**SUPPLEMENTARY MATERIALS**

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

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

[1] Evolutionary Medicine Group, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse UMR 5288 CNRS, Université Toulouse III, Université de Toulouse. France.

[2] Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular Biology, Jakarta, Indonesia

[3] Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

[4] Asia Institute, University of Melbourne, Melbourne, Australia

[5] Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia

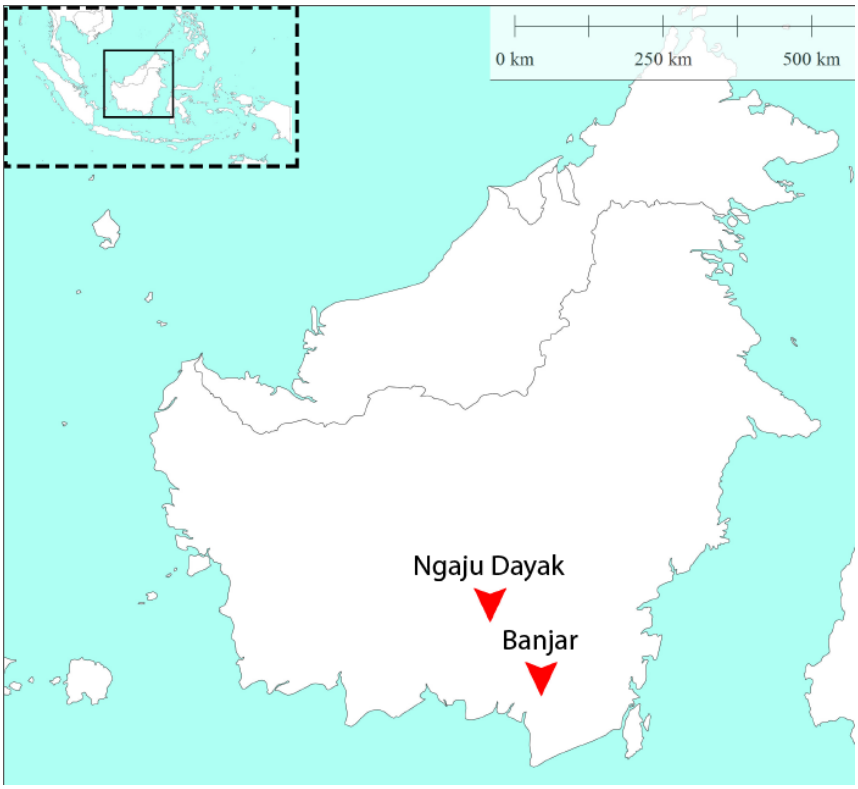
[6] Department of Biological Anthropology, University of Cambridge, Cambridge, United Kingdom

[7] Estonian Biocentre, Tartu, Estonia

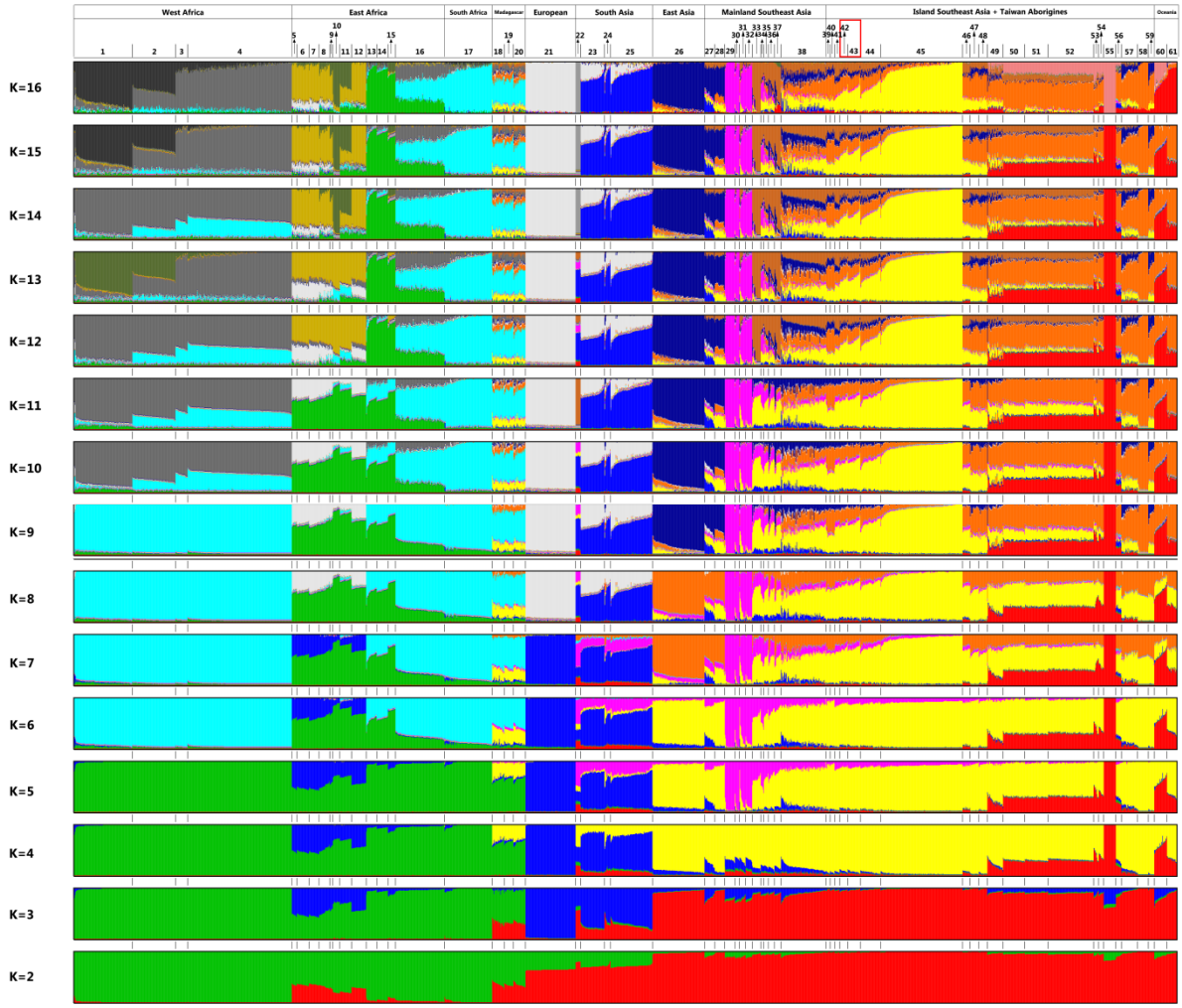
**These authors contributed equally to this work.*

[†]*Corresponding author:* F-X.R (francois-xavier.ricaut@univ-tlse3.fr)

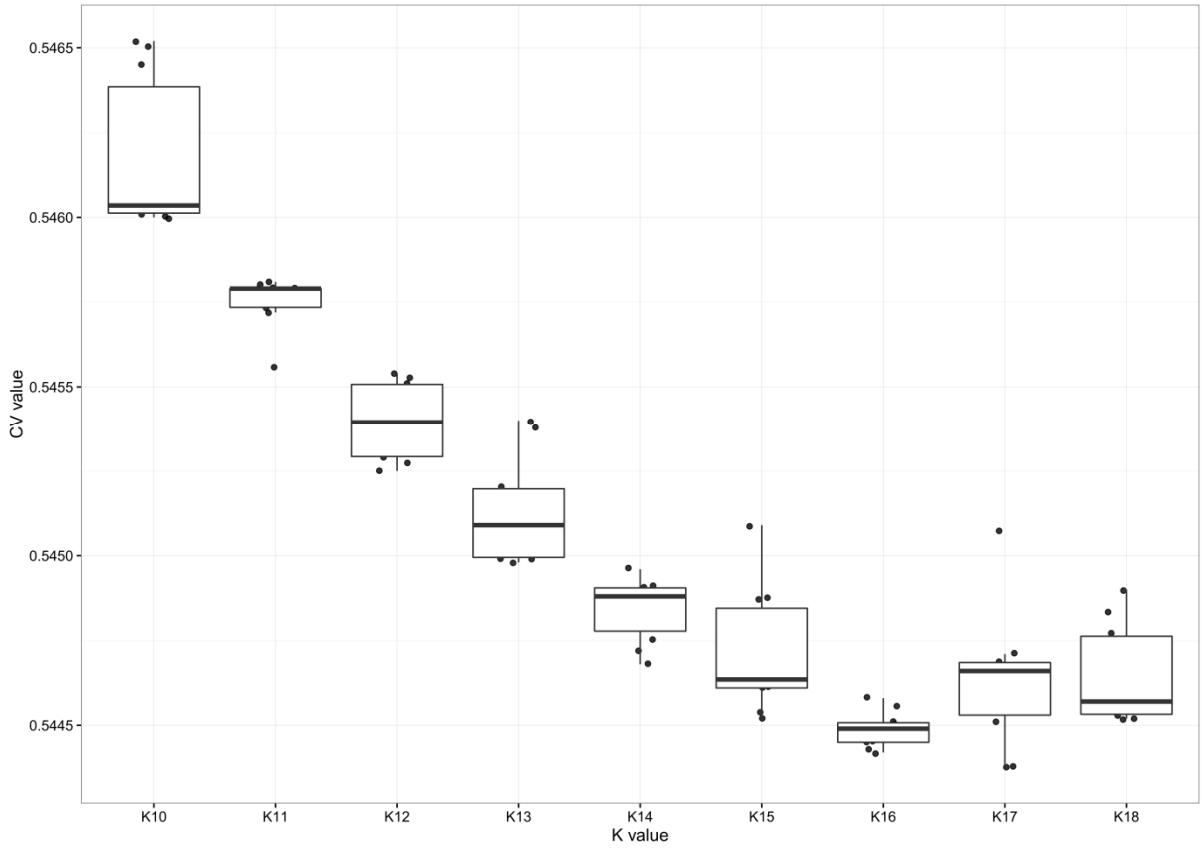
**Supplementary Figure S1** - Locations of the sampled groups in Borneo island in Indonesia. The map was generated using Global Mapper v.15 software. (http://www.bluemarblegeo.com/products/global-mapper.php)



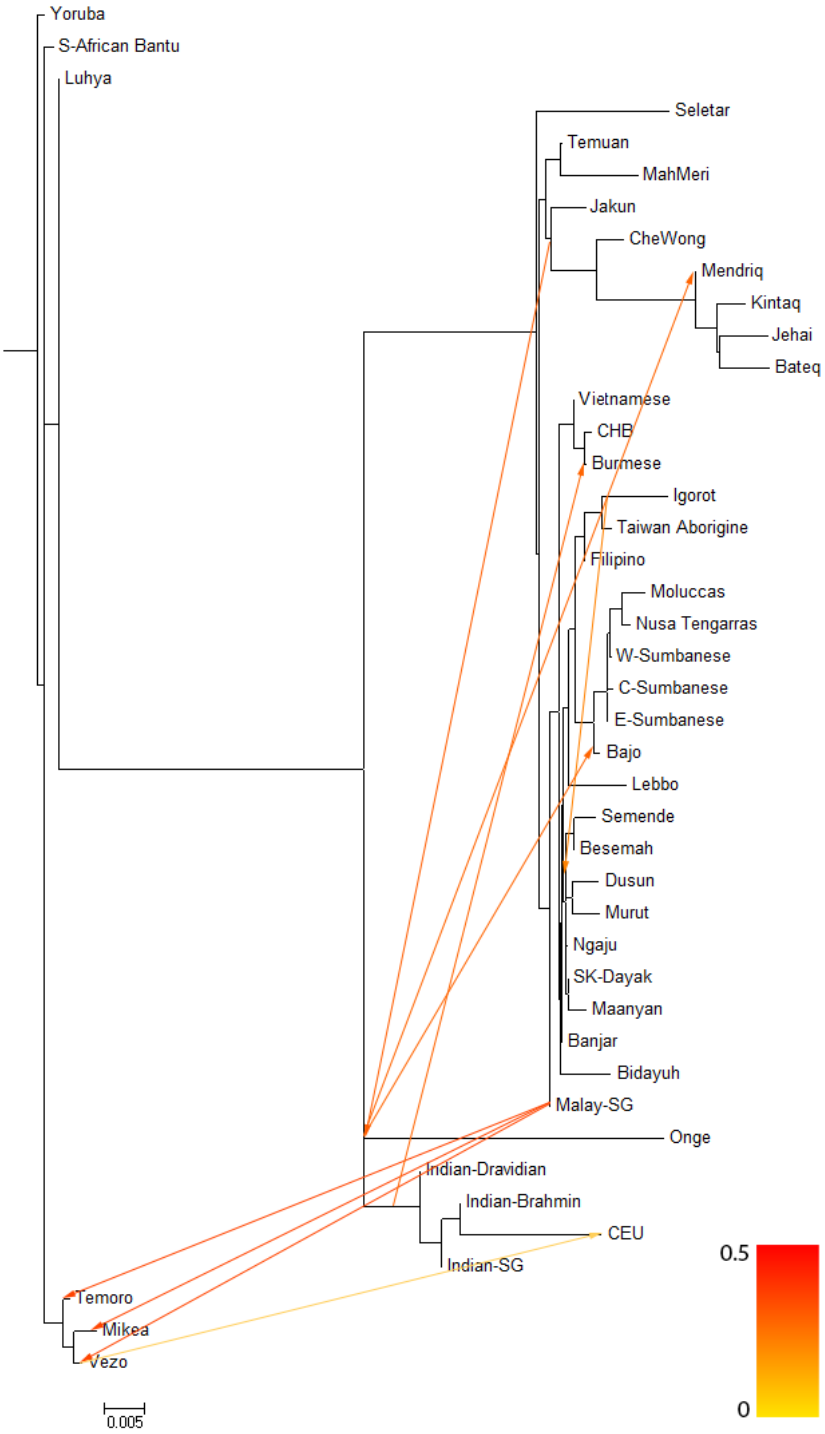
**Supplementary Figure S2** - ADMIXTURE plot from K=2 to K=16 of the low SNP density dataset using ADMIXTURE software (Alexander, et al. 2009). The Banjar and Ngaju populations are identified by the red rectangle, respectively labelled 41 and 42. The studied populations are numerically labelled as follows: 1: Gambian; 2: Mende; 3: Beninese; 4: Yoruba; 5: Afar; 6: Amhara; 7: Tygray; 8: Oromo; 9: Wolayta; 10: Ari-Blacksmith; 11: Ari-Cultivator; 12: Somalian; 13: Sudanese; 14: Anuak; 15: Gumuz; 16: Luhya; 17: South African Bantu; 18: Vezo; 19: Mikea; 20: Temoro; 21: US. European descendant; 22: Onge; 23: Brahmin; 24: Dravidian; 25: Indian from Singapore; 26: Han; 27: Burmese; 28: Vietnamese; 29: Jehai; 30: Bateq; 31: Mendriq; 32: Kintaq; 33: MahMeri; 34: CheWong; 35: Jakun; 36: Temuan; 37: Seletar; 38: Malay; 39: Besemah; 40: Semende; 41: Bidayuh; 42: Banjar; 43: Ngaju; 44: South Kalimantan Dayak; 45: Ma'anyan; 46: Lebbo; 47: Murut; 48: Dusun; 49: Bajo; 50: Central Sumbanese; 51: East Sumbanese; 52: West Sumbanese; 53: Nusa Tenggara; 54: Moluccas; 55: Papua New Guinea Highlander; 56: Philippines Negrito; 57: Filipino; 58: Igorot; 59: Taiwan Aborigenes; 60: Fiji; 61: Polynesia.


West Africa
East Africa
South Africa
Madagascar
European
South Asia
East Asia
Mainland Southeast Asia
Island Southeast Asia + Taiwan Aborigines
Oceania
K=16
K=15
K=14
K=13
K=12
K=11
K=10
K=9
K=8
K=7
K=6
K=5
K=4
K=3
K=2

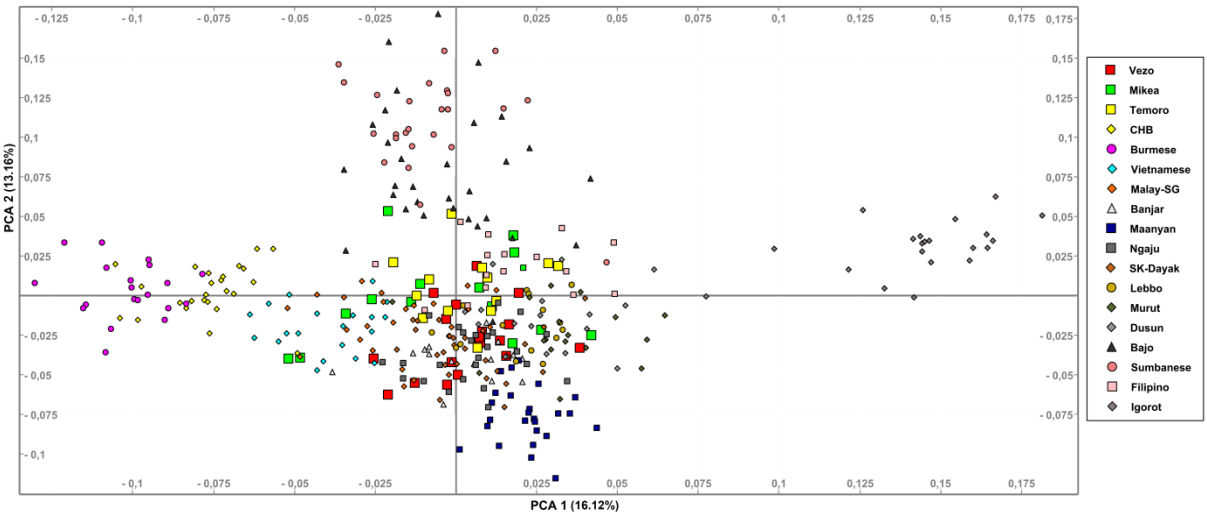
**Supplementary File S3** - Cross-validation plot generated from the ADMIXTURE analysis (Alexander et al. 2009) from K=2 to K=18 on the low density SNP dataset from 10 iterations. It shows that K=16 has the lowest median CV value thereby representing the strongly supported ADMIXTURE plot on the analysis.



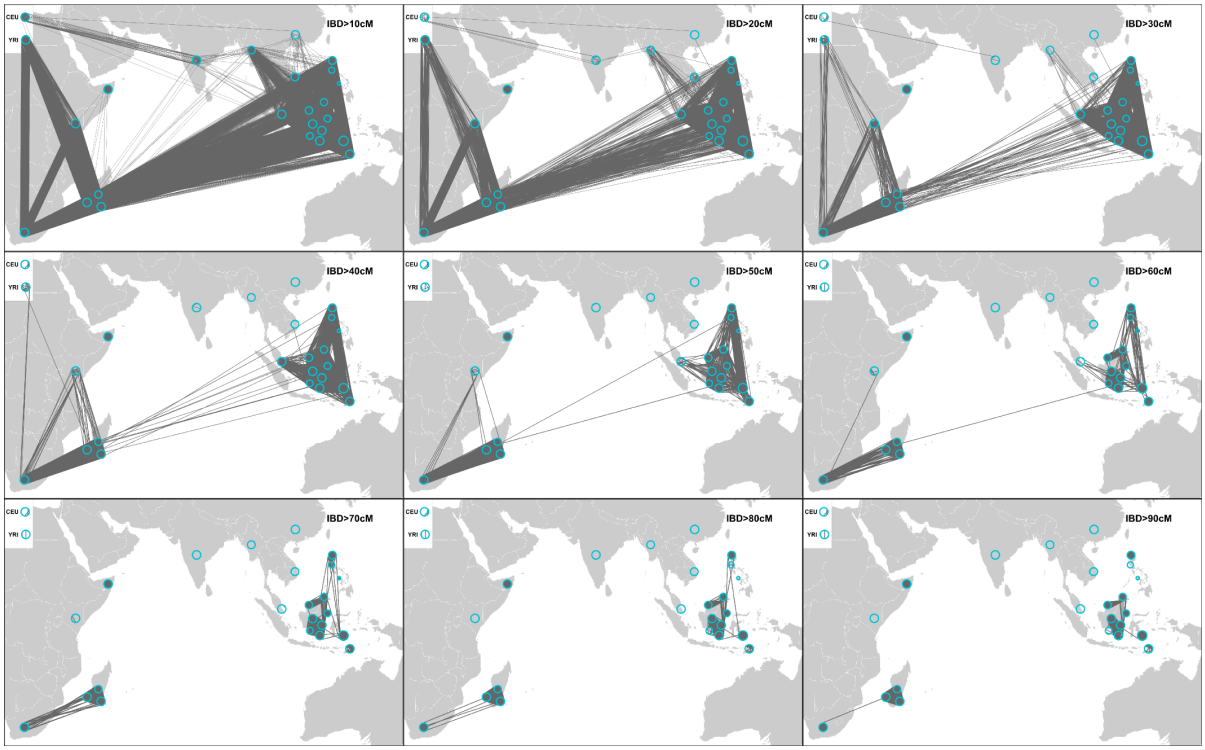
**Supplementary Figure S4** -TreeMix analysis on the low density dataset with five migration nodes (99% variability) (Pickrell and Pritchard 2012). The arrows indicate the direction of the gene flow and their colors its intensity. A strong gene flow is identified from a node clustering Southeast Asian populations into the three Malagasy populations.



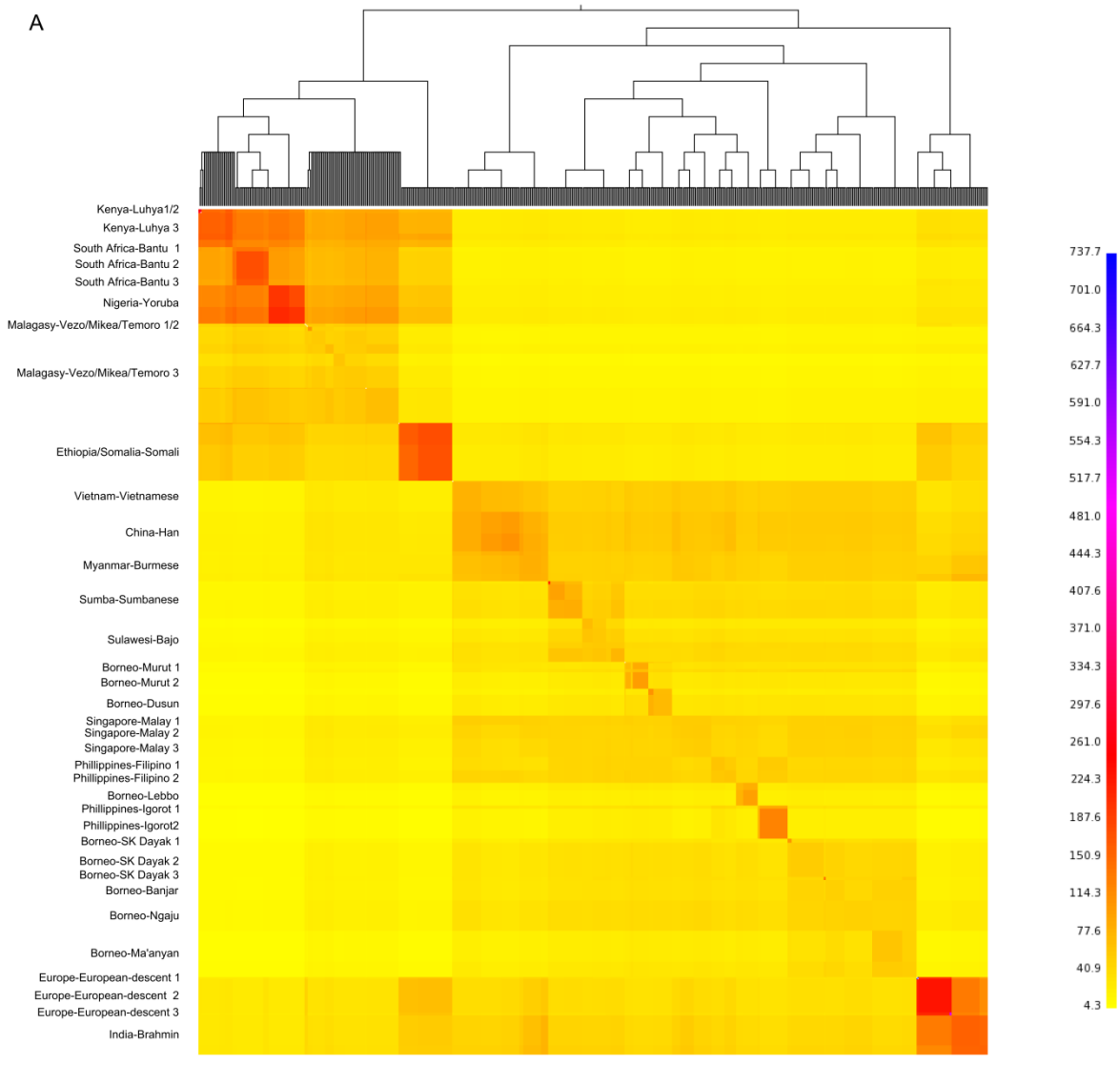
**Supplementary figure S5** - Ancestry-specific Principal Component Analysis based on masked SNPs from individuals of the high density dataset obtained after PCAdmix analysis (Brisbin, et al. 2012) and plot with EIGENSOFT v6.0.1 (Patterson, et al. 2006).

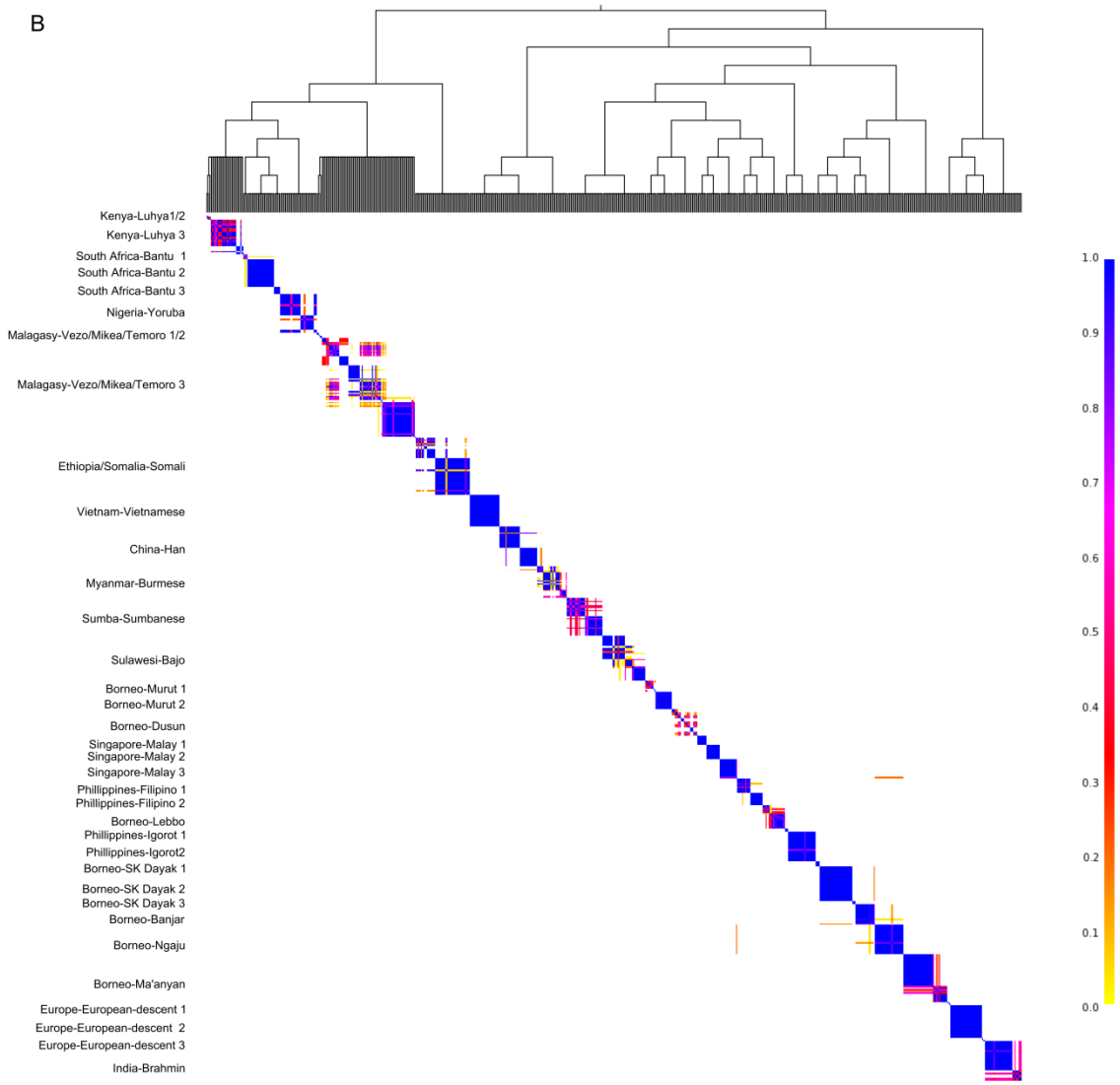


**Supplementary Figure S6** - Shared Identity-By-Descent fragments between pairs of individuals in the high density dataset, filtering for different thresholds of shared IBD fragments (from IBD > 10cM to IBD >90cM), calculated with RefinedIBD (Browning and Browning 2007; Browning and Browning 2013). Populations are represented by a blue circle which diameter is proportional to the number of individuals (from n=4 to n=31). Shared IBD fragments between each pair of individuals are represented by a black line with Cytoscape (Shannon, et al. 2003).



**Supplementary Figure S7** – Coancestry heat map (A) and pairwise coincidence matrix (B) of the individuals of the high SNP density dataset generated by fineSTRUCTURE (Lawson, et al. 2012). The color scale in (A) indicates the chunk counts between each pair of individuals calculated by Chromopainter. The color scale in (B) indicates the pairwise coincidence between each pair of individuals. Nodes are collapsed below 60% probability. For convenience, close labels are merged (e.g., Kenya-Luhya 1/2 refers to population sub-structure in two groups of the Kenya-Luhya).




B
Kenya-Luhya1/2
Kenya-Luhya 3
South Africa-Bantu 1
South Africa-Bantu 2
South Africa-Bantu 3
Nigeria-Yoruba
Malagasy-Vezo/Mikea/Temoro 1/2
Malagasy-Vezo/Mikea/Temoro 3
Ethiopia/Somalia-Somali
Vietnam-Vietnamese
China-Han
Myanmar-Burmese
Sumba-Sumbanese
Sulawesi-Bajo
Borneo-Murut 1
Borneo-Murut 2
Borneo-Dusun
Singapore-Malay 1
Singapore-Malay 2
Singapore-Malay 3
Phillippines-Filipino 1
Phillippines-Filipino 2
Borneo-Lebbo
Phillippines-Igorot 1
Phillippines-Igorot2
Borneo-SK Dayak 1
Borneo-SK Dayak 2
Borneo-SK Dayak 3
Borneo-Banjar
Borneo-Ngaju
Borneo-Ma'anyan
Europe-European-descent 1
Europe-European-descent 2
Europe-European-descent 3
India-Brahmin
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

**Supplementary Table S1** - Table of the populations included in both the low and high density SNP datasets used in the study.

| Population | Country | N (LDD) | N (HDD) | Reference |
|---|---|---|---|---|
| Yoruba | Nigeria | 206 | 25 | (International HapMap, et al. 2010) |
| Mende | Senegal | 85 | - | (International HapMap et al. 2010) |
| Gambian | Gambia | 115 | - | (International HapMap et al. 2010) |
| European | USA | 99 | 25 | (International HapMap et al. 2010) |
| Han | China | 103 | 25 | (International HapMap et al. 2010) |
| Luhya | Kenya | 97 | 25 | (International HapMap et al. 2010) |
| Beninese | Benin | 24 | - | Fortes-Lima, personal communication |
| Afar | Ethiopia | 10 | - | (Pagani, et al. 2012) |
| Tygray | Ethiopia | 20 | - | (Pagani et al. 2012) |
| Amhara | Ethiopia | 24 | - | (Pagani et al. 2012) |
| Wolayta | Ethiopia | 6 | - | (Pagani et al. 2012) |
| Oromo | Ethiopia | 21 | - | (Pagani et al. 2012) |
| Ari-Cutivator | Ethiopia | 21 | - | (Pagani et al. 2012) |
| Ari-Blacksmith | Ethiopia | 14 | - | (Pagani et al. 2012) |
| Gumuz | Ethiopia | 15 | - | (Pagani et al. 2012) |
| Anuak | Ethiopia | 22 | - | (Pagani et al. 2012) |
| Sudanese | Sudan | 21 | - | (Pagani et al. 2012) |
| Somalian | Somalia | 23 | 23 | (Pagani et al. 2012) |
| Bantu | South Africa | 94 | 25 | (May, et al. 2013) |
| Mikea | Malagasy | 18 | 18 | (Pierron, et al. 2014) |
| Temoro | Malagasy | 24 | 24 | (Pierron et al. 2014) |
| Vezo | Malagasy | 24 | 24 | (Pierron et al. 2014) |
| Lebbo | Indonesia (Borneo) | 15 | 15 | (Pierron et al. 2014) |
| Bajo | Indonesia (Sulawesi) | 31 | 31 | (Pierron et al. 2014) |
| Indian | Singapore | 83 | - | (Teo, et al. 2009) |
| Malay | Singapore | 89 | 25 | (Teo et al. 2009) |
| Jehai | Malaysia | 20 | - | (Aghakhanian, et al. 2015) |
| Bateq | Malaysia | 9 | - | (Aghakhanian et al. 2015) |
| Mendriq | Malaysia | 12 | - | (Aghakhanian et al. 2015) |
| Kintaq | Malaysia | 13 | - | (Aghakhanian et al. 2015) |
| MahMeri | Malaysia | 17 | - | (Aghakhanian et al. 2015) |
| CheWong | Malaysia | 6 | - | (Aghakhanian et al. 2015) |
| Jakun | Malaysia | 9 | - | (Aghakhanian et al. 2015) |
| Temuan | Malaysia | 12 | - | (Aghakhanian et al. 2015) |
| Seletar | Malaysia | 13 | - | (Aghakhanian et al. 2015) |
| Onge | Andaman | 10 | - | (Reich, et al. 2011) |

| | | | | |
|---|---|---|---|---|
| Dravidian | India | 12 | - | (Reich et al. 2011) |
| Besemah | Indonesia (Sumatra) | 8 | - | (Reich et al. 2011) |
| Semende | Indonesia (Sumatra) | 9 | - | (Reich et al. 2011) |
| Bidayuh | Malaysia (Borneo) | 10 | - | (Reich et al. 2011) |
| Nusa Tenggara | Indonesia | 10 | - | (Reich et al. 2011) |
| Moluccas | Indonesia | 10 | - | (Reich et al. 2011) |
| PNG Highlander | Papua New Guinea | 24 | - | (Reich et al. 2011) |
| Filipino** | Philippines | 16 | - | (Reich et al. 2011) |
| Philippines Negrito | Philippines | 11 | - | (Reich et al. 2011) |
| Taiwanese Aborigines | Taiwan | 12 | - | (Reich et al. 2011) |
| Fiji | Fiji | 25 | - | (Reich et al. 2011) |
| Polynesian | Polynesia | 19 | - | (Reich et al. 2011) |
| Sumbanese* | Indonesia (Sumba) | 180 | 25 | (Cox, et al. forthcoming) |
| Brahmin | India | 47 | 25 | (Mörseburg, et al. forthcoming) |
| Burmese | Myanmar | 20 | 20 | (Mörseburg et al. forthcoming) |
| Vietnamese | Vietnam | 20 | 20 | (Mörseburg et al. forthcoming) |
| Igorot | Philippines | 21 | 21 | (Mörseburg et al. forthcoming) |
| Filipino** | Philippines | 16 | 16 | (Mörseburg et al. forthcoming) |
| Dusun | Malaysia (Borneo) | 17 | 17 | (Mörseburg et al. forthcoming) |
| Murut | Malaysia (Borneo) | 17 | 17 | (Mörseburg et al. forthcoming) |
| Ma'anyan | Indonesia (Borneo) | 162 | 25 | (Kusuma, et al. 2016) |
| SK.Dayak | Indonesia (Borneo) | 40 | 25 | (Kusuma et al. 2016) |
| Ngaju | Indonesia (Borneo) | 25 | 25 | *this study* |
| Banjar | Indonesia (Borneo) | 16 | 16 | *this study* |

Note: LDD: Low Density of SNP Dataset (40,272 SNPs); HDD: High Density of SNP Dataset (374,189 SNPs). *: Sumbanese of the LDD were sub-grouped in 3 locations of sampling for some analyses: West (W-Sumba), Central (C-Sumba), East (E-Sumba).; **: For some analyses Filipino individuals from (Reich et al. 2011) and (Mörseburg et al. forthcoming) were pooled.

**Supplementary Table S2** - Results of *f3*-statistics showing the Z-score <-2 obtained for each combination on populations in the high density dataset to test for admixture. Only the 15 lowest Z-scores are shown for each tested group.

| Tested group | Parental Pop A | Parental Pop B | f3 | SD | Z-score | Tested group | Parental Pop A | Parental Pop B | f3 | SD | Z-score |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Bajo | Luhya | Igorot | -0,00074 | 0,00015 | -4,83578 | Mikea | SK-Dayak | Yoruba | -0,00995 | 0,00010 | -100,59000 |
| Bajo | S-African_Bantu | Igorot | -0,00075 | 0,00016 | -4,79533 | Mikea | S-African_Bantu | Malay-SG | -0,00946 | 0,00009 | -99,94540 |
| Bajo | Yoruba | Igorot | -0,00073 | 0,00015 | -4,70864 | Mikea | Ngaju | Yoruba | -0,00985 | 0,00010 | -99,50600 |
| Bajo | Eth-Somalian | Igorot | -0,00064 | 0,00015 | -4,20585 | Mikea | Yoruba | Maanyan | -0,01003 | 0,00010 | -99,44410 |
| Bajo | Somalian | Igorot | -0,00063 | 0,00015 | -4,19374 | Mikea | Yoruba | Murut | -0,01017 | 0,00010 | -99,15970 |
| Bajo | Gyan | Igorot | -0,00042 | 0,00014 | -2,91720 | Mikea | S-African_Bantu | Dusun | -0,01009 | 0,00010 | -99,13510 |
| Bajo | CEU | Igorot | -0,00038 | 0,00016 | -2,34123 | Mikea | Vietnamese | S-African_Bantu | -0,00973 | 0,00010 | -98,79300 |
| Banjar | Gyan | Igorot | -0,00117 | 0,00009 | -12,53420 | Mikea | S-African_Bantu | Luzon | -0,00971 | 0,00010 | -98,31260 |
| Banjar | Gyan | Maanyan | -0,00086 | 0,00007 | -11,70650 | Mikea | S-African_Bantu | Banjar | -0,00979 | 0,00010 | -98,24890 |
| Banjar | Somalian | Igorot | -0,00112 | 0,00010 | -11,08290 | Mikea | Malay-SG | Yoruba | -0,00939 | 0,00010 | -97,83200 |
| Banjar | Eth-Somalian | Igorot | -0,00113 | 0,00010 | -10,98400 | Mikea | Dusun | Yoruba | -0,01003 | 0,00010 | -97,10260 |
| Banjar | CEU | Maanyan | -0,00089 | 0,00008 | -10,75080 | Ngaju | Gyan | Maanyan | -0,00033 | 0,00006 | -5,42691 |
| Banjar | CEU | Igorot | -0,00113 | 0,00010 | -10,74080 | Ngaju | CEU | Maanyan | -0,00030 | 0,00007 | -4,34292 |
| Banjar | Luhya | Igorot | -0,00110 | 0,00011 | -10,40430 | Ngaju | Gyan | Igorot | -0,00037 | 0,00009 | -4,23838 |
| Banjar | S-African_Bantu | Igorot | -0,00110 | 0,00011 | -10,30480 | Ngaju | Eth-Somalian | Maanyan | -0,00021 | 0,00006 | -3,27339 |
| Banjar | Yoruba | Igorot | -0,00110 | 0,00011 | -10,24280 | Ngaju | Eth-Somalian | Igorot | -0,00030 | 0,00010 | -3,01367 |
| Banjar | Eth-Somalian | Maanyan | -0,00077 | 0,00008 | -9,87042 | Ngaju | Maanyan | Somalian | -0,00019 | 0,00006 | -2,98514 |
| Banjar | Maanyan | Somalian | -0,00076 | 0,00008 | -9,71601 | Ngaju | Luhya | Igorot | -0,00030 | 0,00010 | -2,95500 |
| Banjar | Gyan | Murut | -0,00074 | 0,00008 | -9,13794 | Ngaju | Somalian | Igorot | -0,00028 | 0,00010 | -2,92635 |
| Banjar | S-African_Bantu | Maanyan | -0,00073 | 0,00008 | -8,92003 | Ngaju | Yoruba | Igorot | -0,00029 | 0,00010 | -2,85123 |
| Banjar | Maanyan | Luhya | -0,00069 | 0,00008 | -8,65876 | Ngaju | S-African_Bantu | Igorot | -0,00028 | 0,00010 | -2,68118 |
| Banjar | CEU | Murut | -0,00081 | 0,00009 | -8,63599 | Ngaju | CEU | Igorot | -0,00026 | 0,00010 | -2,60884 |
| Burmese | Gyan | CHB | -0,00325 | 0,00006 | -56,17750 | Ngaju | S-African_Bantu | Maanyan | -0,00017 | 0,00007 | -2,50506 |
| Burmese | Vietnamese | Gyan | -0,00306 | 0,00006 | -51,65090 | Ngaju | Maanyan | Luhya | -0,00017 | 0,00007 | -2,43966 |
| Burmese | CEU | CHB | -0,00316 | 0,00007 | -43,60350 | Ngaju | Yoruba | Maanyan | -0,00016 | 0,00007 | -2,32433 |
| Burmese | Vietnamese | CEU | -0,00311 | 0,00007 | -42,84930 | Ngaju | Gyan | Murut | -0,00011 | 0,00007 | -1,43415 |
| Burmese | CHB | Somalian | -0,00280 | 0,00007 | -41,56630 | SK-Dayak | Maanyan | Burmese | -0,00031 | 0,00004 | -7,90256 |
| Burmese | Eth-Somalian | CHB | -0,00283 | 0,00007 | -40,63090 | SK-Dayak | Gyan | Maanyan | -0,00044 | 0,00006 | -7,43457 |
| Burmese | Vietnamese | Somalian | -0,00260 | 0,00007 | -37,29110 | SK-Dayak | Maanyan | Sumbanese | -0,00030 | 0,00004 | -7,41106 |
| Burmese | Gyan | Murut | -0,00267 | 0,00007 | -36,73700 | SK-Dayak | Vietnamese | Maanyan | -0,00027 | 0,00004 | -7,27878 |
| Burmese | Vietnamese | Eth-Somalian | -0,00260 | 0,00007 | -36,71400 | SK-Dayak | Malay-SG | Maanyan | -0,00023 | 0,00003 | -7,08378 |
| Burmese | Gyan | Igorot | -0,00290 | 0,00008 | -36,48650 | SK-Dayak | Maanyan | Bajo | -0,00035 | 0,00005 | -7,02357 |
| Burmese | CHB | Luhya | -0,00257 | 0,00007 | -35,51950 | SK-Dayak | CEU | Maanyan | -0,00041 | 0,00007 | -6,08851 |
| Burmese | S-African_Bantu | CHB | -0,00261 | 0,00007 | -35,41820 | SK-Dayak | Maanyan | CHB | -0,00024 | 0,00004 | -5,80104 |
| Burmese | Yoruba | CHB | -0,00257 | 0,00007 | -34,33380 | SK-Dayak | Gyan | Igorot | -0,00049 | 0,00008 | -5,79169 |
| Burmese | SK-Dayak | Gyan | -0,00220 | 0,00006 | -34,22450 | SK-Dayak | Temoro | Maanyan | -0,00028 | 0,00005 | -5,32602 |
| Burmese | Gyan | Maanyan | -0,00233 | 0,00007 | -33,86420 | SK-Dayak | Vezo | Maanyan | -0,00027 | 0,00005 | -5,20714 |
| Luzon | Gyan | Igorot | -0,00226 | 0,00008 | -26,73450 | SK-Dayak | Maanyan | Murut | -0,00021 | 0,00004 | -5,04424 |
| Luzon | Somalian | Igorot | -0,00228 | 0,00009 | -24,00030 | SK-Dayak | Maanyan | Luhya | -0,00034 | 0,00007 | -4,91390 |
| Luzon | CEU | Igorot | -0,00233 | 0,00010 | -23,76760 | SK-Dayak | Luhya | Igorot | -0,00049 | 0,00010 | -4,90414 |
| Luzon | Eth-Somalian | Igorot | -0,00227 | 0,00010 | -23,51470 | SK-Dayak | S-African_Bantu | Maanyan | -0,00034 | 0,00007 | -4,84829 |
| Luzon | Luhya | Igorot | -0,00229 | 0,00010 | -22,86680 | Somalian | Yoruba | CEU | -0,00478 | 0,00010 | -47,79240 |
| Luzon | Yoruba | Igorot | -0,00229 | 0,00010 | -22,43360 | Somalian | CEU | Luhya | -0,00448 | 0,00009 | -47,43610 |
| Luzon | S-African_Bantu | Igorot | -0,00227 | 0,00010 | -22,24460 | Somalian | S-African_Bantu | CEU | -0,00443 | 0,00010 | -44,19020 |
| Luzon | Vezo | Igorot | -0,00161 | 0,00008 | -20,85780 | Somalian | Gyan | Luhya | -0,00221 | 0,00009 | -25,10260 |
| Luzon | Mikea | Igorot | -0,00166 | 0,00008 | -20,39340 | Somalian | Gyan | Yoruba | -0,00237 | 0,00009 | -24,97390 |
| Luzon | Temoro | Igorot | -0,00160 | 0,00008 | -20,18860 | Somalian | S-African_Bantu | Gyan | -0,00205 | 0,00009 | -21,81550 |
| Luzon | Burmese | Igorot | -0,00112 | 0,00006 | -18,23300 | Somalian | Burmese | Luhya | -0,00036 | 0,00010 | -3,47382 |
| Luzon | Igorot | Sumbanese | -0,00073 | 0,00006 | -12,37340 | Somalian | Yoruba | Burmese | -0,00034 | 0,00011 | -2,98567 |
| Luzon | Malay-SG | Igorot | -0,00069 | 0,00006 | -12,16120 | Somalian | CHB | Luhya | -0,00013 | 0,00011 | -1,19885 |
| Luzon | CHB | Igorot | -0,00059 | 0,00006 | -9,63134 | Sumbanese | Luhya | Igorot | -0,00065 | 0,00011 | -5,87579 |
| Luzon | Vietnamese | Igorot | -0,00046 | 0,00006 | -7,76522 | Sumbanese | S-African_Bantu | Igorot | -0,00065 | 0,00011 | -5,85895 |
| Gyan | CEU | Burmese | -0,00234 | 0,00008 | -30,74300 | Sumbanese | Yoruba | Igorot | -0,00065 | 0,00011 | -5,84816 |
| Gyan | Malay-SG | CEU | -0,00235 | 0,00008 | -28,72070 | Sumbanese | Eth-Somalian | Igorot | -0,00047 | 0,00010 | -4,53064 |
| Gyan | CEU | Sumbanese | -0,00240 | 0,00009 | -28,18800 | Sumbanese | Somalian | Igorot | -0,00045 | 0,00010 | -4,34370 |
| Gyan | Ngaju | CEU | -0,00242 | 0,00009 | -27,79990 | Sumbanese | Gyan | Igorot | -0,00021 | 0,00009 | -2,35079 |
| Gyan | SK-Dayak | CEU | -0,00242 | 0,00009 | -27,08320 | Sumbanese | CEU | Igorot | -0,00012 | 0,00011 | -1,08230 |
| Gyan | Banjar | CEU | -0,00235 | 0,00009 | -26,65450 | Temoro | S-African_Bantu | Ngaju | -0,01130 | 0,00009 | -124,99600 |
| Gyan | Vietnamese | CEU | -0,00238 | 0,00009 | -26,49640 | Temoro | S-African_Bantu | SK-Dayak | -0,01141 | 0,00009 | -124,52900 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Gyan | Dusun | CEU | -0,00248 | 0,00009 | -26,47620 | Temoro | SK-Dayak | Yoruba | -0,01133 | 0,00009 | -123,75300 |
| Gyan | CEU | Maanyan | -0,00238 | 0,00009 | -26,35860 | Temoro | Ngaju | Yoruba | -0,01122 | 0,00009 | -123,16300 |
| Gyan | CEU | Murut | -0,00241 | 0,00009 | -25,80560 | Temoro | S-African_Bantu | Maanyan | -0,01147 | 0,00009 | -122,94300 |
| Gyan | CEU | Lebbo | -0,00240 | 0,00009 | -25,68800 | Temoro | S-African_Bantu | Banjar | -0,01118 | 0,00009 | -122,91200 |
| Gyan | CEU | Luzon | -0,00224 | 0,00009 | -25,41830 | Temoro | S-African_Bantu | Malay-SG | -0,01083 | 0,00009 | -122,90400 |
| Gyan | CEU | CHB | -0,00225 | 0,00009 | -24,96840 | Temoro | Banjar | Yoruba | -0,01112 | 0,00009 | -121,74100 |
| Gyan | CEU | Bajo | -0,00235 | 0,00010 | -24,66920 | Temoro | Yoruba | Maanyan | -0,01137 | 0,00009 | -121,48900 |
| Gyan | CEU | Igorot | -0,00231 | 0,00010 | -23,36910 | Temoro | Malay-SG | Yoruba | -0,01073 | 0,00009 | -121,35500 |
| Luhya | Eth-Somalian | Yoruba | -0,00072 | 0,00005 | -15,55730 | Temoro | Yoruba | Murut | -0,01149 | 0,00009 | -121,07700 |
| Luhya | Yoruba | Somalian | -0,00071 | 0,00005 | -15,36080 | Temoro | S-African_Bantu | Murut | -0,01157 | 0,00010 | -120,65900 |
| Luhya | Yoruba | CEU | -0,00101 | 0,00007 | -14,90660 | Temoro | Vietnamese | S-African_Bantu | -0,01114 | 0,00009 | -120,07700 |
| Luhya | S-African_Bantu | Somalian | -0,00062 | 0,00004 | -13,93520 | Temoro | S-African_Bantu | Luzon | -0,01108 | 0,00009 | -119,93300 |
| Luhya | Gyan | Yoruba | -0,00087 | 0,00006 | -13,82580 | Temoro | Yoruba | Luzon | -0,01099 | 0,00009 | -118,51300 |
| Luhya | S-African_Bantu | Eth-Somalian | -0,00059 | 0,00005 | -12,47290 | Vezo | S-African_Bantu | Ngaju | -0,01131 | 0,00009 | -126,64300 |
| Luhya | Yoruba | Burmese | -0,00068 | 0,00007 | -10,42080 | Vezo | S-African_Bantu | SK-Dayak | -0,01140 | 0,00009 | -125,42300 |
| Luhya | Malay-SG | Yoruba | -0,00068 | 0,00007 | -10,24830 | Vezo | S-African_Bantu | Maanyan | -0,01147 | 0,00009 | -124,43700 |
| Luhya | Yoruba | Sumbanese | -0,00067 | 0,00007 | -10,19040 | Vezo | S-African_Bantu | Malay-SG | -0,01079 | 0,00009 | -123,28900 |
| Luhya | Dusun | Yoruba | -0,00070 | 0,00007 | -10,17090 | Vezo | S-African_Bantu | Banjar | -0,01118 | 0,00009 | -122,99600 |
| Luhya | SK-Dayak | Yoruba | -0,00068 | 0,00007 | -10,16430 | Vezo | SK-Dayak | Yoruba | -0,01132 | 0,00009 | -122,00400 |
| Luhya | Ngaju | Yoruba | -0,00068 | 0,00007 | -10,06140 | Vezo | Ngaju | Yoruba | -0,01123 | 0,00009 | -121,63800 |
| Luhya | Yoruba | CHB | -0,00068 | 0,00007 | -10,01850 | Vezo | S-African_Bantu | Murut | -0,01161 | 0,00010 | -121,03500 |
| Luhya | Yoruba | Luzon | -0,00067 | 0,00007 | -9,95678 | Vezo | S-African_Bantu | Luzon | -0,01109 | 0,00009 | -120,72900 |
| Luhya | Vietnamese | Yoruba | -0,00068 | 0,00007 | -9,92264 | Vezo | Yoruba | Maanyan | -0,01138 | 0,00009 | -120,61700 |
| Malay-SG | Gyan | Maanyan | -0,00165 | 0,00006 | -26,43780 | Vezo | S-African_Bantu | Dusun | -0,01146 | 0,00010 | -120,42300 |
| Malay-SG | Gyan | Murut | -0,00190 | 0,00007 | -26,38870 | Vezo | Vietnamese | S-African_Bantu | -0,01113 | 0,00009 | -120,37100 |
| lMalay-SG | SK-Dayak | Gyan | -0,00144 | 0,00005 | -26,33230 | Vezo | Malay-SG | Yoruba | -0,01070 | 0,00009 | -119,67100 |
| Malay-SG | Gyan | Igorot | -0,00202 | 0,00008 | -24,03160 | Vezo | Banjar | Yoruba | -0,01112 | 0,00009 | -119,60800 |
| Malay-SG | CEU | Maanyan | -0,00168 | 0,00007 | -23,49810 | Vezo | Yoruba | Murut | -0,01153 | 0,00010 | -118,59900 |
| Malay-SG | SK-Dayak | CEU | -0,00151 | 0,00007 | -23,18220 | Vietnamese | Ngaju | CHB | -0,00062 | 0,00004 | -16,10290 |
| Malay-SG | CEU | Murut | -0,00196 | 0,00008 | -23,08730 | Vietnamese | SK-Dayak | CHB | -0,00054 | 0,00004 | -14,28810 |
| Malay-SG | Dusun | Gyan | -0,00159 | 0,00007 | -22,56590 | Vietnamese | Banjar | CHB | -0,00053 | 0,00004 | -12,78690 |
| Malay-SG | Ngaju | Gyan | -0,00121 | 0,00006 | -21,92340 | Vietnamese | Maanyan | CHB | -0,00052 | 0,00004 | -12,52290 |
| Malay-SG | SK-Dayak | Somalian | -0,00130 | 0,00006 | -21,53510 | Vietnamese | Malay-SG | CHB | -0,00043 | 0,00004 | -11,57120 |
| Malay-SG | Eth-Somalian | Murut | -0,00172 | 0,00008 | -21,37210 | Vietnamese | CHB | Murut | -0,00042 | 0,00005 | -9,27405 |
| Malay-SG | SK-Dayak | Eth-Somalian | -0,00129 | 0,00006 | -21,28580 | Vietnamese | CHB | Lebbo | -0,00047 | 0,00005 | -9,09139 |
| Malay-SG | Somalian | Murut | -0,00169 | 0,00008 | -21,09710 | Vietnamese | Dusun | CHB | -0,00033 | 0,00005 | -7,09284 |
| Malay-SG | Eth-Somalian | Maanyan | -0,00138 | 0,00007 | -20,51370 | Vietnamese | Burmese | Igorot | -0,00036 | 0,00005 | -6,54163 |
| Malay-SG | Maanyan | Somalian | -0,00137 | 0,00007 | -20,33400 | Vietnamese | CHB | Bajo | -0,00025 | 0,00005 | -4,59526 |
| Mikea | S-African_Bantu | SK-Dayak | -0,01002 | 0,00010 | -103,48400 | Vietnamese | CHB | Sumbanese | -0,00013 | 0,00004 | -2,85187 |
| Mikea | S-African_Bantu | Ngaju | -0,00991 | 0,00010 | -102,49300 | Vietnamese | Gyan | Igorot | -0,00020 | 0,00009 | -2,18635 |
| Mikea | S-African_Bantu | Maanyan | -0,01012 | 0,00010 | -102,06900 | Vietnamese | CEU | Igorot | -0,00012 | 0,00010 | -1,17054 |
| Mikea | S-African_Bantu | Murut | -0,01023 | 0,00010 | -100,83700 | | | | | | |

**Supplementary Table S3** *-f3*-statistics table obtained for the tree configuration (Malagasy population; Asian population; Yoruba) using the Asian-SNP dataset.

| Source 1 | Source 2 | Target | f3 | SE | Z-score | SNPs |
|---|---|---|---|---|---|---|
| Han | Temoro | Yoruba | 0.114 | 0.005 | 21.092 | 1595 |
| Burmese | Temoro | Yoruba | 0.108 | 0.005 | 21.899 | 1596 |
| Vietnamese | Temoro | Yoruba | 0.119 | 0.006 | 20.305 | 1596 |
| Jehai | Temoro | Yoruba | 0.109 | 0.006 | 18.594 | 1592 |
| Bateq | Temoro | Yoruba | 0.106 | 0.006 | 16.485 | 1590 |
| Mendriq | Temoro | Yoruba | 0.11 | 0.006 | 17.136 | 1593 |
| Kintaq | Temoro | Yoruba | 0.108 | 0.006 | 18.793 | 1594 |
| Mah-Meri | Temoro | Yoruba | 0.111 | 0.006 | 19.042 | 1591 |
| CheWong | Temoro | Yoruba | 0.114 | 0.006 | 17.852 | 1585 |
| Jakun | Temoro | Yoruba | 0.117 | 0.006 | 19.377 | 1590 |
| Temuan | Temoro | Yoruba | 0.115 | 0.005 | 21.534 | 1589 |
| Seletar | Temoro | Yoruba | 0.115 | 0.007 | 17.066 | 1593 |
| Malay | Temoro | Yoruba | 0.115 | 0.005 | 22.628 | 1597 |
| Besemah | Temoro | Yoruba | 0.119 | 0.006 | 21.528 | 1591 |
| Semende | Temoro | Yoruba | 0.123 | 0.006 | 21.331 | 1590 |
| Bidayuh | Temoro | Yoruba | 0.118 | 0.006 | 21.085 | 1589 |
| Banjar | Temoro | Yoruba | 0.121 | 0.006 | 21.923 | 1593 |
| Ngaju | Temoro | Yoruba | 0.12 | 0.005 | 22.258 | 1595 |
| SK-Dayak | Temoro | Yoruba | 0.121 | 0.005 | 23.247 | 1596 |
| Ma'anyan | Temoro | Yoruba | 0.122 | 0.006 | 21.937 | 1597 |
| Lebbo | Temoro | Yoruba | 0.117 | 0.005 | 21.457 | 1593 |
| Murut | Temoro | Yoruba | 0.119 | 0.005 | 22.637 | 1593 |
| Dusun | Temoro | Yoruba | 0.119 | 0.006 | 20.81 | 1594 |
| Bajo | Temoro | Yoruba | 0.116 | 0.005 | 21.905 | 1597 |
| C-Sumbanese | Temoro | Yoruba | 0.117 | 0.005 | 21.529 | 1597 |
| E-Sumbanese | Temoro | Yoruba | 0.113 | 0.005 | 20.879 | 1597 |
| W-Sumbanese | Temoro | Yoruba | 0.114 | 0.005 | 21.931 | 1597 |
| Nusa_Tenga | Temoro | Yoruba | 0.107 | 0.006 | 18.523 | 1590 |
| Moluccas | Temoro | Yoruba | 0.111 | 0.006 | 19.981 | 1587 |
| PNG highlander | Temoro | Yoruba | 0.086 | 0.008 | 11.34 | 1578 |
| Philippine Negrito | Temoro | Yoruba | 0.109 | 0.006 | 19.665 | 1587 |
| Filipino | Temoro | Yoruba | 0.12 | 0.006 | 20.97 | 1596 |
| Igorot | Temoro | Yoruba | 0.127 | 0.006 | 22.68 | 1593 |
| Taiwan Aboriginese | Temoro | Yoruba | 0.119 | 0.006 | 20.335 | 1594 |
| Fiji | Temoro | Yoruba | 0.103 | 0.005 | 19.196 | 1594 |
| Polynesia | Temoro | Yoruba | 0.116 | 0.006 | 20.188 | 1591 |
| Han | Vezo | Yoruba | 0.12 | 0.005 | 25.6 | 1620 |
| Burmese | Vezo | Yoruba | 0.112 | 0.005 | 24.617 | 1620 |
| Vietnamese | Vezo | Yoruba | 0.123 | 0.005 | 24.012 | 1621 |
| Jehai | Vezo | Yoruba | 0.111 | 0.005 | 20.357 | 1616 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Bateq | Vezo | Yoruba | 0.108 | 0.006 | 18.361 | 1614 |
| Mendriq | Vezo | Yoruba | 0.113 | 0.006 | 20.44 | 1619 |
| Kintaq | Vezo | Yoruba | 0.11 | 0.006 | 18.975 | 1615 |
| Mah-Meri | Vezo | Yoruba | 0.118 | 0.005 | 21.917 | 1616 |
| CheWong | Vezo | Yoruba | 0.121 | 0.006 | 19.101 | 1605 |
| Jakun | Vezo | Yoruba | 0.119 | 0.006 | 21.084 | 1612 |
| Temuan | Vezo | Yoruba | 0.119 | 0.005 | 23.297 | 1612 |
| Seletar | Vezo | Yoruba | 0.114 | 0.006 | 18.853 | 1615 |
| Malay | Vezo | Yoruba | 0.12 | 0.005 | 24.098 | 1622 |
| Besemah | Vezo | Yoruba | 0.124 | 0.005 | 24.584 | 1614 |
| Semende | Vezo | Yoruba | 0.126 | 0.005 | 24.929 | 1614 |
| Bidayuh | Vezo | Yoruba | 0.124 | 0.006 | 21.471 | 1613 |
| Banjar | Vezo | Yoruba | 0.127 | 0.005 | 25.264 | 1617 |
| Ngaju | Vezo | Yoruba | 0.125 | 0.005 | 24.353 | 1620 |
| SK-Dayak | Vezo | Yoruba | 0.126 | 0.005 | 26.05 | 1622 |
| Ma'anyan | Vezo | Yoruba | 0.128 | 0.005 | 25.407 | 1622 |
| Lebbo | Vezo | Yoruba | 0.122 | 0.006 | 21.471 | 1616 |
| Murut | Vezo | Yoruba | 0.124 | 0.005 | 22.842 | 1619 |
| Dusun | Vezo | Yoruba | 0.127 | 0.005 | 23.517 | 1619 |
| Bajo | Vezo | Yoruba | 0.117 | 0.005 | 24.57 | 1620 |
| C-Sumbanese | Vezo | Yoruba | 0.119 | 0.005 | 23.606 | 1621 |
| E-Sumbanese | Vezo | Yoruba | 0.116 | 0.005 | 23.617 | 1622 |
| W-Sumbanese | Vezo | Yoruba | 0.115 | 0.005 | 24.193 | 1622 |
| Nusa_Tenga | Vezo | Yoruba | 0.112 | 0.006 | 20.193 | 1615 |
| Moluccas | Vezo | Yoruba | 0.115 | 0.005 | 22 | 1613 |
| PNG highlander | Vezo | Yoruba | 0.082 | 0.008 | 10.777 | 1603 |
| Philippine Negrito | Vezo | Yoruba | 0.113 | 0.005 | 21.12 | 1609 |
| Filipino | Vezo | Yoruba | 0.123 | 0.005 | 24.999 | 1619 |
| Igorot | Vezo | Yoruba | 0.13 | 0.005 | 23.754 | 1618 |
| Taiwan Aboriginese | Vezo | Yoruba | 0.125 | 0.006 | 21.543 | 1618 |
| Fiji | Vezo | Yoruba | 0.106 | 0.005 | 19.416 | 1616 |
| Polynesia | Vezo | Yoruba | 0.121 | 0.006 | 21.102 | 1614 |
| Han | Mikea | Yoruba | 0.116 | 0.005 | 22.42 | 1621 |
| Burmese | Mikea | Yoruba | 0.11 | 0.005 | 22.506 | 1622 |
| Vietnamese | Mikea | Yoruba | 0.118 | 0.006 | 21.511 | 1621 |
| Jehai | Mikea | Yoruba | 0.106 | 0.005 | 20.259 | 1619 |
| Bateq | Mikea | Yoruba | 0.103 | 0.006 | 16.336 | 1616 |
| Mendriq | Mikea | Yoruba | 0.107 | 0.006 | 18.533 | 1619 |
| Kintaq | Mikea | Yoruba | 0.108 | 0.006 | 18.301 | 1616 |
| Mah-Meri | Mikea | Yoruba | 0.113 | 0.005 | 20.671 | 1615 |
| CheWong | Mikea | Yoruba | 0.115 | 0.006 | 19.484 | 1612 |
| Jakun | Mikea | Yoruba | 0.118 | 0.006 | 20.018 | 1616 |
| Temuan | Mikea | Yoruba | 0.115 | 0.005 | 21.211 | 1615 |
| Seletar | Mikea | Yoruba | 0.109 | 0.006 | 19.324 | 1617 |
| Malay | Mikea | Yoruba | 0.115 | 0.005 | 22.626 | 1622 |
| Besemah | Mikea | Yoruba | 0.122 | 0.006 | 21.96 | 1615 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Semende | Mikea | Yoruba | 0.12 | 0.006 | 20.337 | 1616 |
| Bidayuh | Mikea | Yoruba | 0.121 | 0.006 | 21.17 | 1616 |
| Banjar | Mikea | Yoruba | 0.122 | 0.005 | 22.223 | 1619 |
| Ngaju | Mikea | Yoruba | 0.119 | 0.006 | 21.055 | 1622 |
| SK-Dayak | Mikea | Yoruba | 0.123 | 0.005 | 23.21 | 1622 |
| Ma'anyan | Mikea | Yoruba | 0.123 | 0.006 | 21.426 | 1622 |
| Lebbo | Mikea | Yoruba | 0.121 | 0.006 | 21.713 | 1618 |
| Murut | Mikea | Yoruba | 0.118 | 0.005 | 21.946 | 1621 |
| Dusun | Mikea | Yoruba | 0.124 | 0.006 | 20.771 | 1621 |
| Bajo | Mikea | Yoruba | 0.116 | 0.005 | 21.556 | 1622 |
| C-Sumbanese | Mikea | Yoruba | 0.116 | 0.006 | 19.959 | 1622 |
| E-Sumbanese | Mikea | Yoruba | 0.113 | 0.006 | 19.793 | 1621 |
| W-Sumbanese | Mikea | Yoruba | 0.115 | 0.005 | 21.267 | 1621 |
| Nusa_Tenga | Mikea | Yoruba | 0.114 | 0.006 | 18.352 | 1618 |
| Moluccas | Mikea | Yoruba | 0.116 | 0.006 | 17.925 | 1614 |
| PNG highlander | Mikea | Yoruba | 0.089 | 0.008 | 10.992 | 1603 |
| Philippine Negrito | Mikea | Yoruba | 0.109 | 0.006 | 17.881 | 1613 |
| Filipino | Mikea | Yoruba | 0.121 | 0.006 | 21.286 | 1621 |
| Igorot | Mikea | Yoruba | 0.126 | 0.006 | 20.052 | 1618 |
| Taiwan Aboriginese | Mikea | Yoruba | 0.12 | 0.006 | 20.126 | 1620 |
| Fiji | Mikea | Yoruba | 0.107 | 0.006 | 18.074 | 1619 |
| Polynesia | Mikea | Yoruba | 0.12 | 0.006 | 18.604 | 1618 |

**Supplementary Table S4** - $F_{ST}$ genetic distances (lower diagonal) between 'Asian-SNP' Malagasy and data from the Asian populations of the high density dataset, and the corresponding standard deviation (upper diagonal).

| FST \ S.D. | Vezo | Mikea | Temoro | CHB | Burmese | Vietnamese | Malay-SG | Banjar | Ma'anyan | Ngaju | SK-Dayak | Lebbo | Murut | Dusun | Bajo | Sumbanese | Filipino | Igorot |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Vezo** | 0 | 0.005 | 0.006 | 0.004 | 0.004 | 0.003 | 0.004 | 0.003 | 0.004 | 0.003 | 0.003 | 0.004 | 0.004 | 0.004 | 0.003 | 0.003 | 0.003 | 0.003 |
| **Mikea** | 0.01 | 0 | 0.006 | 0.005 | 0.004 | 0.005 | 0.004 | 0.005 | 0.004 | 0.004 | 0.004 | 0.005 | 0.005 | 0.005 | 0.004 | 0.004 | 0.005 | 0.005 |
| **Temoro** | 0.007 | 0.016 | 0 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.005 | 0.004 | 0.005 | 0.004 | 0.004 | 0.004 | 0.005 |
| **CHB** | 0.035 | 0.041 | 0.031 | 0 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 |
| **Burmese** | 0.032 | 0.034 | 0.024 | 0.009 | 0 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 |
| **Vietnamese** | 0.026 | 0.032 | 0.021 | 0.007 | 0.009 | 0 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 |
| **Malay-SG** | 0.019 | 0.028 | 0.015 | 0.015 | 0.009 | 0.007 | 0 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 |
| **Banjar** | 0.021 | 0.029 | 0.016 | 0.02 | 0.015 | 0.009 | 0.004 | 0 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 |
| **Ma'anyan** | 0.026 | 0.036 | 0.021 | 0.028 | 0.024 | 0.017 | 0.009 | 0.008 | 0 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 |
| **Ngaju** | 0.021 | 0.028 | 0.016 | 0.021 | 0.015 | 0.011 | 0.003 | 0.004 | 0.008 | 0 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 |
| **SK-Dayak** | 0.018 | 0.025 | 0.015 | 0.019 | 0.014 | 0.01 | 0.002 | 0.002 | 0.005 | 0.001 | 0 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 |
| **Lebbo** | 0.043 | 0.046 | 0.036 | 0.042 | 0.036 | 0.034 | 0.024 | 0.027 | 0.034 | 0.027 | 0.025 | 0 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.003 |
| **Murut** | 0.033 | 0.041 | 0.027 | 0.028 | 0.028 | 0.021 | 0.016 | 0.018 | 0.024 | 0.016 | 0.015 | 0.035 | 0 | 0.002 | 0.001 | 0.002 | 0.001 | 0.002 |
| **Dusun** | 0.033 | 0.041 | 0.026 | 0.028 | 0.026 | 0.02 | 0.017 | 0.016 | 0.022 | 0.019 | 0.015 | 0.036 | 0.023 | 0 | 0.001 | 0.002 | 0.001 | 0.002 |
| **Bajo** | 0.028 | 0.032 | 0.018 | 0.02 | 0.016 | 0.014 | 0.008 | 0.011 | 0.019 | 0.011 | 0.009 | 0.028 | 0.02 | 0.019 | 0 | 0.001 | 0.001 | 0.002 |
| **Sumbanese** | 0.031 | 0.037 | 0.019 | 0.025 | 0.02 | 0.019 | 0.011 | 0.015 | 0.021 | 0.014 | 0.012 | 0.032 | 0.025 | 0.024 | 0.008 | 0 | 0.001 | 0.002 |
| **Filipino** | 0.023 | 0.028 | 0.017 | 0.013 | 0.015 | 0.008 | 0.006 | 0.008 | 0.013 | 0.009 | 0.006 | 0.028 | 0.013 | 0.012 | 0.007 | 0.012 | 0 | 0.002 |
| **Igorot** | 0.051 | 0.056 | 0.043 | 0.047 | 0.05 | 0.039 | 0.037 | 0.038 | 0.042 | 0.038 | 0.037 | 0.057 | 0.04 | 0.041 | 0.038 | 0.044 | 0.026 | 0 |

**Supplementary Table S5** – GLOBETROTTER inferred dates (in both generations from present and years, bootstrap 95% CIs given in parenthesis; YBP = years before present), admixing sources, and percentage (%) of admixture for each population of interest (Hellenthal, et al. 2014). We assume a generation time of 25 years. $R^2$ corresponds to the goodness-of-fit of the tested model. $FQ_1$ and $FQ_2$ correspond respectively to the fit of a single admixture event and the fit of the first two principal components capturing the admixture events. M corresponds to the additional $R^2$ explained by adding a second date versus assuming only a single date of admixture (M>0.35 to infer multiple dates event). The admixture models presented in the table correspond to the best fit models considering the 'best guess'. P: p-value of evidence of any detectable admixture event obtained after 100 bootstrap resamplings of the NULL procedure. Date confidence intervals are based on 100 bootstrap replicates of the date inference.

| Target Group | Analysis | P | $R^2$ | $FQ_1$ | $FQ_2$ | M | Best-guess | 1-date (gen.) | 1-date (YBP) | 1-date Best Surrogate 1 | 1-date Best Surrogate 2 | 2-date (gen.) | 2-date (YBP) | 2-date Best Surrogate 1 | 2-date Best Surrogate 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Vezo | main | < 0.01 | 0.996 | 1 | 1 | 0.095 | one-date | 25 (24-27) | 625 (600-675) | Banjar (39%) | S.A.Bantu (61%) | - | - | - | - |
| | main.null | < 0.01 | 0.996 | 1 | 1 | 0.012 | one-date | 25 (24-27) | 625 (600-675) | Banjar (38%) | S.A.Bantu (62%) | - | - | - | - |
| Mikea | main | < 0.01 | 0.993 | 1 | 1 | 0.142 | one-date | 27 (25-29) | 675 (625-725) | Banjar (36%) | S.A.Bantu (64%) | - | - | - | - |
| | main.null | < 0.01 | 0.991 | 1 | 1 | 0.255 | one-date | 27 (25-29) | 675 (625-725) | Banjar (36%) | S.A.Bantu (64%) | - | - | - | - |
| Temoro | main | < 0.01 | 0.996 | 1 | 1 | 0.076 | one-date | 30 (29-31) | 750 (725-775) | Banjar (37%) | S.A.Bantu (63%) | - | - | - | - |
| | main.null | < 0.01 | 0.995 | 1 | 1 | 0.102 | one-date | 30 (28-31) | 750 (725-775) | Banjar (37%) | S.A.Bantu (63%) | - | - | - | - |
| Banjar | main | < 0.01 | 0.621 | 1 | 1 | 0.078 | one-date | 17 (11-20) | 425 (275-500) | Ma'anyan (23%) | Malay (77%) | - | - | - | - |
| | main.null | < 0.01 | 0.615 | 1 | 1 | 0.042 | one-date | 23 (14-32) | 575 (350-800) | Ma'anyan (24%) | Malay (76%) | - | - | - | - |
| Ngaju | main | < 0.01 | 0.521 | 1 | 1 | 0.082 | one-date | 16 (9-19) | 400 (225-475) | Ma'anyan (19%) | Malay (81%) | - | - | - | - |
| | main.null | < 0.01 | 0.554 | 1 | 1 | 0.045 | one-date | 20 (17-23) | 500 (425-575) | Ma'anyan (43%) | Malay (57%) | - | - | - | - |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SK Dayak | main | < 0.01 | 0.463 | 1 | 1 | 0.127 | one-date | 12 (7-16) | 300 (175-400) | Malay (21%) | Ma'anyan (79%) | - | - | - | - |
| | main.null | < 0.01 | 0.615 | 1 | 1 | 0.084 | one-date | 19 (11-35) | 475 (275-875) | Malay (27%) | Ma'anyan (73%) | - | - | - | - |
| Malay | main | < 0.01 | 0.959 | 0.972 | 0.989 | 0.481 | multiple-dates | 7 (2-10) | 175 (50-250) | Burmese (24%) | Filipino (76%) | 42 (29-58) | 1050 (725-1450) | Brahmin (14%) | Filipino (86%) |
| | main.null | < 0.01 | 0.94 | 0.979 | 0.992 | 0.514 | multiple-dates | 9 (4-15) | 225 (100-375) | Burmese (16%) | Filipino (84%) | 66 (43-111) | 1650 (1075-2775) | Brahmin (19%) | Filipino (81%) |
| Bajo | main | < 0.01 | 0.946 | 0.882 | 0.998 | 0.22 | one-date-multiway | 35 (32-40) | 875 (800-1000) | Sumbanese (23%), Brahmin (12%) | Banjar (25%), Filipino (25%), Malay (11%), Ngaju (6%) | - | - | - | - |
| | main.null | < 0.01 | 0.953 | 0.96 | 0.999 | 0.22 | one-date-multiway | 47 (43-54) | 1175 (1075-1350) | Sumbanese (27%), Brahmin (15%) | Banjar (23%), Filipino (27%), Ngaju (8%) | - | - | - | - |

**Supplementary Table S6** – Estimated dates of admixture in Malagasy ethnic groups using Banjar and South African Bantu data as parental populations with ALDER (Loh, et al. 2013).

| Test Pop | Ref A | Ref B | p-value | 2-ref z-score | 1-ref z-score A | 1-ref z-score B | max decay diff (%) | 2-ref decay (gen.) | 2-ref amp_exp |
|---|---|---|---|---|---|---|---|---|---|
| Temoro | Banjar | South Africa Bantu | 5.1e-81 | 19.34 | 17.87 | 22.90 | 0.03 | 29.16 +/- 1.08 | 0.0018 +/- 0.0001 |
| Mikea | Banjar | South Africa Bantu | 6.9e-85 | 19.79 | 18.70 | 11.12 | 0.09 | 25.21 +/- 1.27 | 0.0017 +/- 0.0001 |
| Vezo | Banjar | South Africa Bantu | 2.6e-69 | 17.89 | 18.88 | 15.81 | 0.01 | 22.29 +/- 1.25 | 0.0017 +/- 0.0001 |

## Supplementary Methods

*Sample collection and ethics*

A total of 41 DNA samples were analyzed from two groups in Southeast Borneo: the Banjar (n=16) and the Ngaju (n=25) that were collected in Banjarmasin city (Supplementary Figure S1). Blood samples were collected from healthy adult donors, all of whom provided written informed consent. DNA was extracted using a standard salting-out procedure. All participants were surveyed for language affiliation, current residence, familial birthplaces, and a genealogy of four generations to establish ancestry. This study was approved by the Research Ethics Commission of the Eijkman Institute for Molecular Biology (Jakarta, Indonesia). Genome-wide SNP genotyping for individuals from the two groups were performed using the Illumina Human Omni Express Bead Chip-24 v1.0 (Illumina Inc., San Diego, CA), which characterizes 730,525 single nucleotide markers regularly spaced across the genome. The accession number for the Southeast Borneo samples genotyped for this study is EGA: EGAS00001001841.

*Dataset*

We gathered data from previously published studies on populations from Madagascar, Southeast Asia, South Asia, East Asia, East Africa, South Africa, and Europe (Supplementary Table S1). Two datasets were compiled respective to their analytical use: a low SNP density dataset of populations covering a large geographical area; and a high SNP density dataset of populations composed by a subset of populations of the latter dataset. To avoid any statistical bias that could be induced by a size effect due to populations over-represented in the high SNP density

dataset, we randomly selected a maximum of 31 individuals in each group, such as each population has a number of individuals between 15 and 31. Quality controls were applied using Plink v1.9 (Chang, et al. 2015) to filter for i) close relatives, using an Identity-by-Descent (IBD) estimation with upper threshold of 0.25 (second degree relatives); ii) SNPs that failed the Hardy-Weinberg exact (HWE) test (P < 10-6) were excluded; iii) samples with a call rate <0.99 and displayed missing rates >0.05 across all samples in each population were excluded; and iv) variants in high linkage disequilibrium ($r^2 > 0.5$) were also removed for the low density dataset. After filtering, the low SNP density dataset included 2183 individuals from 61 populations genotyped for 40,272; and the SNPs high SNP density dataset was composed of 551 individuals from 24 populations genotyped for 374,189 SNPs. All genotypes of the high SNP density dataset were then phased together with SHAPEIT v2.r790 (Delaneau, et al. 2012) using the 1000Genomes phased data (Delaneau, et al. 2014) as reference panel and the HapMap phase 2 genetic map (International HapMap 2005).

*Statistical Analyses*

The low density dataset was described by the following analyses. Principal Components Analysis was computed with EIGENSOFT v6.0.1 (Patterson et al. 2006). ADMIXTURE v1.23 (Alexander et al. 2009) was used to estimate the profile of individual genomic ancestries using maximum likelihood for components K = 2 to K = 18. Ten replicates were run at each value of K with different random seeds, then merged and assessed for clustering quality using CLUMPP (Jakobsson and Rosenberg 2007), and the cross-validation value was calculated to determine the optimal number of genomic components (here, K = 16). ADMIXTURE and PCA plots

were generated with Genesis (Buchmann and Hazelhurst 2014). Three-population (*f*3) statistics (Patterson, et al. 2012) were computed for each trios of populations of the low SNP density dataset to identify groups showing potential recent admixture event. TreeMix v1.12 analysis (Pickrell and Pritchard 2012) was performed with all Asian populations and three representative of African groups (Yoruba, Luhya, South African Bantu), to estimate gene flows from Asian groups and their relative intensity, with blocks of 200 SNPs to account for linkage disequilibrium and migration edges added sequentially until the model explained 99% of the variance (the TreeMix outputs in Newick format were visualized with MEGA v7.0.14 (Tamura, et al. 2013)).

Population structure of the phased high density dataset was evaluated using the fineSTRUCTURE v2.07 package (Lawson et al. 2012). It uses the detection of shared IBD fragments between each pair of individuals, without self-copying, calculated with Chromopainter v2.0 (Lawson et al. 2012) to perform a model-based Bayesian clustering of genotypes. From the results, a coancestry heat map and a dendrogram were inferred to visualize the number of clusters statistically defined that would describe relevantly the data. FineSTRUCTURE v2 identified 37 groups of individuals that can be statistically defined as natural populations (Lawson et al. 2012), according to their shared Identity-By-Descent (IBD) (bootstrapped nodes>60%; Supplementary figure S7). Most of these correspond to anthropologically defined populations, such as the Banjar or the Ngaju, with few exceptions (e.g., the sub-structuration of the Kenyan Luhya group). Although each Malagasy individual has its closest connections with other Malagasy, regardless of their ethnicity, these ethnic groups were treated separately for anthropological interest. Haplotype sharing between pairs of individuals was estimated from the phased high SNP density dataset by the Refined IBD algorithm of Beagle v4.0

(Browning and Browning 2007; Browning and Browning 2013), filtering for detected fragments with a logarithm of odds ratio (LOD)>3. 10 iterations were realized, randomizing the seed number for each run, and the overlapping shared fragments were merged to favour the detection of long shared IBD. Detected fragments between the same pairs of individuals were summed up and visualized with Cytoscape v3.2.1 (Shannon et al. 2003). IBD sharing data with Malagasy individuals was also averaged per Asian population to geographically plot a gradient map with Surfer v12.0 using the Kriging method (GoldenSoftware 2014). All maps used in the present study were generated using Global Mapper v.15 software. (http://www.bluemarblegeo.com/products/global-mapper.php). Local ancestry analysis in Malagasy individuals was performed with PCAdmix v1.0 (Brisbin et al. 2012) using two parental metapopulations of 100 individuals of African ancestry (randomly selected in Yoruba, South African Bantu, Kenyan Luhya and Somali groups), and of Asian ancestry (randomly selected in Han, Igorot, Ma'anyan and Malay groups). The phased Malagasy data were screened using the linkage disequilibrium information to define the probability of common ancestry of each Malagasy haplotype with each 'parental' metapopulation. The Viterbi algorithm was then used to mask all haplotypes according to one or the other ancestry in the Malagasy individuals. Ancestry-specific PCAs and $F_{ST}$ calculations were realized with EIGENSOFT v6.0.1 (Patterson et al. 2006). Three-population (*f*3) statistics (Patterson et al. 2012) and TreeMix v1.12 analysis (Pickrell and Pritchard 2012) were also performed on this masked dataset (1,664 SNPs for 18 populations). Haplotype 'painting' with Chromopainter v2 (Lawson et al. 2012) was realized on the high density of SNP dataset, defining each cluster of populations as target or donor/surrogate according to the anthropological question addressed. Mutational

rates and Ne parameters were first estimated with an Estimation-Maximization (EM) algorithm running Chromopainter v2 on all 22 autosomes for the entire dataset with 10 iterations (Lawson et al. 2012). The weighted average of these parameters, according to the SNP coverage of each used chromosomes and the number of individuals, were then used to compute the chromosomal painting. Each cluster of populations has been successively identified as target and the others as surrogates (at the exclusion of the Malagasy cluster which has not been used as surrogate). The obtained painted chromosomes for each cluster were used in GLOBETROTTER v1.0 (Hellenthal et al. 2014) to estimate the ratios and the dates of the potential admixture events characterizing them. Coancestry curves were estimated with and without standardization with a 'NULL' individual, and consistency between each estimated parameters was checked. 100 bootstrap resamplings were realized to estimate the p-value of the admixture events (considering the 'NULL' individual) and the 95% confidence interval for the obtained dates (without the 'NULL' individual). The 'best-guess' scenario given by GLOBETROTTER v1.0 (Hellenthal et al. 2014) was considered for each target population. Using the parental populations given by GLOBETROTTER v1.0 (Hellenthal et al. 2014), dates of admixture were also estimated by ALDER v1 (Loh et al. 2013).

## REFERENCES


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

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

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

Browning BL, Browning SR 2013. Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics 194: 459-471. doi: 10.1534/genetics.113.150029

Browning SR, Browning BL 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. The American Journal of Human Genetics 81: 1084-1097. doi: 10.1086/521987

Buchmann R, Hazelhurst S. 2014. Genesis Manual. University of the Witwatersrand, Johannesburg.

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

Cox MP, Hudjashov G, Sim A, Savina O, Karafet TM, Sudoyo H, Lansing JS forthcoming. Small traditional human communities sustain genomic diversity over microgeographic sscales despite linguistic isolation. Molecular Biology and Evolution.

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

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

GoldenSoftware. 2014. Surfer v.12.01.

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

International HapMap C 2005. A haplotype map of the human genome. Nature 437: 1299-1320. doi: 10.1038/nature04226

International HapMap C, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Gibbs RA, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarrol SA, Nemesh J, Dermitzakis E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghori MJ, McGinnis R, McLaren W, Pollack S, Price AL, Schaffner SF,

Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE 2010. Integrating common and rare genetic variation in diverse human populations. Nature 467: 52-58. doi: 10.1038/nature09298

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

Kusuma P, Brucato N, Cox MP, Pierron D, Razafindrazaka H, Adelaar A, Sudoyo H, Letellier T, Ricaut F-X 2016. Contrasting Linguistic and Genetic Influences during the Austronesian Settlement of Madagascar. Scientific Reports 6:26066. doi: doi: 10.1038/srep26066

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

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

May A, Hazelhurst S, Li Y, Norris SA, Govind N, Tikly M, Hon C, Johnson KJ, Hartmann N, Staedtler F, Ramsay M 2013. Genetic diversity in black South Africans from Soweto. BMC Genomics 14: 644. doi: 10.1186/1471-2164-14-644

Mörseburg A, Pagani L, Ricaut F-X, Yngvadottir B, Harney E, Castillo C, Antao T, Hoogervorst T, Kusuma P, Razafindrazaka H, Cardona A, Pierron D, Letellier T, Wee J, Abdullah M, Metspalu M, Kivisild T forthcoming. Multi-layered population

structure and local adaptations in Island Southeast Asians. European Journal of Human Genetics.

Pagani L, Kivisild T, Tarekegn A, Ekong R, Plaster C, Gallego Romero I, Ayub Q, Mehdi SQ, Thomas MG, Luiselli D, Bekele E, Bradman N, Balding DJ, Tyler-Smith C 2012. Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am J Hum Genet 91: 83-96. doi: 10.1016/j.ajhg.2012.05.015

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

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

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

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

Reich D, Patterson N, Kircher M, Delfin F, Nandineni MR, Pugach I, Ko AM, Ko YC, Jinam TA, Phipps ME, Saitou N, Wollstein A, Kayser M, Paabo S, Stoneking M 2011. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. Am J Hum Genet 89: 516-528. doi: 10.1016/j.ajhg.2011.09.005

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

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30: 2725-2729. doi: 10.1093/molbev/mst197

Teo YY, Sim X, Ong RT, Tan AK, Chen J, Tantoso E, Small KS, Ku CS, Lee EJ, Seielstad M, Chia KS 2009. Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations. Genome Res 19: 2154-2162. doi: 10.1101/gr.095000.109