10.1101/201178

1002 Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G,
1003 Sudmant PH, de Filippo C, Li H, Mallick S, Dannemann M, Fu Q, Kircher M, Kuhlwilm M,
1004 Lachmann M, Meyer M, Ongyerth M, Siebauer M, Theunert C, Tandon A, Moorjani P,
1005 Pickrell J, Mullikin JC, Vohr SH, Green RE, Hellmann I, Johnson PL, Blanche H, Cann H,
1006 Kitzman JO, Shendure J, Eichler EE, Lein ES, Bakken TE, Golovanova LV, Doronichev VB,
1007 Shunkov MV, Derevianko AP, Viola B, Slatkin M, Reich D, Kelso J, Paabo S 2014. The
1008 complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505: 43-
1009 49. doi: 10.1038/nature12886

1010 Redd AJ, Stoneking M 1999. Peopling of Sahul: mtDNA variation in aboriginal Australian and
1011 Papua New Guinean populations. *Am J Hum Genet* 65: 808-828. doi: 10.1086/302533

1012 Reesink G, Singer R, Dunn M 2009. Explaining the linguistic diversity of Sahul using
1013 population models. *PLoS Biol* 7: e1000241.

1014 Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel
1015 U, Johnson PL, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer
1016 M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin JJ,
1017 Kelso J, Slatkin M, Paabo S 2010. Genetic history of an archaic hominin group from
1018 Denisova Cave in Siberia. *Nature* 468: 1053-1060. doi: 10.1038/nature09710

1019 Reich D, Patterson N, Kircher M, Delfin F, Nandineni MR, Pugach I, Ko AMS, Ko YC, Jinam TA,
1020 Phipps ME, Saitou N, Wollstein A, Kayser M, Paabo S, Stoneking M 2011. Denisova
1021 Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania.
1022 *American Journal of Human Genetics* 89: 516-528. doi: 10.1016/j.ajhg.2011.09.005

1023 Sankararaman S, Mallick S, Patterson N, Reich D 2016. The Combined Landscape of
1024 Denisovan and Neanderthal Ancestry in Present-Day Humans. *Curr Biol* 26: 1241-1247.
1025 doi: 10.1016/j.cub.2016.03.037

1026 Schiffels S, Durbin R 2014. Inferring human population size and separation history from
1027 multiple genome sequences. *Nat Genet* 46: 919-925. doi: 10.1038/ng.3015

1028 Schiffels S, Wang K. 2020. MSMC and MSMC2: the multiple sequentially markovian
1029 coalescent. In. *Statistical population genomics: Humana*, New York, NY. p. 147-166.

1030 Shaw B, Field JH, Summerhayes GR, Coxe S, Coster ACF, Ford A, Haro J, Arifeae H, Hull E,
1031 Jacobsen G, Fullagar R, Hayes E, Kealhofer L 2020. Emergence of a Neolithic in highland
1032 New Guinea by 5000 to 4000 years ago. *Sci Adv* 6: eaay4573. doi: 10.1126/sciadv.aay4573

1033 Skoglund P, Posth C, Sirak K, Spriggs M, Valentin F, Bedford S, Clark GR, Reepmeyer C,
1034 Petchey F, Fernandes D, Fu Q, Harney E, Lipson M, Mallick S, Novak M, Rohland N,
1035 Stewardson K, Abdullah S, Cox MP, Friedlaender FR, Friedlaender JS, Kivisild T, Koki G,
1036 Kusuma P, Merriwether DA, Ricaut FX, Wee JT, Patterson N, Krause J, Pinhasi R, Reich D
1037 2016. Genomic insights into the peopling of the Southwest Pacific. *Nature* 538: 510-513.
1038 doi: 10.1038/nature19844

1039 Summerhayes GR, Field JH, Shaw B, Gaffney D 2017. The archaeology of forest exploitation
1040 and change in the tropics during the Pleistocene: The case of Northern Sahul (Pleistocene
1041 New Guinea). *Quaternary International* 448: 14-30. doi: 10.1016/j.quaint.2016.04.023

1042 Summerhayes GR, Leavesley M, Fairbairn A, Mandui H, Field J, Ford A, Fullagar R 2010.
1043 Human adaptation and plant use in highland New Guinea 49,000 to 44,000 years ago.
1044 Science 330: 78-81. doi: 10.1126/science.1193130

1045 Teixeira JC, Cooper A 2019. Using hominin introgression to trace modern human dispersals.
1046 Proc Natl Acad Sci U S A 116: 15327-15332. doi: 10.1073/pnas.1904824116

1047 Terrell JE 2004. The 'sleeping giant' hypothesis and New Guinea's place in the prehistory of
1048 Greater Near Oceania. World Archaeology 36: 601-609.

1049 Tobler R, Rohrlach A, Soubrier J, Bover P, Llamas B, Tuke J, Bean N, Abdullah-Highfold A,
1050 Agius S, O'Donoghue A, O'Loughlin I, Sutton P, Zilio F, Walshe K, Williams AN, Turney CSM,
1051 Williams M, Richards SM, Mitchell RJ, Kowal E, Stephen JR, Williams L, Haak W, Cooper A
1052 2017. Aboriginal mitogenomes reveal 50,000 years of regionalism in Australia. Nature
1053 544: 180-184. doi: 10.1038/nature21416

1054 Torrence R, Swadling P 2008. Social networks and the spread of Lapita. Antiquity 82: 600-
1055 616.

1056 Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM, Dannemann M, Grote S,
1057 McCoy RC, Norton H, Scheinfeldt LB, Merriwether DA, Koki G, Friedlaender JS, Wakefield
1058 J, Paabo S, Akey JM 2016. Excavating Neandertal and Denisovan DNA from the genomes
1059 of Melanesian individuals. Science 352: 235-239. doi: 10.1126/science.aad9416

1060 Wallace AR. 1869. The Malay Archipelago-Volume II.(of II.).

1061 Westaway KE, Louys J, Awe RD, Morwood MJ, Price GJ, Zhao JX, Aubert M, Joannes-Boyau R,
1062 Smith TM, Skinner MM, Compton T, Bailey RM, van den Bergh GD, de Vos J, Pike AWG,
1063 Stringer C, Saptomo EW, Rizal Y, Zaim J, Santoso WD, Trihascaryo A, Kinsley L, Sulistyanto
1064 B 2017. An early modern human presence in Sumatra 73,000-63,000 years ago. Nature
1065 548: 322-325. doi: 10.1038/nature23452

1066

1067

1068 **Figures captions**

1069 Figure 1. Genetic diversity of individuals from New Guinea, Wallacea, the Bismarck
1070 Archipelago and northeast Australia. (A) Geographic locations of all 58 genetic sub-groups
1071 defined by the fineSTRUCTURE analysis. (B) Principal Component Analysis representing
1072 26.7% of the total variability of all individual genomes from the 58 sub-groups, masked for
1073 Asian genetic ancestry. The two first components are represented (PC1 and PC2). (C)
1074 ADMIXTURE plot for eight genetic components summarized by genetic sub-groups (based on
1075 K=8 from Supplementary Fig. 7). The grey component combines four components mainly
1076 present in African and Eurasian groups. Numbers above bar plots identify genetic sub-groups
1077 referred in Supplementary table 2. (D) TREEMIX plot for all 58 sub-groups (based on
1078 Supplementary Fig. 7). All four panels used the same color scheme to identify the 58 genetic
1079 sub-groups.

1080

1081 Figure 2. Scenario of the settlement of Sahul. (A) f_3 -outgroup analysis for all Sahul groups
1082 compared to populations from the North Maluku Islands (y-axis) or Lesser Sunda Islands (x-
1083 axis). Populations are identified by the same color scheme as in Fig. 1. (B) Best fitting
1084 qpGraph models considering one migration to Sahul (model A; Z-score=-2.570) or two
1085 migrations to Sahul (model B; Z-score=2.301). One subgroup was used as representant of a
1086 region: populations from North Maluku (Wallacea subgroup 23, following Figure 1C and
1087 defined in Supplementary table 2), Lesser Sunda Islands (Wallacea subgroup 13), New
1088 Guinea (New Guinea subgroup 16), the Bismarck Archipelago (Bismarck subgroup 5),
1089 northeast Australia (Australia subgroup 1) and Africa (Africa_5). Solid arrows are labeled
1090 with branch lengths in 1000 times drift units. Dotted arrows indicate gene flows whose
1091 proportions are indicated by percentages. (C) Relative cross-coalescence rate curves
1092 estimated by MSMC2 between populations from Wallacea (Lesser Sunda Islands, Wallacea
1093 subgroup 13), New Guinea (New Guinea subgroup 16), the Bismarck Archipelago (Bismarck
1094 subgroup 5) and northeast Australia (Australia subgroup 1).

1095

1096 Figure 3. Scenario of the settlement of New Guinea. (A) Best fitting qpGraph model for
1097 populations from New Guinean southeast lowlands (New Guinea subgroup 16, following
1098 Figure 1C and defined in Supplementary table 2), north lowlands (New Guinea subgroup 28),
1099 south lowlands (New Guinea subgroup 39), east highlands (New Guinea subgroup 45), west

1100 highlands (New Guinea subgroup 55) and Africa (Africa_5)(Z-score=2.759). Solid arrows are
1101 labeled with branch lengths in 1000 times drift units. Dotted arrows indicate gene flows
1102 whose proportions are indicated by percentages. (B) Relative cross-coalescence rate curves
1103 estimated by MSMC2 between populations from New Guinean lowlands and highlands.

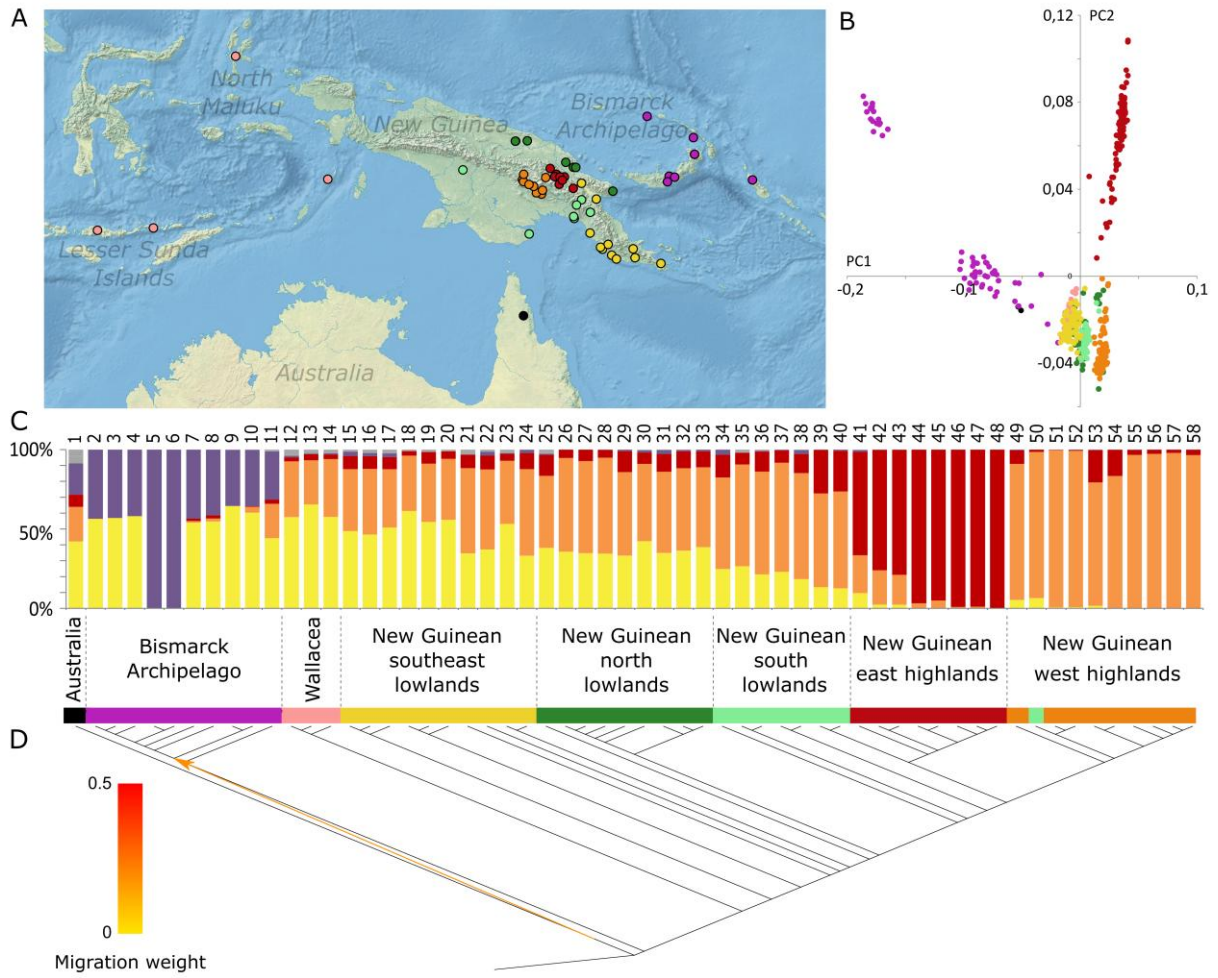
1104

1105 Figure 4. Genomic scenario of the human dynamics in north Sahul during the Upper
1106 Pleistocene. Light blue arrows represent the settlement of north Sahul following a scenario
1107 with one migration to Sahul (model A). The dark blue arrows represent the settlement of
1108 north Sahul following a scenario with two migrations to Sahul (model B). The dotted dark
1109 blue arrow represents a gene flow following a scenario with two migrations to Sahul (model
1110 B). Yellow arrows represent the settlement of New Guinea. The dotted orange arrows
1111 represent gene flows after the initial settlement of Sahul. The black dots represent
1112 approximate locations of archaeological sites older than 40kya (O'Connor 2007;
1113 Summerhayes, et al. 2017). The dashed black lines represent ecological lines in Wallacea.
1114 The dark grey areas represent current land area, with light grey areas representing
1115 estimated land area around 50,000 years ago based on paleogeological reconstructions
1116 (Coller 2009).

1117

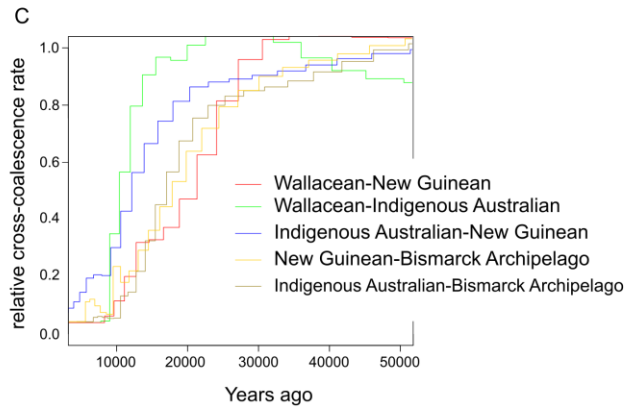
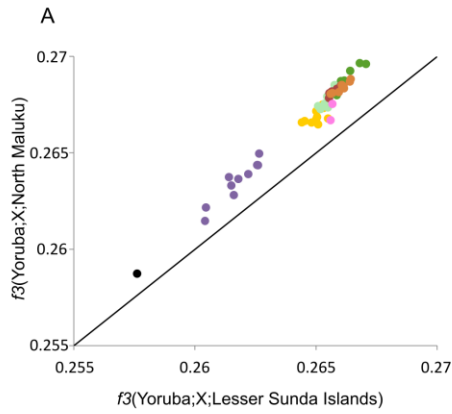
1118 **Figures**

1119

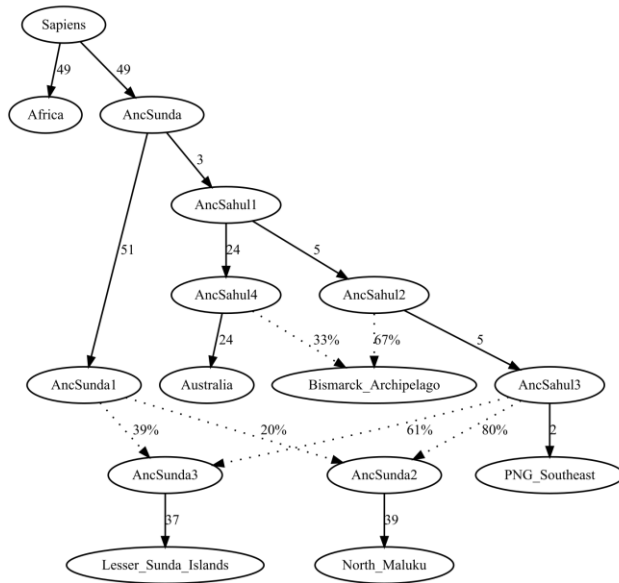


1120

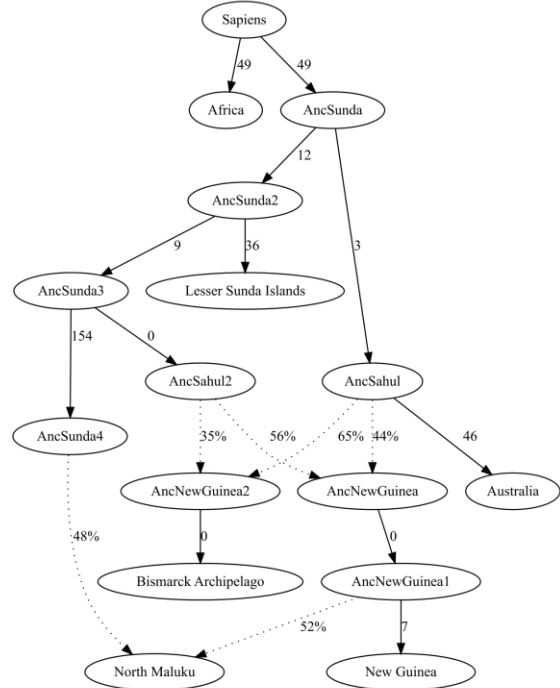
1121 **Figure 1**



B **Model A: One migration to Sahul**

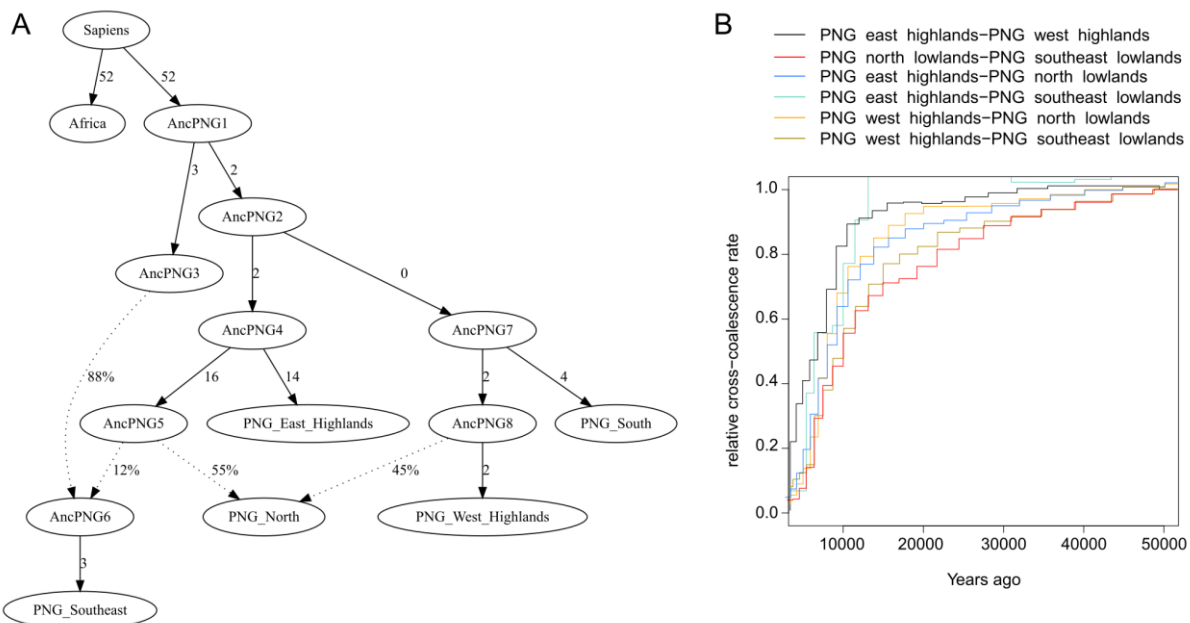


Model B: Two migrations to Sahul



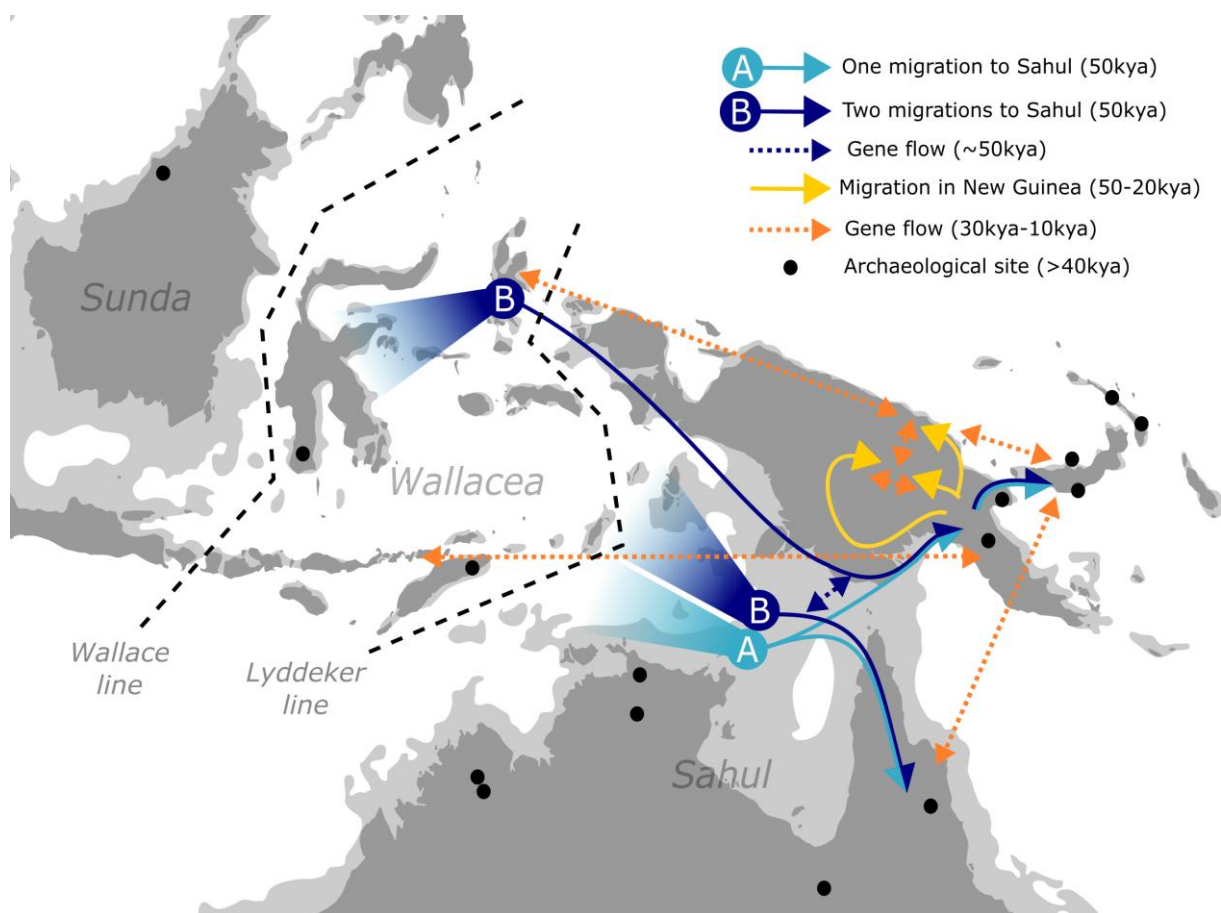
1122

1123 Figure 2



1124
1125
1126
1127

Figure 3



1128
1129

Figure 4