2

Assessing human genome-wide variation in the Massim region of Papua New Guinea and implications for the Kula trading tradition

Dang Liu[1,2], Benjamin M. Peter[1], Wulf Schiefenhövel[3], Manfred Kayser[4]* and Mark Stoneking[1,5]*

1. Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
2. Human Evolutionary Genetics Unit, Institut Pasteur, UMR 2000, CNRS, Paris, France
3. Human Ethology Group, Max Planck Institute for Ornithology, Seewiesen, Germany
4. Department of Genetic Identification, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands
5. Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive, UMR 5558, Villeurbanne, France

*Corresponding authors: Mark Stoneking, stoneking@eva.mpg.de; Manfred Kayser, m.kayser@erasmusmc.nl

19

Conflict of interest: The authors declare no conflict of interest.

24

Running title: Genome-wide variation in the Massim region

Keywords: Massim; Kula Ring; Genome-wide data; Austronesian; Papuan

27

Abstract

The Massim, a cultural region that includes the southeastern tip of mainland Papua New Guinea (PNG) and nearby PNG offshore islands, is renowned for a trading network called Kula, in which different valuable items circulate in different directions among some of the islands. While the Massim has been a focus of anthropological investigation since the pioneering work of Malinowski in 1922, the genetic background of its inhabitants remains relatively unexplored. To characterize the Massim genomically, we generated genome-wide SNP data from 192 individuals from 15 groups spanning the entire region. Analyzing these together with comparative data, we found that all Massim individuals have variable Papuan-related (indigenous) and Austronesian-related (arriving ~3,000 years ago) ancestries. Individuals from Rossel Island in southern Massim, speaking an isolate Papuan language, have the highest amount of a distinct Papuan ancestry. We also investigated the recent contact via sharing of identical by descent (IBD) genomic segments and found that Austronesian-related IBD tracts are widely distributed geographically, but Papuan-related tracts are shared exclusively between the PNG mainland and Massim, and between the Bismarck and Solomon Archipelagoes. Moreover, the Kula-practicing groups of the Massim show higher IBD sharing among themselves than do groups that do not participate in Kula. This higher sharing predates the formation of Kula, suggesting that extensive contact between these groups since the Austronesian settlement may have facilitated the formation of Kula. Our study provides the first comprehensive genome-wide assessment of Massim inhabitants and new insights into the fascinating Kula system.

1 **Introduction**

2 New Guinea has a long history of human occupation and harbors extensive ethnolinguistic

3 diversity. Modern humans first colonized Sahul (the connected Australia – New Guinea land

4 mass) at least 47 thousand years ago (kya), and perhaps as long ago as 65 kya (Roberts, et al.

5 1990; Summerhayes, et al. 2010; O'Connell and Allen 2015; Clarkson, et al. 2017), while ~3 kya

6 the Austronesian expansion/settlement brought a second wave of human migration into the

7 region (Bellwood 2007; Kirch 2017). This long-term isolation of human occupation and different

8 episodes of human migration promoted diverse regional culture developments, including an

9 independent invention of farming in the New Guinea highlands at least 4-5 kya (Denham 2004;

10 Shaw, Field, et al. 2020) and the Lapita culture, associated with the expansion of Austronesians

11 from Near Oceania into the Pacific (Spriggs 1995; Kirch 2017).

12 Geographically, the Massim encompasses the southeastern tip of mainland Papua New

13 Guinea (PNG) and nearby offshore islands (Fig. 1A) and is considered by anthropologists to be a

14 culturally-defined region (Seligmann 1910; Shaw 2019). The Massim region is well-known for

15 the Kula Ring tradition (or Kula), a network trading system in which two types of unique

16 necklaces circulate among the islands in opposite directions, facilitating inter-island social and

17 economic relationships (Malinowski 1922; Ziegler 2017; Irwin, et al. 2018; Kuehling 2021). The

18 Kula connects the islands of the northern and the western Massim (including the eastern tip of

19 Mainland PNG) and also includes Misima, the most northern island of the southern Massim,

20 while other islands of the southern Massim are not involved in the Kula (Fig. 1A). Due to

21 seasonal wind conditions, traders on Kula voyages often spend long periods of time on different

22 islands (Malinowski 1922; Irwin, et al. 2018), which might further facilitate interactions,

23 including genetic (Schiefenhövel 2004), between the groups participating in Kula.

24 Since the initial description of Kula by Malinowski in 1922 (Malinowski 1922), the

25 region has attracted the attention of anthropologists, archaeologists, and linguists, and

26 chronologies of Massim have been proposed (Shaw 2016; Shaw, Coxe, Kewibu, et al. 2020).

27 The earliest evidence for human occupation of the Massim is dated to 17.3 kya, on Paneati Island

28 in the Louisiades Archipelago (Shaw, Coxe, Haro, et al. 2020); undated obsidian and stone

29 artifacts that are similar to those dated to the Late Pleistocene or Early Holocene elsewhere in

30 New Guinea further support human occupation around this time, when lowered sea levels

31 connected much of the Massim to the PNG mainland (Shaw, et al. 2016; Shaw 2017). After

3

settlement by Lapita people ~2.5-3 kya, the northern and southern Massim exhibited separate
cultural developments, until ~500-200 years ago with the formation of Kula (Shaw 2016).
Linguistically, almost all Massim inhabitants speak languages belonging to the Austronesian
language family, except for a few groups living on mainland PNG who speak Papuan languages
belonging to the Trans-New Guinea language family (Eberhard, et al. 2021), and the Rossel
Islanders on the most eastern tip of the southern Massim. Rossel Islanders speak a Papuan
language that has been classified either as a language isolate (Levinson and Majid 2013;
Stebbins, et al. 2017) or as belonging to the Yele-West New Britain language family, linking
Rossel linguistically with the Bismarck Archipelago (Ross 2005; Eberhard, et al. 2021).

While the Massim region has been well studied by archaeologists, anthropologists, and
linguists, genetic variation in the Massim remains largely unexplored, even though it occupies a
key position in connecting the northern and southern coasts of New Guinea as well as connecting
New Guinea with the neighboring Solomon Islands and other regions of Oceania. The most
comprehensive human genetic study of the Massim to date analyzed mitochondrial DNA
(mtDNA) and Y chromosome variation, and found regional genetic-geographic population
structure for mtDNA but not for the Y chromosome (van Oven, et al. 2014). This was interpreted
as a potential signature of the Kula, as the travel between islands to perform the trading of the
goods is mostly mediated by males, which could reduce inter-island genetic differences for the Y
chromosome but not for mtDNA. However, studies of genome-wide data provide much richer
information concerning admixture and population history, as shown by two recent genome-wide
studies of PNG (Bergström, et al. 2017; Brucato, et al. 2021), but genome-wide studies of the
Massim are lacking as of yet. Here, we report the results of comprehensive genome-wide
analyses of the Massim region with resulting insights into the genetic history of Oceania in
general and implications for the Kula tradition in particular.

Results

Overall genetic variation of Austronesian and Papuan ancestries in the Massim

We generated genome-wide single nucleotide polymorphism (SNP) array data on the Illumina
Infinium Multi-Ethnic Global Array (MEGA; ~1.5 million SNPs) for 192 individuals from 15
groups spanning the entire Massim region including northern, southern and western Massim as
well as the mainland parts from Collingwood Bay (Fig. 1A). We merged our data with published

4

array and WGS data encompassing East Asia and Oceania and additionally included Africans and Europeans as more distant comparative groups (Fig. 1A; Supplementary Fig. S1; Supplementary Table S1). To obtain an overview of the genetic variation and genetic-geographic population structure revealed with our compiled genomic dataset, we first performed principal component analysis (PCA). In a PCA plot focusing on the East Asian and Oceanian groups (Fig. 1B; Supplementary Fig. S2), we observed a striking cline with East Asians at one pole and PNG highlanders at the other; the Massim groups fall in between together with groups from the Central Province of PNG, the Bismarck Archipelago (in short, Bismarcks), and the Solomon Archipelago (i.e. Bougainville and the Solomon Islands; in short, Solomons). Groups from the PNG lowlands are placed close to PNG highlanders, except for the Central group, which comprises more Austronesian speakers. Bellona, Renell and Tikopia are Polynesian outliers in the Solomons, i.e. groups that migrated back to the Solomons from Polynesia, which explains their placement further towards East Asians than any other Near Oceanian group studied. The northern Massim groups are closer to the East Asian pole, while the southern Massim group from Rossel is closest to the PNG highlander pole, and other southern Massim as well as all western Massim and Collingwood Bay groups fall in between. Notably, there is a division in the PCA within southern Massim, with Rossel being closest to the PNG mainland and highland groups and Sudest, which geographically is located next to Rossel, being close to Rossel, while the other southern Massim groups fall together with northern and western Massim groups and those from Collingwood Bay.

Since the PCA suggests mixed East Asian-Papuan ancestry in the Massim groups, we explored this further in an ADMIXTURE analysis. At K=2, one component (pink) is most enriched in the East Asian Austronesian groups while the other component (blue) is enriched in the Papuan highlanders (Supplementary Fig. S3); all other Oceanian groups (including the Massim) exhibit a mixture of these two genetic ancestry components, albeit in different proportions. The northern Massim individuals, who speak Austronesian languages, have the highest amount of Austronesian-related ancestry (~52% pink), while the Papuan-speaking Rossel islanders from southern Massim have the lowest amount (~20% pink), and hence the highest amount of Papuan (~80% blue) ancestry (Supplementary Table S2). However, several Austronesian-speaking groups in the Massim, including those from Sudest, the western Massim,

5

and Collingwood Bay, also have more Papuan than Austronesian ancestry, suggesting substantial contact between Papuan and Austronesian groups.

This large range in variation of Austronesian (~20-52%) vs. Papuan genomic ancestry in the Massim exceeds that of nearby, larger island regions: Austronesian ancestry varies between ~25-36% in the Bismarcks and between ~38-54% in the Solomons (excluding Santa Cruz and Polynesian outliers) (Supplementary Fig. S3). We therefore investigated the Austronesian vs. Papuan ancestry in the Massim in more detail via outgroup f3 statistics. A plot of the f3 values measuring shared drift between Oceanian populations and East Asian Austronesians, vs. the f3 values measuring shared drift between Oceanian populations and PNG highlanders (Supplementary Fig. S4), gives results similar to the PCA plot (Fig. 1B), namely that Massim groups harbor both Austronesian and Papuan ancestry with different proportions depending on their locations within the Massim region. As before, northern Massim groups have more Austronesian ancestry while Rossel Islanders from southern Massim have more Papuan ancestry. As in the PCA plot, the Massim groups are distributed along a linear cline between PNG highlanders with high amounts of Papuan ancestry at one pole, and Polynesian outliers in the Solomons with high amounts of Austronesian ancestry at the other pole. Moreover, the shape of the f3 plot suggests a single major admixture episode rather than continuous gene flow or a tree-like history (Bergstrom, et al. 2020).

In addition to the drift/allele sharing analyses, a haplotype-based method (GLOBETROTTER) also suggests a single pulse of admixture for all the Massim groups except for Rossel (Supplementary Fig. S5); Rossel exhibits an unclear admixture signal probably due to the low amount of Austronesian ancestry. Moreover, it is likely that the incoming population was already admixed (Supplementary Fig. S5) i.e., had both Austronesian and Papuan ancestry. Consistently, the percentage of modeled Austronesian-related ancestry is highest in the northern Massim groups (~42%), and the Papuan-related ancestry is highest in the Rossel group from southern Massim (~73%; Supplementary Table S2). GLOBETROTTER also infers admixture dates of ~1-3 kya for the Massim groups, and similar dates are inferred with another method, MALDER (Supplementary Fig. S6); additionally, MALDER does not find evidence for more than one admixture event. The inferred dates are not correlated with the amount of Austronesian ancestry (Supplementary Fig. S7), suggesting that the higher amounts of Austronesian ancestry

6

1 in some Massim groups cannot be explained by a longer period of contact, but instead must

2 reflect other social circumstances.

3

4 **Austronesian and Papuan ancestries specific to the Massim**

5 In the ADMIXTURE analysis, the lowest cross-validation error (and hence, the best-supported

6 number of ancestry components) occurred at K=8 (Supplementary Fig. S8A). Six ancestry

7 components were found in the Oceanians (Fig. 2); the other two were restricted to Africans, and

8 to Europeans and East Asians, respectively. The ancestry components found in Oceanians

9 include:

10 1) a light green component at highest frequency in Chimbu and Madang highlanders and also

11 enriched in Western and Eastern highlanders; this component is present at uniformly low

12 frequencies across the Massim.

13 2) a blue component at highest frequency in Southern PNG highlanders and Enga and enriched

14 in other mainland PNG groups; this component is present up to medium frequency in the

15 Massim, Bismarcks and Santa Cruz.

16 3) a pink component enriched in Central PNG that exists in low frequency in a few other lowland

17 PNG groups, the Solomons (including the Polynesian outlier Tikopia, but not the other

18 Polynesian outlier, Bellona_Rennell), and the Bismarcks; and is the major component in all

19 Massim groups except Rossel and neighboring Sudest.

20 4) a black component that is at highest frequency in the Rossel group, second highest in the

21 neighboring group from Sudest, and at lower frequency in several other groups from the western

22 and southern Massim.

23 5) a dark green component enriched in Bougainville, the Bismarcks and the Solomons (including

24 the Polynesian outlier Tikopia, but not the other Polynesian outlier, Bellona_Rennell); this

25 component is present at very low frequency in some PNG mainland groups but notably absent

26 from the Massim.

27 6) a peach component that is the only component in the Polynesian outlier Bellona_Rennell, and

28 is found in high frequency in Tikopia and in low frequency in other Solomons and in Bismarck

29 groups; this component is not found in the Massim.

30 The blue and pink components detected at K=8 (Fig. 2) overlap largely, but not completely, with

31 the results for the Oceanian groups observed at K=2 (Supplementary Fig. S3). All of the groups

7

exhibiting any of the four components newly detected with K=8 (light green, black, dark green, and peach) in the Oceanians have more Papuan than Austronesian ancestry detected at K=2 (Supplementary Fig. S3) except for the Polynesian outliers (peach), which suggests substantial variation in Papuan ancestry as well as a drifted Austronesian ancestry in the Polynesian outliers. Similar results were obtained with a more strict threshold for kinship quality control, suggesting this observed pattern is not due to cryptically related individuals (Supplementary Figs. S8B and S9). Hence, we further investigated the Papuan and Austronesian ancestry-specific variation in the Massim.

We first calculated f4 statistics of the form f4(Test groups, East Asian Austronesians; Chimbu/Bougainville/Rossel, Southern Highlanders) to test for different affinities of each test group (all other Oceanian groups) with Chimbu, Bougainville, or Rossel-related ancestry when compared to Southern Highlanders. The results of this f4 statistic will be influenced by both differences in Papuan-related ancestry as well as different amounts of Austronesian-related ancestry in the test groups. Therefore, to control for the latter, we plotted this f4 statistic against an f4 statistic of the form f4(Test groups, Southern Highlanders; East Asian Austronesians, Australians). This f4 statistic was previously shown mathematically to provide a measure of Austronesian-related ancestry in the test groups (Lipson, et al. 2020). Groups exhibiting significant deviations from the regression line in each plot would indicate differential Papuan ancestries in those groups. The results for the comparison of Chimbu vs. Southern Highlanders shows no significant deviations from the regression line for any Oceanian group (Supplementary Fig. S10A), indicating that all of the groups tested have equal affinities with Chimbu and Southern Highlanders ancestry. In the comparison between Bougainville and Southern Highlanders (Supplementary Fig. S10B), several of the groups from the Bismarcks and the Solomons exhibit stronger affinities with Bougainville, as noted previously (Pugach, et al. 2018), while several PNG groups exhibit stronger affinities with the Southern Highlanders; however, none of the Massim groups show any preferential affinities. In the comparison between Rossel and Southern Highlanders (Supplementary Fig. S10C), again several PNG groups exhibit stronger affinities with Southern Highlanders, along with the Polynesian Outliers from the Solomons. Interestingly, there is variation in genomic ancestry among the Massim groups: southern Massim groups share more ancestry with Rossel; Collingwood Bay groups (from

8

mainland PNG) share more ancestry with the Southern Highlanders; and the remaining Massim groups do not share excess ancestry with either Rossell or Southern Highlanders.

We then carried out a haplotype-based method of analysis (ChromoPainter) in which we restricted Massim groups to be recipients and not donors (with the exception of Rossel, in order to have a source of the putative distinct Papuan ancestry in the southern Massim). The results (Supplementary Fig. S11) are consistent with the f4 analyses and strongly support the existence of more than one distinct Papuan-related ancestry in the Massim groups (one related to Rossel, and at least one more related to other Papuan groups).

Switching Australians to non-Austronesian East Asians in the f4(Test groups, Southern Highlanders; East Asian Austronesians, Australians) further confirmed that the non-Papuan ancestry in the Massim groups is associated with Austronesian ancestry (Supplementary Figs. S10D-E). We next used f4 statistics to investigate variation in Austronesian-related ancestry in the Massim groups and found that Massim groups are equally-related to the most differentiated Austronesian-related ancestries, East Asian Austronesians vs. Polynesian outliers (Supplementary Fig. S12), suggesting no significant differences in the source(s) of Austronesian-related ancestry among the Massim groups.

To study the ancestry-specific variation we detected in the Massim in more detail, we applied local ancestry inference via RFMix and extracted Austronesian and Papuan ancestry-specific segments. The amount of Austronesian-related ancestry inferred by RFMix is slightly lower than that inferred by the ADMIXTURE and GLOBETROTTER analyses, while similar proportions of Papuan ancestries are inferred by all three methods (Supplementary Table S2). This suggests that the Papuan ancestry-specific segments are identified with high confidence, but there is more uncertainty in inferring Austronesian-related ancestry segments, probably due to a poorer proxy for this source. A PCA based on Papuan ancestry-specific segments shows three poles, consisting of: the Bismarck and Solomon groups, Rossel, and Papuan highlanders (Fig. 3A; Supplementary Fig. S13A), suggesting that the Papuan ancestry associated with Rossel probably reflects strong genetic drift, as further evidenced by the high degree of sharing of identity-by-descent (IBD) fragments and runs of homozygosity (ROH) within Rossel (Supplementary Fig. S14). The Bismarck and Solomon groups are separated from Rossel and Papuan highlanders at PC1, suggesting the Papuan ancestry in Rossel is closer to Papuan highlanders than to the Papuan ancestry of Bismarck and Solomon groups. This is further

9

supported by Papuan ancestry-specific ADMIXTURE and TreeMix results (Supplementary Figs. S15 and S16). In contrast to the Papuan ancestry-specific PCA, in the Austronesian ancestry-specific PCA all of the Oceanian groups fall together in a cluster quite distinct from East Asian groups (Fig. 3B and Supplementary Fig. S13B) with Polynesian outliers at the extreme pole; this lack of distinction in the Austronesian-related ancestry of Massim groups is consistent with the results from the allele-sharing statistics (f3 and f4) and further supported by an Austronesian ancestry-specific ADMIXTURE analysis (Supplementary Fig. S17).

Recent contact involving Massim groups and implications for the Kula tradition

To investigate in more detail genetic contacts involving the Massim groups in the recent past, we analyzed sharing of identify-by-descent (IBD) segments. Different size ranges of IBD segments are informative about contact at different times, extending back to ~4 kya (Ralph and Coop 2013; Al-Asadi, et al. 2019). In the range of 1 to 5 cM, which roughly corresponds to an average of ~2.7 kya, i.e. around the time of the arrival of Austronesians in the region (Spriggs 1995; Shaw 2016), there is strong IBD sharing among Oceanian groups, in particular among the Massim, the Bismarcks and Solomons, and among the Papuan highlanders (Fig. 4A; Supplementary Fig. S18). In the range of 5 to 10 cM, roughly corresponding to ~675 ya, the strong connection between mainland Papua New Guinea and the offshore islands in Massim is no longer there, but the connection between northern and southern Massim remains. In the range of over 10 cM or roughly 225 ya, strong sharing is only seen within some parts of the Papuan highlands, northern Massim, and southern Massim, respectively. To further test the idea that the extensive IBD sharing in the range of 1 to 5 cM reflects the Austronesian settlement, we extracted and analyzed Austronesian ancestry-specific and Papuan ancestry-specific IBD segments. The results show that in the range of 1 to 5 cM, Austronesian ancestry-specific IBD segments account for most of the sharing between regions, while Papuan ancestry-specific IBD sharing shows a clear distinction between the PNG mainland and Massim groups, vs. the Bismarcks and Solomons (Fig. 4B and 4C; Supplementary Figs. S19 and S20). In particular, the strong sharing between many Massim groups and the Central Province group in the ChromoPainter analysis (Supplementary Fig. S11) can be attributed to Austronesian-related ancestry, as this accounts for the IBD-sharing between them (Supplementary Figs. S18 and S19). And, the closer relationship in Papuan-related ancestry between the PNG mainland and Massim

10

groups, vs. either of these and the Bismarck/Solomon groups, is consistent with the PCA/ADMIXTURE/TreeMix based on Papuan-related ancestry (Fig. 3A; Supplementary Figs. S15 and S16).

Finally, to investigate the potential relationship between the Kula tradition and genomic variation, we compared the IBD sharing between Kula-practicing vs. non-Kula-practicing groups (in short, Kula vs. non-Kula). We measured IBD similarities among Kula/non-Kula taking the differences in IBD sharing within groups into account by dividing the mean IBD sharing between groups by the mean IBD sharing within groups. A plot of IBD similarity vs. geographic distance, for different lengths of IBD segments (corresponding to different time periods) shows that Kula groups exhibit higher IBD similarities than non-Kula groups separated by the same geographic distance (Supplementary Fig. S21). To test for statistical significance and further controlling for background sharing through time by additionally including groups outside Massim, we then computed a relative IBD similarity statistic that takes into account the overall IBD sharing among all Oceanians for different IBD segment size bins. We find that relative IBD similarity statistics are significantly positive for all Massim groups (i.e. both Kula and non-Kula groups), indicating that the Massim region overall shows evidence of more contact among groups than does the rest of Oceania. Moreover, Kula groups show constantly and significantly higher relative IBD similarities than non-Kula groups (Fig. 5; Supplementary Fig. S22), even before the formation of the Kula Ring tradition ~500 ya (Shaw 2016; Irwin, et al. 2018).

Discussion

Our analyses of genome-wide data from an extensive sampling of individuals from across the Massim, together with data from other regions, indicate that all Massim groups share both Austronesian-related and Papuan-related ancestry, in agreement with the previous study of uniparental markers (van Oven, et al. 2014). However, we also find important regional distinctions within the Massim with respect to various aspects of these ancestries.

First, there is more Austronesian-related ancestry in the northern Massim (average based on ADMIXTURE of 52%; Supplementary Fig. S3; Supplementary Table S2) than in the other regions (average of 34%, with Rossel having the lowest amount, 20%). In general, there is also more Austronesian-related ancestry for the uniparental markers in the northern Massim than in the other regions, although Rossel does not have a particularly low amount (van Oven, et al.

11

2014). While the higher amount of Austronesian-related ancestry may reflect additional pulses of admixture or more prolonged contact with Austronesians in the northern Massim, the estimated admixture dates do not correlate with the amount of Austronesian ancestry (Supplementary Fig. S7). Moreover, the GLOBETROTTER results suggest single pulses of admixture (Supplementary Fig. S5), as does the linear relationship in the f3 values between the Massim and Papuans vs. those between the Massim and Austronesians (Supplementary Fig. S4) (Bergstrom, et al. 2020). We also do not see any differences in the Austronesian-related ancestry of different Massim groups (Fig. 3B; Supplementary Figs. S12 and S17). These results therefore suggest differences in the contact relationships between the indigenous Papuan groups and the incoming Austronesians in the northern Massim vs. elsewhere in the Massim, with Rossel being the least-impacted by the Austronesians, possibly due to its more remote location. It is also notable that while all of the Massim groups studied (with the exception of Rossel and 5 individuals from groups on mainland PNG) speak Austronesian languages, Austronesian ancestry is in the minority in most of the groups, and reaches a maximum of 52% in the northern Massim.

Second, the Papuan-related ancestry of Papuan-speaking Rossel Islanders is distinct from Papuan-related ancestries identified previously and elsewhere in Near Oceania. Previous studies of genome-wide data have shown a major distinction between Papuan ancestries in the PNG Highlands vs. Bougainville (Pugach, et al. 2018), and in the western vs. eastern regions of the PNG Highlands (Bergström, et al. 2017; Brucato, et al. 2021). We find support for both of these distinctions, as well as for a distinct Papuan-related ancestry on Rossel (Figs. 2 and 3; Supplementary Figs. S10C and S15) that is also present in lower amounts in other Massim groups, especially from the southern Massim. In addition to this Rossel-related Papuan ancestry, Massim groups (except Rossel) also have, at frequencies of 1-63%, another Papuan-related ancestry that is at highest frequency in the western PNG Highlands and the southern PNG lowlands (Fig. 2; Supplementary Fig. S15). Whether this ancestry predates, is associated with, or postdates the arrival of the Austronesians cannot be determined from our data. The existence of a distinct Papuan-related ancestry on Rossel could reflect initial colonization by people with this ancestry, subsequent isolation and genetic drift, or both. Further attesting to its isolation is the fact that the unique Papuan-related (non-Austronesian) language, Yélî Dnye, is spoken on Rossel; Yélî Dnye has been variously classified as either a language isolate (Levinson and Majid 2013; Stebbins, et al. 2017) or as possibly related to Anêm and Ata, two languages of West New

12

Britain in the Bismarck Archipelago (Ross 2005; Eberhard, et al. 2021). Moreover, pottery was introduced relatively late on Rossel, around 500-550 years ago, compared to around 2.8 kya elsewhere in the southern Massim (Shaw, et al. 2016). The very high amount of sharing of IBD segments within Rossel (Supplementary Fig. S15) further supports isolation and genetic drift as responsible for the development of the distinct Papuan-related ancestry on Rossel. We further note that the genetic isolation of Rossel in comparison to other Massim groups was not as apparent in a previous study of mtDNA and Y-chromosome variation in these same samples (van Oven, et al. 2014), attesting to the value of genome-wide data for studies of human population history.

Third, there is a striking contrast in patterns of sharing of IBD segments of Austronesian vs. Papuan ancestry between groups. There is extensive sharing of short IBD segments (1 to 5 cM) across the studied Massim region (Fig. 4A; Supplementary Fig. S18), and Austronesian segments are widely shared among the PNG lowland and island groups but not with the highlands (Fig. 4B), in keeping with previous observations of a lack of Austronesian-associated ancestry in the PNG highlands (Stoneking, et al. 1990; Bergström, et al. 2017; Pugach, et al. 2018; Brucato, et al. 2021). However, sharing of Papuan segments is strictly either among mainland PNG and Massim groups, or among groups from the Bismarcks and Solomons; remarkably, there is no sharing of Papuan-related IBD segments between any mainland PNG or Massim group and any group from the Bismarcks or Solomons (Fig. 4C). The Papuan ancestry-specific PCA/ADMIXTURE/TreeMix results also suggest a closer relationship between the mainland PNG and Massim groups compared to the Bismarcks and Solomons (Fig. 3A; Supplementary Figs. S15 and S16). The extensive sharing of Austronesian-related IBD segments of 1-5 cM, which roughly corresponds to a time of ~2.7 kya (Al-Asadi, et al. 2019), is in keeping with the large impact of the Austronesian expansion, which spread rapidly across the lowland and island regions of Near Oceania shortly after its arrival around 3 kya (Spriggs 1995; Shaw 2016; Kirch 2017). The distinct geographic pattern in the sharing of Papuan-related ancestry – which, to the best of our knowledge, has not been noted previously - must have a different explanation; it cannot reflect the spread of people with both Austronesian and Papuan ancestry as then there should not be a distinction between patterns of IBD sharing for Austronesian-related vs. Papuan-related segments (although we cannot rule out that the spread of Austronesian-related ancestry included a small amount of Papuan-related ancestry). We speculate that the arrival of

13

the Austronesians may have impacted the indigenous Papuan societies, as documented in a recent archaeological study (Shaw, et al. 2022), resulting in enhanced movement of Papuans within these two geographic regions (mainland/Massim, and Bismarck/Solomon Archipelagos). However, IBD sharing of segments of 1 to 5 cM reflects a time span of a few thousand years, and so the results in Fig. 4C could also reflect movement of people prior to the arrival of the Austronesians, or movement unrelated to the arrival of the Austronesians (Brucato, et al. 2021). For example, the flooding of the shallow continental shelf east of New Guinea, following the Last Glacial Maximum, appears to have led to depopulation of the current islands of the Massim region, with subsequent recolonization after ~5 kya (Shaw, Coxe, Haro, et al. 2020); perhaps the IBD-sharing results reflect these movements.

Finally, we found increased sharing of IBD segments between Massim groups practicing Kula vs. Massim groups that do not participate in this specific trading exchange ritual in the Massim (Fig. 5; Supplementary Figs. S21 and S22). Kula, made famous by Malinowski in his classic work *Argonauts of the Western Pacific* (1922), is an example of "... gift exchange with delayed reciprocity. Two kinds of objects are passed between a chain of partners in a large maritime region (the 'Massim'), providing strong networks of support, a competitive element between the participants and the thrill of adventure" (Kuehling 2021). The objects in question consist of decorated shell armbands (*mwali)* and decorated shell necklaces (*bagi* or *souvlava*) which travel in opposite directions through the participating islands (Malinowski 1922; Kuehling 2021); long-distance travel between islands is largely restricted to Kula-related voyages, leading to the expectation that Kula would have an impact on patterns of gene flow between islands. Indeed, a previous study of mtDNA and Y chromosome variation in the same samples studied here found less genetic structure between islands for the Y chromosome than for mtDNA, suggesting a potential impact of male-mediated Kula voyages (van Oven, et al. 2014). The increased sharing of IBD segments for Kula vs. non-Kula groups in the Massim is further evidence for the impact of this cultural practice on patterns of gene flow. However, we note that the increased sharing of IBD segments for Kula vs. non-Kula groups occurs across all size classes of IBD segments (Fig. 5), and thus has an approximate associated time depth of a few thousand years (Ralph and Coop 2013; Al-Asadi, et al. 2019), whereas archaeological evidence suggests a time depth for Kula of ~500 years (Shaw 2016; Irwin, et al. 2018), as well as the existence of other inter-island exchange networks that preceded the formation of Kula (Shaw

2016; Shaw and Langley 2017) and from which Kula may have developed. Thus, both archaeological and genetic evidence suggest that Kula should be viewed as arising out of a previous history of enhanced contact among the islands involved, rather than necessarily introducing novel avenues of contact that did not exist before. We observed a decline in the relative amount of IBD sharing with longer IBD segments (Fig. 5; Supplementary Figs. S21 and S22) and no apparent change in IBD sharing at the time of Kula ring formation. This may reflect either that the development of Kula was not associated with any increase in the intensity of contact between participating islands, or a lack of power to detect any such increase due to the nature of IBD sharing being restricted to fewer individuals in more recent times (Ralph and Coop 2013; Al-Asadi, et al. 2019). Nonetheless, we identified some strong sharing of long IBD segments within northern and southern Massim, suggesting extensive regional interactions in the last few hundred years.

In conclusion, our results concerning the Massim region of Papua New Guinea fill an important lacuna in genetic studies of Near Oceania. Austronesian-related ancestry varies across the region but overall is in the minority, even though Austronesian languages are in the majority. We demonstrate the existence of a distinct Papuan-related ancestry that is associated with Rossel Island and probably arose as a consequence of its isolation. In addition to the expected signal of a rapid and widespread dispersion of Austronesian-related ancestry across coastal and island Near Oceania, we also found an unexpected signal of substantial movement of people with Papuan-related ancestry that was geographically restricted, occurring exclusively between mainland PNG and the Massim region, and between the Bismarck and Solomon Archipelagoes. We speculate that these movements of people with Papuan-related ancestry may reflect a disruptive impact of the arrival of Austronesian people. Finally, we document the effect of a cultural trait, the Kula, on relative amounts of IBD sharing among participating vs. non-participating groups; the Kula thus joins other examples of cultural traits, such as residence pattern (Seielstad, et al. 1998; Oota, et al. 2001) and social stratification (Bamshad, et al. 1998), that can influence human genetic diversity.

Materials and Methods

Sample and data information

1 Saliva samples were collected in 2001 with the approval of the Medical Board of PNG and with

2 support from the Diocese of Alotau, PNG (Missionaries of the Sacred Heart, M.S.C.),

3 particularly Fr Joe Ensing (M.S.C.) and then Bishop Desmond Charles Moore (M.S.C.). Written

4 informed consent was obtained from each donor, after the project was explained and all

5 questions answered to the satisfaction of the donor. Genetic work within this study was

6 additional approved by the Ethics Commission of the University of Leipzig Medical Faculty.

7 MtDNA and Y chromosome data were published in a previous study (van Oven, et al. 2014).

8 Here, we generated genome-wide data (~1.6 million SNPs) for 255 (192 from Massim, 33 from

9 Gulf Province, and 30 from Central Province) individuals on the Illumina Infinium Multi-Ethnic

10 Global Array (MEGA); genotyping was carried out by the Laboratorio de Servicios Genómicos

11 at the Laboratorio Nacional de Genómica para la Biodiversidad, Irapuato, Mexico on our request.

12 We first merged our newly generated data with published data from Papuan populations on the

13 same SNP array (Bergström, et al. 2017), and then with published whole-genome sequencing

14 (WGS) data from Oceanians, East Asians, Europeans, and Africans (Mallick, et al. 2016; Vernot,

15 et al. 2016; Choin, et al. 2021). For the WGS data, we obtained the jointly re-called genotypes

16 from Choin et al. (Choin, et al. 2021). To avoid batch effects between the array and WGS data,

17 we extracted the overlapping sites (including both monomorphic and polymorphic sites) and

18 removed sites whose reference vs. alternative alleles were inconsistent between the array and

19 WGS data. For sites with more than two alleles, we first flipped the WGS data and then removed

20 those flipped sites that still had more than two alleles. Merging was done using PLINK v1.9

21 (Purcell, et al. 2007). For quality control, we first excluded sites with more than 5% missing data

22 in the entire dataset and sites with more than 50% missing data and/or Hardy-Weinberg

23 equilibrium p values below 0.00005 within a population group (except for groups with only one

24 individual). Then, we removed individuals with more than 5% missing data or with parents

25 speaking different languages or coming from different locations. We also filtered out individuals

26 to exclude up to 1st degree kinship pairs. Data missingness and Hardy-Weinberg equilibrium

27 were calculated using PLINK v1.9 while the individuals to be removed to avoid 1st degree

28 relatedness (kinship coefficient $>=$ 0.177) were inferred using KING (Manichaikul, et al. 2010),

29 as implemented in PLINK v2 (Chang, et al. 2015). We used KING to remove 2nd degree

30 relatedness (kinship coefficient $>=$ 0.0884) to further confirm that the components seen in

31 ADMIXTURE results were not affected by cryptic kinship. There are 776 individuals and

16

1 1,408,767 SNPs remaining after quality control (Supplementary Table S1 indicates the

2 individuals that were removed and why).

3   The sampled Massim individuals were assigned to one of 14 groups according to parental

4 geographic origins as described in the previous study (van Oven, et al. 2014): (1) Trobriand; (2)

5 Gawa; (3) Woodlark (also known as Muyuw); (4) Laughlan (also known as Budibudi); (5)

6 Fergusson (also known as Moratau), including a few individuals from nearby Dobu (also known

7 as Watoa) and Goodenough (also known as Nidula); (6) Normanby (also known as Duau); (7)

8 Milne Bay mainland eastern tip (Mainland eastern tip); (8) Misima, including some individuals

9 from nearby Paneati, Panapompom and Kimuta; (9) Western Calvados (including: Motorina,

10 Bagaman, Utian or Brooker Island, and Panaumala); (10) Eastern Calvados (including: Dadahai,

11 Kuanak or Abaga Gaheia Island, Nimoa, Panatinane or Joannet Island, Panawina, Sabarl, and

12 Wanim or Grass Island); (11) Sudest (also known as Vanatinai or Tagula); (12) Rossel (also

13 known as Yela); (13) Wanigela (and nearby settlements); and (14) Airara (and nearby

14 settlements). Together with a previously-studied group from the region (Northern, from North

15 Collingwood Bay) from Bergström et al. (Bergström, et al. 2017), there are data from 15 groups

16 over the entire Massim region including mainland and island parts. Sub-region groups were

17 defined according to geography as in the previous study (van Oven, et al. 2014): Collingwood

18 Bay (Northern, Wanigela, Airara), western Massim (Mainland eastern tip, Normanby,

19 Fergusson), northern Massim (Trobriand, Gawa, Woodlark, Laughlan), and southern Massim

20 (Misima, Western Calvados, Eastern Calvados, Sudest, Rossel). We assigned Massim groups as

21 participating in the Kula tradition or as non-participants according to previous studies (van Oven,

22 et al. 2014; Irwin, et al. 2018). The locations of the newly-studied and the reference groups are

23 shown in Fig. 1A (and Supplementary Fig. S1).

24

25 **Population structure analyses**

26 For population structure analyses, variants were pruned beforehand for linkage disequilibrium

27 using PLINK v1.9, excluding one variant from pairs with $r^2 > 0.4$ within windows of 200

28 variants and a step size of 25 variants. There were 366,193 SNPs left after pruning. Principal

29 component analysis (PCA) was done with smartpca v16000 (Patterson, et al. 2006). Two

30 individuals (papuan6278328 and papuan6278360) from Bergström et al. (Bergström, et al. 2017)

31 were identified as PCA outliers and excluded from this study. We then ran ADMIXTURE v1.3.0

17

1 (Alexander, et al. 2009) for K = 2 to K = 15 with 100 replicates for each K with random seeds.

2 We used pong v1.4.7 (Behr, et al. 2016) to visualize the 20 ADMIXTURE replicates with the

3 highest likelihoods for the major mode at each K. We also plotted K = 2 (to visualize the

4 variation of East Asian- and Papuan-related ancestries) and K = 8 (the K with the lowest cross-

5 validation error; Supplementary Fig. S8A) on a map in R v4.0.3. To support the results not due

6 to cryptically unexcluded kinship, we further analyzed results of 20 independent ADMIXTURE

7 replicates for each K from K = 2 to K = 11 (the lowest cross-validation error still occurred at K =

8 8; Supplementary Fig. S8B) using data with 2nd degree kinship excluded.

9

10 **F3 and F4 statistics**

11 We used admixr v0.9.1 (Petr, et al. 2019) from ADMIXTOOLS v7.0.2 (Patterson, et al. 2012) to

12 compute f3- and f4-statistics, with significance assessed through block jackknife resampling

13 across the genome. For f3- and f4-statistics, the African groups Mbuti and Yoruba were used

14 together as the outgroup.

15

16

1 **Data phasing**

2 We used consHap (Al Bkhetan, et al. 2019) to obtain a consensus of phasing from the results of

3 SHAPEIT v2 (Delaneau, et al. 2013), BEAGLE v5.1 (Browning, Zhou, et al. 2018), and EAGLE

4 v2 (Loh, et al. 2016) using the genetic map from the 1000 Genomes Project Phase3 (Genomes

5 Project, et al. 2015). In general, phasing accuracy can be increased by increasing the number of

6 iterations and conditioning states on which haplotype estimation is based (Browning and

7 Browning 2011). We therefore ran SHAPEIT v2 with options --burn 10, --prune 10 and --main

8 30 for iteration number with 500 conditioning states, leaving other parameters as default;

9 BEAGLE v5.1 with options burnin=12, iterations=24, phase-states=24, and ne=800000 (this is

10 smaller than the default value, but is recommended by the authors for populations with a smaller

11 effective population size, as expected for island populations in Oceania); EAGLE v2 with --

12 Kpbwt 40000 (which determines the number of conditioning haplotypes). We did not use a

13 reference panel for phasing as 1000 Genomes or other available WGS datasets lack

14 representative cohorts for Oceanian populations.

15

16 **Identity by descent (IBD) analyses**

17 We identified shared IBD blocks between each pair of individuals and homozygous-by-descent

18 (HBD; same as runs of homozygosity, ROH) blocks within each individual using refinedIBD

19 (Browning and Browning 2013). Both identified IBD and HBD blocks are considered as IBD

20 blocks in our analyses, which is analogous to pairwise shared coalescence (PSC) segments in a

21 previous study (Al-Asadi, et al. 2019). The IBD blocks within a 0.6 cM gap were merged using

22 the program merge-ibd-segments from BEAGLE utilities

23 (https://faculty.washington.edu/browning/refined-ibd), allowing only 1 inconsistent genotype

24 between the gap and block regions. We used IBD blocks at least 2 cM in length shared by

25 individuals within a population to investigate the demography of each population group. Then,

26 we used IBD blocks in 1-5 cM, 5-10 cM, and over 10 cM to investigate the sharing between

27 individuals from different populations for different time periods (Ralph and Coop 2013; Al-

28 Asadi, et al. 2019). For network visualization of the sharing between populations, the pairs with

29 average sharing $\geq$ 0.5 (i.e., on average at least half of the pairs share IBD blocks) were kept to

30 reduce noise and false positives. We summarize the patterns of shared IBD length within a

population by averaging over all comparisons between individuals, i.e. we define the average of summed IBD length L as

$$L(X) \ = \ \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} ibd(X_i, X_j)$$

here, n is the number of individuals in population X and ibd(X_i, X_j) is the length of IBD shared between individuals X_i and X_j. For the number of blocks, an analogous equation applies. Similarly, for sharing between two populations X, Y we define the average of summed IBD length L as

$$L(X,Y) \ = \ \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} ibd(X_i, Y_j)$$

where m is the number of individuals in population Y. In order to compare IBD sharing between Massim groups (Kula vs non-Kula), we use a statistic motivated by FST (Bhatia, et al. 2013). In particular, we define the similarity statistic S as

$$S(X,Y) = \frac{2\,L(X,Y)}{L(X) \ + \ L(Y)}$$

This statistic will be zero if there is no IBD sharing between X and Y (since L(X,Y) is zero), and it will be one if IBD sharing is independent of population structure, i.e L(X) = L(Y) = L(X,Y). For our analysis, we compute a matrix of pairwise S-statistics for all pairs of populations. In order to test whether there is more recent gene flow within the Kula/non-Kula group relative to the background, we use the relative similarity statistic R as

$$R = \frac{S_{within} \ - \ S_{all}}{1 - S_{all}}$$

where S_{within} and S_{all} are the average pairwise S-statistics within Kula/non-Kula populations and all Oceanian populations, respectively. If the IBD-sharing among the Kula/non-Kula groups and among all Oceanians were equal then R will be zero, and the more shared migration the Kula/non-Kula groups have, the larger R will be. We calculate the standard error of the R-statistic by jackknife resampling the chromosomes. The significance of the R-statistic between Kula- vs. non-Kula groups was estimated by comparing the observed value to the distribution of simulated values generated by randomly assigning Massim groups to either Kula- or non-Kula groups and calculating the difference in relative IBD similarity. Scripts made for analyzing IBD results are available from https://github.com/dangliu/Massim_project.

20

ChromoPainter and GLOBETROTTER analyses

To study haplotyple sharing, ChromoPainter v2 (Lawson, et al. 2012) was run on the phased dataset with sample sizes for each group randomly down-sampled to 5 (all individuals were used for groups with size < 5). We began with 10 iterations of the EM (expectation maximization) process to estimate the switch rate and global mutation probability, using chromosomes 1, 5, 10, 15, and 20. With the estimated switch and global mutation rates, we ran the chromosomal painting process for all chromosomes, which then gave the output for downstream analyses. We first attempted to paint the Massim chromosomes using them only as recipients and all of the other individuals as both donors and recipients. The EM estimation of switch rate and global mutation probability were ~142.27 and ~0.0002, respectively, which were then used as the starting values for these parameters for all donors in the painting process. To investigate differential affinities of Massim groups to PNG highlanders vs. Rossel, we also performed another run using the Massim (except for Rossel) only as recipients and all of the other individuals (including Rossel) as both donors and recipients. The EM estimation of switch rate and global mutation probability for this analysis were ~142.21 and ~0.0002, respectively.

To investigate the admixture of Austronesian- and Papuan-related ancestries, GLOBETROTTER (Hellenthal, et al. 2014) was run on the ChromoPainter output with all Massim as recipients and East Asians (including both Austronesian and non-Austronesians to increase power) and Papuan Highlanders (who showed no more than 0.0001 East Asian-related ancestry in the ADMIXTURE results for K=2) as surrogates. We first tested the certainty and potential waves of admixture events, and then estimated the major and minor sources as well as the dates of admixture. The Rossel group showed an unclear signal for admixture inference, while a single pulse of admixture was inferred for the other groups. The distributions of admixture dates were estimated via 100 bootstraps.

21

1 **MALDER admixture dating**

2 We used MALDER (https://github.com/joepickrell/malder; last access on 25th May 2022) with

3 default settings, using East Asian Austronesian and Southern Highland groups as reference

4 sources, to date the time of admixture of Austronesian and Papuan ancestries in Massim groups.

5

6 **Local ancestry inference and ancestry-specific analyses**

7 We ran the PopPhased mode of RFMix v1.5.4 (Maples, et al. 2013) to infer local ancestry across

8 the genome for each individual in our dataset with options -e 3 (three EM iterations), -G 85 (85

9 generations since the admixture event; as suggested in a previous study (Bergström, et al. 2017)),

10 -n 5 (minimum five reference haplotypes per tree node; as suggested by the authors of RFMix), -

11 -use-reference-panels-in-EM (the reference samples were used in EM iterations), and the East

12 Asians (including both Austronesian and non-Austronesians to increase power) and Papuan

13 highlanders (who showed no more than 0.0001 East Asian-related ancestry in the ADMIXTURE

14 results for K=2) as reference panels. The Papuan highlanders were randomly down-sampled to

15 50 individuals, to be the same as the sample size of East Asians. We used 0.95 as the cutoff of

16 forward-backward probability (i.e. the posterior probability of an inferred ancestry at a SNP in a

17 haplotype). We interpreted the inferred East Asian-related segments as Austronesian-related

18 segments, as indicated in f4 results (Supplementary Fig. S10). Global ancestry for each Massim

19 group was calculated from the local ancestry profiles of the corresponding individuals. Papuan-/

20 Austronesian-related ancestry-specific PCA was performed with PCAmask (Moreno-Estrada, et

21 al. 2013). Analyses of local ancestry inference, global ancestry calculation, and ancestry-specific

22 PCA were done using a pipeline from a previous study (Martin, et al. 2017); scripts for preparing

23 input files for these analyses were modified from the pipeline (available from

24 https://github.com/dangliu/Massim_project). We generated Papuan-/ Austronesian-related

25 ancestry masked genomes using our own script (available from

26 https://github.com/dangliu/Massim_project) for ancestry-specific ADMIXTURE and TreeMix

27 analyses. We ran ADMIXTURE v1.3.0 for K = 2 to K = 7 and for K = 2 to K = 6 with 20

28 replicates for each K on Austronesian and Papuan ancestry masked data, respectively. We ran

29 TreeMix v1.12 (Pickrell and Pritchard 2012) only for Austronesian ancestry masked data (as too

30 few SNPs in Papuan ancestry masked data for this analysis) with 0-2 migration events and ten

31 independent runs, and selected the topology with the highest likelihood for further investigation.

22

1 Papuan/Austronesian ancestry-specific PCA, ADMIXTURE, and TreeMix analyses were all

2 performed on East Asian and Oceanian individuals with at least 33% Papuan-/ Austronesian-

3 related ancestry based on the ADMIXTURE results for K=2, respectively. The same local

4 ancestry-inferred results were applied to IBD results to study Papuan-/ Austronesian-related

5 ancestry-specific IBD. Ancestry-specific IBD analyses were carried out by following the pipeline

6 described in a previous study (Browning, Browning, et al. 2018).

7

Data Availability

9 The genome-wide SNP array data generated in this study are available from the European

10 Genome-Phenome Archive (EGA; https://ega-archive.org), under accession code

11 EGAS00001006010. Scripts written/modified for ancestry specific analyses and IBD analyses

12 are available from https://github.com/dangliu/Massim_project.

13

Acknowledgements

24

25

1 References

2 Al Bkhetan Z, Zobel J, Kowalczyk A, Verspoor K, Goudey B. 2019. Exploring effective
3 approaches for haplotype block phasing. BMC Bioinformatics 20:540.

4 Al-Asadi H, Petkova D, Stephens M, Novembre J. 2019. Estimating recent migration and
5 population-size surfaces. PLoS genetics 15:e1007908.

6 Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in
7 unrelated individuals. Genome research 19:1655-1664.

8 Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu JM, Prasad BV, Rasanayagam
9 A, Hammer MF. 1998. Female gene flow stratifies Hindu castes. Nature 395:651-652.

10 Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S. 2016. Pong: Fast analysis and
11 visualization of latent clusters in population genetic data. Bioinformatics 32:2817-2823.

12 Bellwood P. 2007. Prehistory of the Indo-Malaysian Archipelago : Revised Edition.

13 Bergstrom A, Frantz L, Schmidt R, Ersmark E, Lebrasseur O, Girdland-Flink L, Lin AT, Stora J,
14 Sjogren KG, Anthony D, et al. 2020. Origins and genetic legacy of prehistoric dogs. Science
15 370:557-564.

16 Bergström A, Oppenheimer SJ, Mentzer AJ, Auckland K, Robson K, Attenborough R, Alpers
17 MP, Koki G, Pomat W, Siba P, et al. 2017. A Neolithic expansion, but strong genetic structure,
18 in the independent history of New Guinea. Science 357:1160-1163.

19 Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting FST: the
20 impact of rare variants. Genome Res 23:1514-1521.

21 Browning BL, Browning SR. 2013. Improving the accuracy and efficiency of identity-by-descent
22 detection in population data. Genetics 194:459-471.

23 Browning BL, Zhou Y, Browning SR. 2018. A One-Penny Imputed Genome from Next-
24 Generation Reference Panels. American Journal of Human Genetics 103:338-348.

25 Browning SR, Browning BL. 2011. Haplotype phasing: Existing methods and new
26 developments. Nature Reviews Genetics 12:703-714.

27 Browning SR, Browning BL, Daviglus ML, Durazo-Arvizu RA, Schneiderman N, Kaplan RC,
28 Laurie CC. 2018. Ancestry-specific recent effective population size in the Americas. PLoS
29 Genetics 14:1-22.

30 Brucato N, Andre M, Tsang R, Saag L, Kariwiga J, Sesuki K, Beni T, Pomat W, Muke J, Meyer
31 V, et al. 2021. Papua New Guinean Genomes Reveal the Complex Settlement of North Sahul.
32 Mol Biol Evol 38:5107-5121.

33 Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation
34 PLINK: Rising to the challenge of larger and richer datasets. GigaScience 4:1-16.

35 Choin J, Mendoza-Revilla J, Arauna LR, Cuadros-Espinoza S, Cassar O, Larena M, Ko AMS,
36 Harmant C, Laurent R, Verdu P, et al. 2021. Genomic insights into population history and
37 biological adaptation in Oceania. Nature 592:583-+.

38 Clarkson C, Jacobs Z, Marwick B, Fullagar R, Wallis L, Smith M, Roberts RG, Hayes E, Lowe
39 K, Carah X, et al. 2017. Human occupation of northern Australia by 65,000 years ago. Nature
40 547:306-310.

41 Delaneau O, Zagury JF, Marchini J. 2013. Improved whole-chromosome phasing for disease and
42 population genetic studies. Nature Methods 10:5-6.

43 Denham T. 2004. Early agriculture in the highlands of New Guinea: an assessment of Phase 1 at
44 Kuk Swamp. In A Pacific Odyssey: Archaeology and Anthropology in the Western Pacific.
45 Papers in Honour of Jim Specht. Records of the Australian Museum, Supplement 29:47–57.

24

Eberhard DM, Simons GF, Fennig CD. 2021. Ethnologue: Languages of the World, Twenty-fourth edition. SIL International.

Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. 2015. A global reference for human genetic variation. Nature 526:68-74.

Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, Myers S. 2014. A Genetic Atlas of Human Admixture History. Science 343:747-751.

Irwin G, Shaw BEN, McAlister A. 2018. The origins of the Kula Ring: Archaeological and maritime perspectives from the southern Massim and Mailu areas of Papua New Guinea. Archaeology in Oceania 54:1-16.

Kirch PV. 2017. On the road of the winds : an archaeological history of the Pacific islands before European contact: University of California Press.

Kuehling S. 2021. Dobu: Ethics of Exchange on a Massim Island, Papua New Guinea: University of Hawaii Press.

Larena M, Sanchez-Quinto F, Sjodin P, McKenna J, Ebeo C, Reyes R, Casel O, Huang JY, Hagada KP, Guilay D, et al. 2021. Multiple migrations to the Philippines during the last 50,000 years. Proc Natl Acad Sci U S A 118.

Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of Population Structure using Dense Haplotype Data. Plos Genetics 8.

Levinson SC, Majid A. 2013. The island of time: yeli dnye, the language of rossel island. Front Psychol 4:61.

Lipson M, Spriggs M, Valentin F, Bedford S, Shing R, Zinger W, Buckley H, Petchey F, Matanik R, Cheronet O, et al. 2020. Three Phases of Ancient Migration Shaped the Ancestry of Human Populations in Vanuatu. Curr Biol 30:4846-4856 e4846.

Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, Schoenherr S, Forer L, McCarthy S, Abecasis GR, et al. 2016. Reference-based phasing using the Haplotype Reference Consortium panel. Nature Genetics 48:1443-1448.

Malinowski B. 1922. Argonauts of the Western Pacific. London: Routledge and Kegan Paul.

Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao MY, Chennagiri N, Nordenfelt S, Tandon A, et al. 2016. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538:201-+.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. 2010. Robust relationship inference in genome-wide association studies. Bioinformatics 26:2867-2873.

Maples BK, Gravel S, Kenny EE, Bustamante CD. 2013. RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. American Journal of Human Genetics 93:278-288.

Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE. 2017. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am J Hum Genet 100:635-649.

Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR, Ortiz-Tello PA, Martinez RJ, Hedges DJ, Morris RW, et al. 2013. Reconstructing the population genetic history of the Caribbean. PLoS Genet 9:e1003925.

O'Connell JF, Allen J. 2015. The process, biotic impact, and global implications of the human colonization of Sahul about 47,000 years ago. Journal of Archaeological Science 56:73-84.
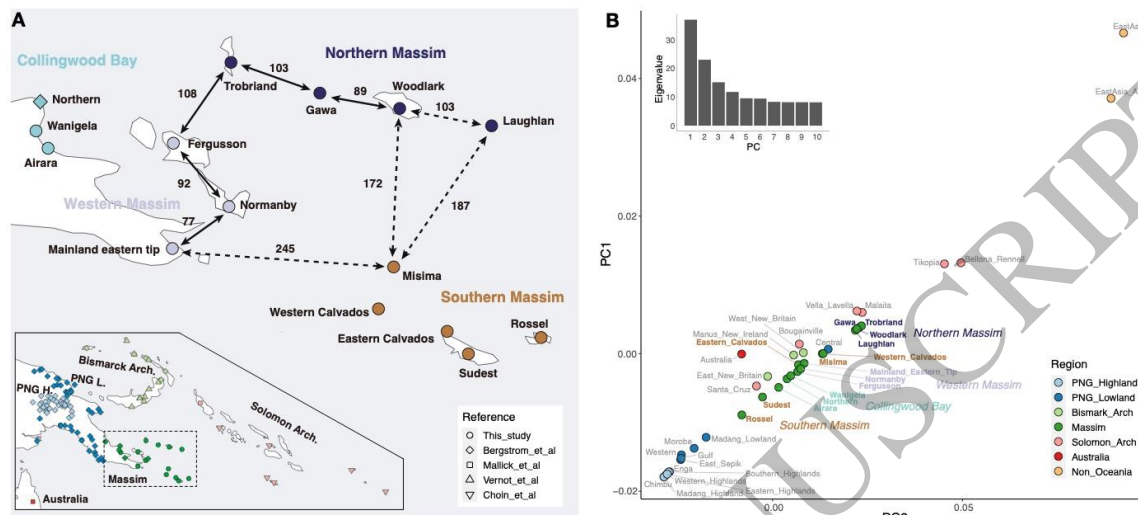
Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nature Genetics 29:20-21.

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

Patterson N, Price AL, Reich D. 2006. Population Structure and Eigenanalysis. PLoS genetics 2:e190.

Petr M, Vernot B, Kelso J. 2019. admixr —R package for reproducible analyses using ADMIXTOOLS Bioinformatics:1-2.

Pickrell JK, Pritchard JK. 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLoS Genetics 8.

Pugach I, Duggan AT, Merriwether DA, Friedlaender FR, Friedlaender JS, Stoneking M. 2018. The gateway from near into remote oceania: New insights from genome-wide data. Molecular Biology and Evolution 35:871-886.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81:559-575.

Ralph P, Coop G. 2013. The Geography of Recent Genetic Ancestry across Europe. PLoS Biology 11.

Roberts RG, Jones R, Smith MA. 1990. Thermoluminescence dating of a 50,000-year-old human occupation site in northern Australia. Nature 345:153-156.

Ross M. 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In: Pawley A, Attenborough R, Golson J, Hide R, editors. Papuan pasts: Cultural, linguistic and biological histories of Papuan-speaking peoples. Canberra Australia: Pacific Linguistics. p. 15-66.

Schiefenhövel W. 2004. Trobriands. In: Ember C, R., Ember M, editors. Encyclopedia of sex and gender: men and women in the world's cultures. New York, Boston, Dordrecht, London, Moscow: Kluwer Academic/Plenum Publishers. p. 912-921.

Seielstad MT, Minch E, Cavalli-Sforza LL. 1998. Genetic evidence for a higher female migration rate in humans. Nat Genet 20:278-280.

Seligmann CG. 1910. The Melanesians of British New Guinea. Cambridge Cambridge University Press.

Shaw B. 2019. Archaeology of the Massim Island Region, Papua New Guinea. In. Encyclopedia of Global Archaeology. Cham: Springer International Publishing. p. 1-15.

Shaw B. 2017. Late Pleistocene Colonisation of the Eastern New Guinea Islands? The Potential Implications of Robust Waisted Stone Tool Finds from Rossel Island on the Long Term Settlement Dynamics in the Massim Region. Journal of Pacific Archaeology 8:1-16.

Shaw B. 2016. The Massim Region of Papua New Guinea: A review and proposed chronology. Journal of Pacific Archaeology 7.

Shaw B, Coxe S, Haro J, Privat K, Haberle S, Hopf F, Hull E, Hawkins S, Jacobsen G. 2020. Smallest Late Pleistocene inhabited island in Australasia reveals the impact of post-glacial sea-level rise on human behaviour from 17,000 years ago. Quaternary Science Reviews 245.

Shaw B, Coxe S, Kewibu V, Haro J, Hull E, Hawkins S. 2020. 2500-year cultural sequence in the Massim region of eastern Papua New Guinea reflects adaptive strategies to small islands and changing climate regimes since Lapita settlement. Holocene 30:1075-1090.
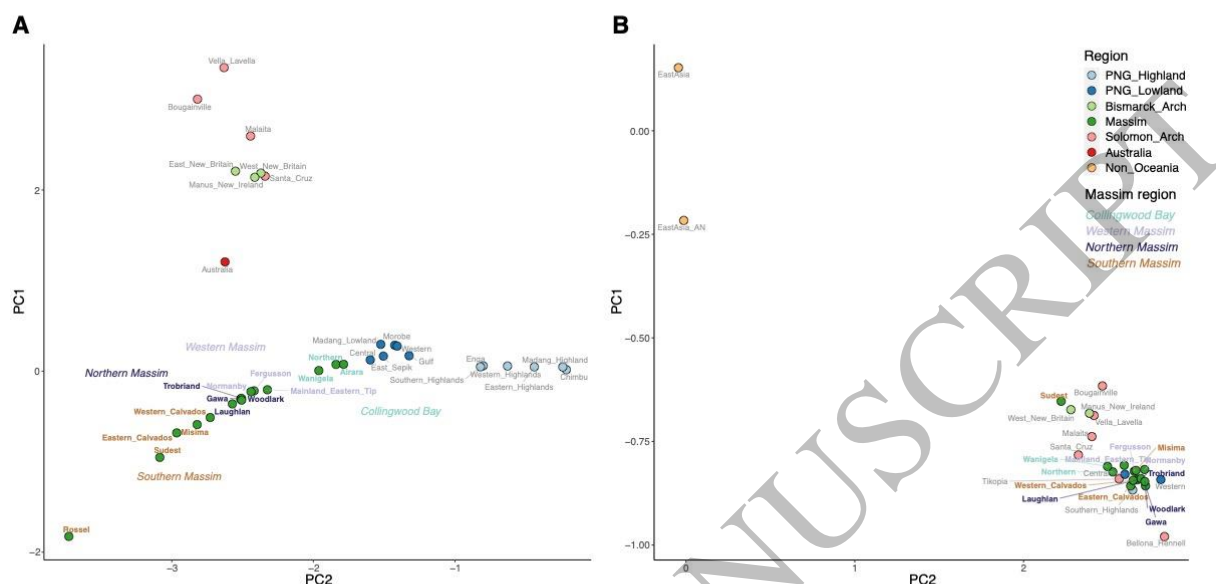
Shaw B, Field JH, Summerhayes GR, Coxe S, Coster ACF, Ford A, Haro J, Arifeae H, Hull E, Jacobsen G, et al. 2020. Emergence of a Neolithic in highland New Guinea by 5000 to 4000 years ago. Sci Adv 6:eaay4573.

Shaw B, Hawkins S, Becerra-Valdivia L, Turney CSM, Coxe S, Kewibu V, Haro J, Miamba K, Leclerc M, Spriggs M, et al. 2022. Frontier Lapita interaction with resident Papuan populations set the stage for initial peopling of the Pacific. Nature Ecology & Evolution.

Shaw B, Langley MC. 2017. Investigating the development of prehistoric cultural practices in the Massim region of eastern Papua New Guinea: Insights from the manufacture and use of shell objects in the Louisiade Archipelago. Journal of Anthropological Archaeology 48:149-165.

Shaw B, Leclerc M, Dickinson W, Spriggs M, Summerhayes GR. 2016. Identifying prehistoric trade networks in the Massim region, Papua New Guinea: Evidence from petrographic and chemical compositional pottery analyses from Rossel and Nimowa Islands in the Louisiade Archipelago. Journal of Archaeological Science: Reports 6.

Spriggs M. 1995. The Lapita culture and Austronesian prehistory in Oceania. In: Bellwood P, Fox JJ, Tyron D, editors. The Austronesians. Canberra ACT 0200, Australia: ANU E Press. p. 119-135.

Stebbins T, Evans B, Terrill A. 2017. The Papuan languages of Island Melanesia. In: Palmer B, editor. The Languages and Linguistics of the New Guinea Area : A Comprehensive Guide: De Gruyter, Inc.

Stoneking M, Jorde LB, Bhatia K, Wilson AC. 1990. Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics 124:717-733.

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

van Oven M, Brauer S, Choi Y, Ensing J, Schiefenhövel W, Stoneking M, Kayser M. 2014. Human genetics of the Kula Ring: Y-chromosome and mitochondrial DNA variation in the Massim of Papua New Guinea. European journal of human genetics : EJHG 22:1393-1403.

Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM, Dannemann M, Grote S, McCoy RC, Norton H, et al. 2016. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science 352:235-239.

Ziegler R. 2017. The Kula Ring of Bronislaw Malinowski: Simulating the Co-Evolution of an Economic and Ceremonial Exchange System Social dilemmas, institutions, and the evolution of cooperation. In: Jann B, Przepiorka W, editors: De Gruyter Oldenbourg. p. 39-60.
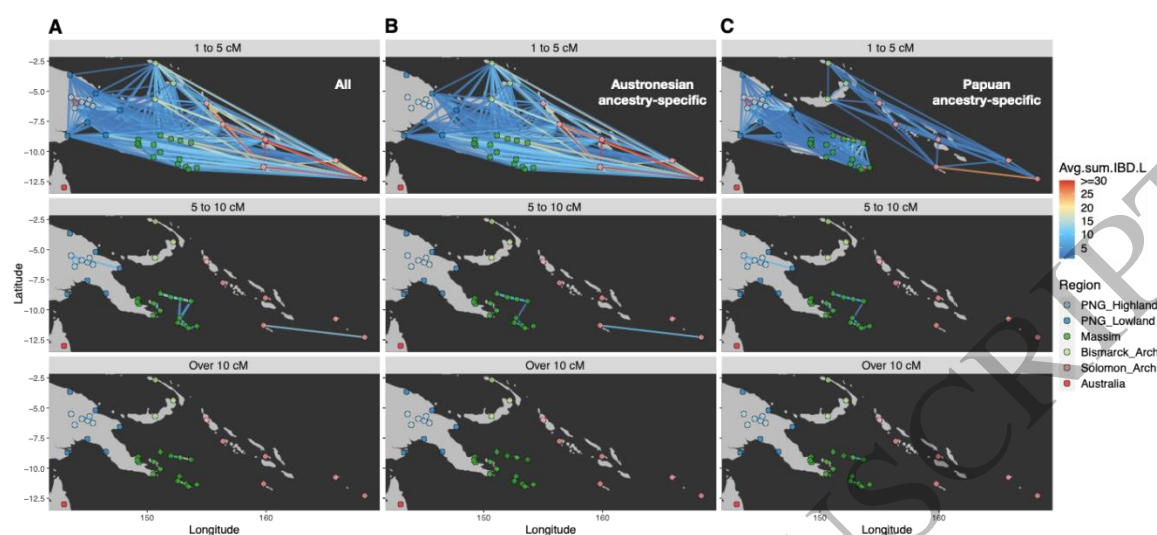
1 **Figures**

2

3

4 **Figure 1. Map showing the location of the newly-studied Massim groups and groups in the**
5 **reference dataset and Principal component analyses (PCA) of the East Asian and Oceanian**
6 **groups.** (A) The main map shows the location of Massim groups, color-coded according to sub-
7 region, while the inset map shows the location of reference groups. Dot shapes indicate the
8 corresponding publication. The solid arrows connect groups frequently involved in a Kula
9 relationship, while the dashed arrows connect groups that marginally involved in a Kula
10 relationship. The numbers indicate distances (in kilometers) between islands. (B) Plot of PC1 vs.
11 PC2 of the median position of the East Asian and Oceanian groups from Fig. S2, colored
12 according to region; Massim regions are further indicated. The eigenvalues from PC1 to PC10
13 are shown on the top left.

14

1

2

3 **Figure 2. ADMIXTURE analyses focusing on Oceanian structure.** ADMIXTURE results for

4 K=8 (the optimal K with the lowest cross-validation error; Supplementary Fig. S8A) on the top,

5 plotted on a map at the bottom. Color bars on the top indicate the region (RG), language family

6 (LF) for the Oceanian groups and Massim region (MR) for the Massim groups. The size of the

7 circle for each group on the map is proportional to the sample size.
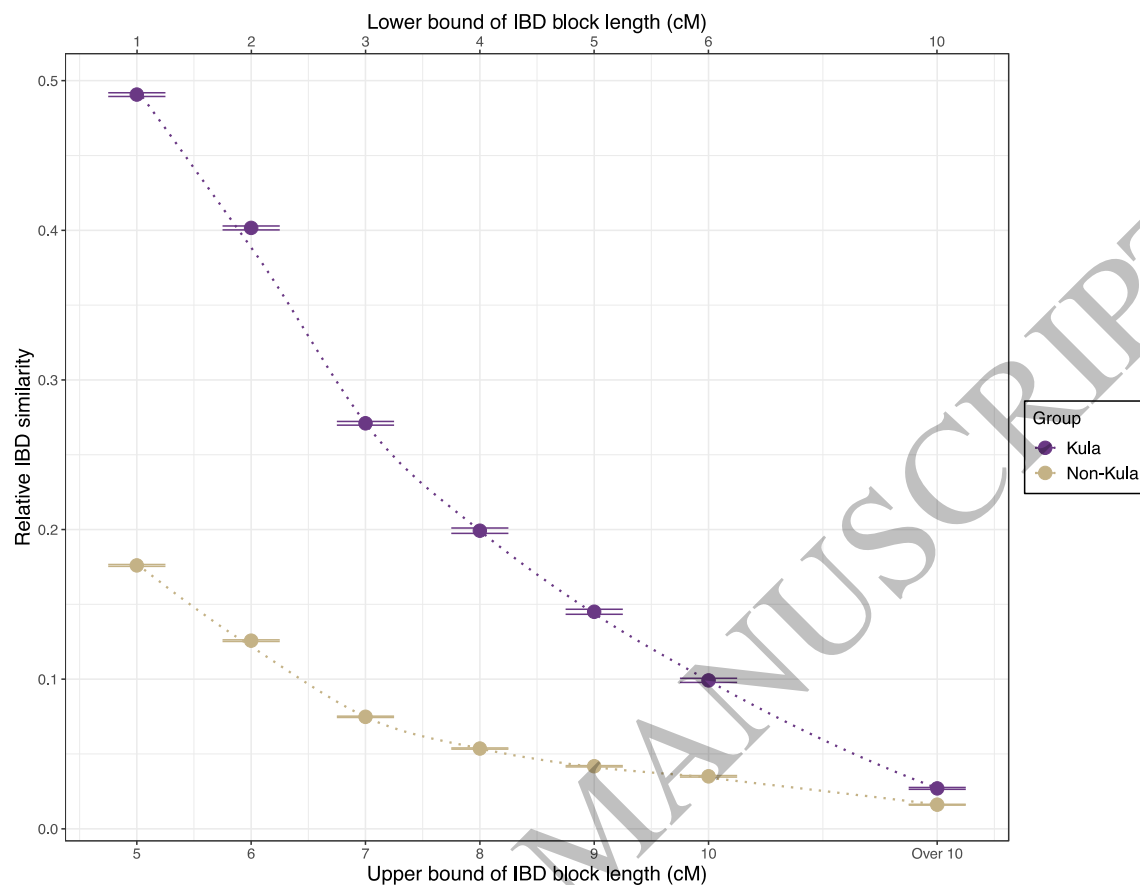
8

1

2 **Figure 3. Papuan/Austronesian ancestry-specific PCA of Oceanian (and East Asian)**
3 **groups.** Plot of median position of PC1 vs. PC2 for individuals from Oceanian groups (see Fig.
4 S13 for the PCA plot of all Oceanian individuals) and using only (A) Papuan or (B) Austronesian
5 segments identified by RFMix, colored according to region; Massim regions are further
6 highlighted.

7

1

Figure 4. IBD-sharing network for Oceanians. Network visualizations of the mean summed length of: (A) all shared IBD blocks, (B) shared Papuan ancestry-specific IBD blocks; and (C) shared Austronesian ancestry-specific IBD blocks. Identified IBD blocks are in the range of 1 to 5 cM, 5 to 10 cM, and over 10 cM, with the mean number of shared IBD blocks equal to 0.5 or more. The group points are colored according to region. The heat plot segments are proportional to the average of the summed IBD length (cM).

8

1

2 **Figure 5. Relative IBD similarities of Kula vs. non-Kula groups.** The lower and upper bound

3 lengths of the IBD blocks used in the calculations are indicated at the top and the bottom of the

4 plot, respectively. These intervals roughly correspond to ~2.7, ~1.5, ~1.1, ~0.8, ~0.7, ~0.6, and

5 ~0.2 thousand year ago (kya). The distance between the two horizontal lines on each point

6 indicate +/- three standard errors.

7