HAL open science

Evidence of Austronesian Genetic Lineages in East Africa and South Arabia: Complex Dispersal from Madagascar and Southeast Asia

Nicolas Brucato, Verónica Fernandes, Pradiptajati Kusuma, Viktor Černý, Connie Mulligan, Pedro Soares, Teresa Rito, Céline Besse, Anne Boland, Jean-François Deleuze, et al.

To cite this version:

Nicolas Brucato, Verónica Fernandes, Pradiptajati Kusuma, Viktor Černý, Connie Mulligan, et al.. Evidence of Austronesian Genetic Lineages in East Africa and South Arabia: Complex Dispersal from Madagascar and Southeast Asia. Genome Biology and Evolution, 2019, 11 (3), pp.748-758. 10.1093/gbe/evz028 . hal-02112685

HAL Id: hal-02112685

https://hal.science/hal-02112685v1

Submitted on 27 Apr 2019

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

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

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Evidence of Austronesian Genetic Lineages in East Africa and South Arabia: Complex Dispersal from Madagascar and Southeast Asia

Nicolas Brucato[1], Veronica Fernandes[2,3], Pradiptajati Kusuma[4], Viktor Černý[5], Connie J. Mulligan[6], Pedro Soares[3,7], Teresa Rito[3,8,9], Céline Besse[10], Anne Boland[10], Jean-Francois Deleuze[10], Murray P. Cox[11], Herawati Sudoyo[4,12], Mark Stoneking[13], Luisa Pereira[2,3], and François-Xavier Ricaut[1,*]

[1]Laboratoire Évolution & Diversité Biologique (EDB UMR 5174), Université de Toulouse Midi-Pyrénées, CNRS, IRD, UPS, Toulouse, France

[2]Instituto de Investigação e Inovação em Saúde, Universidade do Porto (i3S), Porto, Portugal

[3]Instituto de Patologia e Imunologia Molecular da Universidade do Porto (Ipatimup), Porto, Portugal

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

[5]Department of Anthropology, Faculty of Natural Sciences, Comenius University, Bratislava, Slovakia

[6]Department of Anthropology, University of Florida

[7]Centro de Biologia Molecular e Ambiental (CBMA), Departamento de Biologia, Universidade do Minho, Braga, Portugal

[8]Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal

[9]Life and Health Sciences Research Institute (ICVS), School of Medicine & ICVS/3B, PT Government Associate Laboratory, University of Minho, Braga, Portugal

[10]Centre National de Recherche en Génomique Humaine (CNRGH), Institut de Biologie François Jacob, CEA, Université Paris-Saclay, Evry, France

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

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

[13]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

*Corresponding author: E-mail: francois-xavier.ricaut@univ-tlse3.fr.

Accepted: January 30, 2019

Data deposition: This project has been deposited at the European Genome-phenome Archive (https://www.ebi.ac.uk/ega/home) under the accession EGAS00001003425 for the genotyping array data, and on GenBank (https://www.ncbi.nlm.nih.gov/genbank/) under the accession number MK128798–MK128899 for the 102 new complete mtDNA sequences.

Abstract

The Austronesian dispersal across the Indonesian Ocean to Madagascar and the Comoros has been well documented, but in an unexplained anomaly, few to no traces have been found of the Austronesian expansion in East Africa or the Arabian Peninsula. To revisit this peculiarity, we surveyed the Western Indian Ocean rim populations to identify potential Austronesian genetic ancestry. We generated full mitochondrial DNA genomes and genome-wide genotyping data for these individuals and compared them with the Banjar, the Indonesian source population of the westward Austronesian dispersal. We find strong support for Asian genetic contributions to maternal lineages and autosomal variation in modern day Somalia and Yemen. Surprisingly, this input reveals two apparently different geographic origins and timings of admixture for the Austronesian contact; one at a very early phase (likely associated with the early Austronesian dispersals), and a later movement dating to the end of nineteenth century. These Austronesian gene flows come, respectively, from Madagascar and directly from an unidentified location in Island Southeast Asia. This result reveals a far more complex dynamic of Austronesian dispersals through the Western Indian Ocean than has previously been understood and suggests that Austronesian movements within the Indian Ocean may have been part of a lengthy process, probably continuing well into the modern era.

Key words: Madagascar, Austronesian, Polynesian motif, genome-wide data, mitochondrial DNA.

© The Author(s) 2019. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/gbe/article-abstract/11/3/748/5306180 by scd - universit Toulouse III user on 20 March 2019

Introduction

The Austronesian dispersal eastward across the Pacific Ocean and westward across the Indian Ocean is increasingly well documented from historical, archaeological, and genetic perspectives (Beaujard 2012a, 2012b; Duggan et al. 2014; Brucato et al. 2016, 2018; Crowther et al. 2016; Skoglund et al. 2016; Bellwood 2017; Pierron et al. 2017). On the western edge of the Austronesian expansion, the Indian Ocean trading network during the last two millennia led to an exchange of ideas, goods, and also people among Africa, the Middle East, and Asia (Beaujard 2012a, 2012b). Recent studies have reconciled historical, linguistic, and genetic data to reconstruct this interaction network, identified the Asian population that is the source of the Austronesian dispersal (the Banjar from Southeast Borneo in Indonesia), and proposed a robust hypothesis for the timing (between the eighth and thirteenth centuries) and the nature of admixture processes in the East African offshore Islands of Madagascar and the Comoros (Pierron et al. 2014; Brucato et al. 2016, 2017, 2018).

Recent results based on genome-wide data from populations around the Indian Ocean rim support the scenario of a "direct route" from Southeast Borneo to the Comoros and Madagascar, as no significant Austronesian gene flow to other Western Indian Ocean populations was detected (Brucato et al. 2018). However, the global scale of these data might have prevented the tracking of minor genetic contributions from Austronesian migrants, potentially related to geographically and chronologically different events than the main early Austronesian dispersal. In this perspective, uniparental markers such as mitochondrial DNA (mtDNA) are an especially informative tool to identify traces of genetic inheritance, as minor genetic contributions with clear geographic provenance can become insignificant in the autosomal genome after a few generations due to recombination, but can still be maintained or even spread (by drift or sex-biased admixture) for uniparental markers that do not undergo recombination. This question is particularly interesting for the key genetic marker of the westward Austronesian expansion—the mtDNA haplogroup B4a1a1b (Malagasy motif; Razafindrazaka et al. 2010), the Indian Ocean variant of the Polynesian motif (B4a1a1 haplogroup; Soodyall et al. 1995), which has been previously identified in Madagascar and is likely present in the Comoros (Msaidie et al. 2011; Mazières et al. 2018). This lineage is the only Austronesian-specific marker that is both the main Asian maternal lineage found in Madagascar and the main lineage associated with the Austronesian expansion throughout the Indian Ocean and Oceania (Razafindrazaka et al. 2010; Cox et al. 2012; Tumonggor et al. 2013; Duggan et al. 2014; Kusuma et al. 2015).

The Malagasy motif (characterized by polymorphisms C1473T and T3423A) has only been found in Madagascar (ranging from 10% to 50% in frequency) (Tofanelli et al. 2009; Razafindrazaka et al. 2010; Cox et al. 2012; Pierron et al. 2014, 2017) and to date has not been found anywhere else, including within Indonesia and in other regions influenced by the Austronesian dispersal. Its direct precursor, the Polynesian motif (characterized by the polymorphisms A14022G, T16217C, A16247G, and C16261T) is largely restricted to the east of the Wallace's line, from eastern Indonesia to the Pacific islands, with sporadic occurrences further west (e.g., Bali and Borneo; Cox et al. 2012; Tumonggor et al. 2013; Kusuma et al. 2015).

Here, we report the first mitogenomes and associated autosomal genome-wide data of individuals from East Africa and the Arabian Peninsula who have been identified to carry the Malagasy motif. To investigate broader questions about Austronesian gene flow within the Indian Ocean, we also present an analysis of the mitogenomes of the Banjar population of Southeast Borneo, who are considered to be descendants of the ancestral population that migrated to Madagascar and Comoros, and explore how this population connects with the newly discovered Malagasy motif lineages in Somalia and Yemen.

Materials and Methods

Ethics

This study was approved by the Research Ethics Commission of the Eijkman Institute for Molecular Biology (Jakarta, Indonesia) under Research Ethic clearance number 90 for the study of Indonesia Human Genome Diversity and Diseases, and by the French Ethics Committees (Committees of Protection of Persons). Biological sampling was conducted by the Eijkman Institute for Molecular Biology, with the assistance of local Public Health clinic staff, following protocols for the protection of human subjects established by the Eijkman Institute. For Yemeni and Somali samples, the study was approved by the Yemen Center for Studies and Research, (Sana'a, Yemen). All samples were collected with informed consent from unrelated individuals.

Sampling and mtDNA Analysis

The samples analyzed in this study are from populations from Oceania, Island South East Asia (ISEA), and the western part of the Indian Ocean (WIO). Our data set of 1,086 sequences (supplementary table S1, Supplementary Material online) includes only mtDNA sequences affiliated with mtDNA haplogroup B4a1a1 and its subclades and was compiled from 1) newly collected samples for which the whole mtDNA was sequenced (Banjar population from Borneo, Indonesia), 2) newly generated whole mtDNA sequences from selected samples with published HVS-I sequences affiliated with the Polynesian motif from the Western Indian Ocean (1 Somali,

Downloaded from https://academic.oup.com/gbe/article-abstract/11/3/748/5306180 by scd - universit Toulouse III user on 20 March 2019

and 2 Yemeni from Abyan and Hadramawt region) (Černý et al. 2008, 2016), and 3) previously published studies.

Samples from the Southeast Borneo Banjar ($n = 99$) were collected from healthy unrelated adult donors during the 2013 field season. We followed sampling and DNA extraction procedures as described previously (Kusuma et al. 2016). Complete mtDNA sequences were generated for all Banjar samples following the protocol described in Brucato et al. (2017). Briefly, double bar-coded libraries were prepared and enriched for mtDNA, as described previously (Maricic et al. 2010; Kircher et al. 2012). Consensus sequences were obtained after base-calling, quality filtering, and further quality control steps to obtain consensus sequences, as described previously (Arias et al. 2018).

Published databases were screened for mtDNA HVS-I sequences affiliated with haplogroup B4a1a1 and its subclades. In total, 14,461 samples from 186 populations were screened from East Africa and Southwestern Eurasia (Černý et al. 2016) and 2,785 samples from 35 populations from Southeast Asia (Kusuma et al. 2015, 2016, 2017). We identified two samples from Yemen and one sample from Somalia affiliated with the Polynesian motif (B4a1a1) based on HVS-I sequences (Černý et al. 2016). These samples were whole genome amplified with the Illustra GenomiPhi V2 kit (GE Healthcare) following the manufacturer's instructions. The amplified samples underwent whole mtDNA sequencing following the methodology and checking procedures described in Brandão et al. (2016) for the Somali sample and in Brucato et al. (2017) for the Yemeni samples.

A comparative data set was built by compiling all of the published complete mtDNA sequences affiliated with haplogroup B4a1a1 and its subclades by screening the main web-based mtDNA databases (DDBJ/EMBL/GenBank international nucleotide sequence database; Phylotree [van Oven and Kayser 2009]; Family Tree DNA https://www.familytreedna.com/). The final data set included 1,086 mitogenomes from haplogroup B4a1a1, among which 7 belong to haplogroup B4a1a1b (the Malagasy motif), including the 3 new sequences from Somalia and Yemen (supplementary table S1, Supplementary Material online).

All new Yemeni, Somali, and Banjar sequences ($n = 102$) and comparative sequences were then analyzed and aligned against the revised Cambridge Reference Sequence (Andrews et al. 1999) using MAFFT aligner v.7 (Katoh and Standley 2013). Mitochondrial haplogroups were determined with the HaploGrep webtool (Kloss-Brandstätter et al. 2011) based on Phylotree Build 17 (van Oven and Kayser 2009). Maps showing the geographic distribution of the haplogroups were generated with Surfer (version 8, Golden Software, Inc., Golden, CO; https://www.goldensoftware.com/).

A maximum parsimony tree was constructed using the B4a1a1b mitogenomes from the new Yemeni and Somali samples and the published sequences (supplementary table S1, Supplementary Material online) guided by published principles (van Oven and Kayser 2009). In order to estimate the Time to the Most Recent Common Ancestor (TMRCA) of the clades, we used maximum likelihood and the ρ statistic. We performed maximum likelihood estimates of branch lengths using PAML v.4 (Yang 1997), assuming the HKY85 mutation model (excluding indels, and hotspot mutations) with gamma-distributed rates as done previously (Soares et al. 2009). We considered two partitions so as to differentiate the fast evolving HVS-I and HVS-II regions from the rest of the mtDNA genome. We also used the ρ statistic with a mutation rate estimate for the complete mtDNA sequence of one substitution every 3,624 years (Soares et al. 2009) and a mutation rate for synonymous mutations only at one substitution every 7,884 years. Standard errors were calculated based on 95% confidence intervals ($\rho \pm 1.96 \times \sigma^2$) as in Saillard et al. (2000).

Genome-Wide Analysis

The three new samples (two from Yemen and one from Somalia) that carry a maternal lineage affiliated with the Malagasy motif were assayed for genome-wide single nucleotide polymorphism (SNP) genotypes using the Illumina Human Omni5 Bead Chip (Illumina), which surveys 4,284,426 single nucleotide markers regularly spaced across the genome.

We gathered comparative genome-wide data from previously published studies of populations from Africa, Madagascar, the Middle East, ISEA, Southeast Asia, South Asia, East Asia, and Europe (supplementary table S2, Supplementary Material online). Before analysis, data quality controls were performed using PLINK v.1.9 (Chang et al. 2015): 1) to avoid close relatives, relatedness was measured between all pairs of individuals within each population using an identity-by-descent estimation with an upper threshold of 0.25 (second-degree relatives); 2) SNPs that failed the Hardy–Weinberg exact test ($P < 10^{-6}$) in each group were excluded; and 3) samples with an overall call rate <0.99 and individual SNPs with missing rates >0.05 across all samples within each population were excluded. Two data sets were subsequently constituted. For frequency-based analyses (ADMIXTURE, *f*3-statistics), our data set included 3,480 individuals from 193 populations genotyped for 171,728 SNPs, obtained after pruning for variants in high linkage disequilibrium (LD) with Plink v.1.9 ($r^2 > 0.5$; 50 SNP sliding windows). For haplotype-based analyses (PCAdmix), the data set included 411,442 SNPs for 524 individuals, representing 5 metapopulations (Southeast Africa, East Africa, Middle East, South Asia, and ISEA), the Yemeni and Somali populations, and the two Yemeni individuals and the Somali individual of interest. The five metapopulations are composed of 50 individuals with Southeast African Bantu ancestry (randomly selected from Kenyan Luhya, South African Bantu, and Swahili groups), East African (horn of Africa) ancestry (randomly selected

Downloaded from https://academic.oup.com/gbe/article-abstract/11/3/748/5306180 by scd - universit Toulouse III user on 20 March 2019

from Oromo, Ethiopian from Somalia), Middle Eastern ancestry (randomly selected from Oman and Saudi Arabia), South Asia (randomly selected from Gujarat Brahmin), and Indonesian ancestry (randomly selected from Indonesian Banjar, Samihim, and Malay). Genotypes were then phased with SHAPEIT v.2 (Delaneau et al. 2011) using the 1000 Genomes Project phased data (Delaneau et al. 2014) as a reference panel and the HapMap phase II genetic map.

To address specific questions regarding the ancestries of these three individuals, we performed ADMIXTURE analyses (Alexander et al. 2009), with default settings, for components $K = 2$ to $K = 30$. Ten iterations with randomized seeds were run and compiled with CLUMPAK v.1 (Kopelman et al. 2015). The minimum average cross-validation value was used to define the most informative K component (here, $K = 26$).

To estimate a lower bound of the time when the ancestors of these specific Yemeni (Y115 and Y270) and Somali (S25) individuals inherited the Asian component, an exponential decay function was used. This approach estimates the decline in total genome-wide Asian ancestry proportion due to backcrossing with non-Asian marriage partners and should not be confused with LD-based dating approaches that also employ a decay function. The method used here assumes a single admixture event in the ancestry of a given individual t generations in the past, followed by backcrossing at each subsequent generation to individuals with no Asian ancestry. Formally:

$$N_t = N_0 \left(\frac{1}{2}\right)^{t/t_{1/2}},$$

where N_t is the frequency of the total genome-wide Asian ancestry proportion at time t, N_0 is the initial Asian ancestry proportion (either 100% for an unadmixed individual originating in Indonesia or 37% for an individual originating in Madagascar [Brucato et al. 2016; Pierron et al. 2017]), t is the time interval (in generations, comprising a generation time of 29 years; Fenner 2005) years and $t_{1/2}$ is the half-life (here, defined as 1 due to backcrossing to an entirely non-Asian background at each generation).

To identify potential admixture events, three-population ($f3$) statistics (Pickrell and Pritchard 2012) were computed using each new sample as recipient and two population sources from the genome-wide data set, with one of the sources representing the Asian input (Comoros, Madagascar, or Banjar) and the other source representing one of their respective parental or neighboring populations. The population trios yielding a Z-score smaller than -2 were considered significantly admixed.

Local ancestry analysis in the three individuals was performed with PCAdmix v.1.050 with the five parental metapopulations (Southeast African Bantu, East African, Middle Eastern, South Asia, and Indonesian). The phased data were screened with LD information so that the probability of common ancestry of each haplotype with each "parental" metapopulation could be defined. The Viterbi algorithm was then used to identify local ancestries of all haplotypes in the two Yemeni and one Somali individuals, as well as in the Yemeni and Somali populations (excluding the three individuals of interest). This allows estimation of the proportion of Malagasy ancestry (defined as Southeast African Bantu and Indonesian associated fragments, based on Malagasy genetic diversity history, Brucato et al. 2016) and Indonesian ancestry (Indonesian fragments isolated from Southeast African fragments) in these three individuals compared with their respective parental populations. A Z-score was computed to estimate the deviation of Malagasy/Asian proportions in the three individuals of interest in comparison to the rest of their population.

Results

mtDNA Analyses

The geographic distribution of complete mtDNA sequences affiliated with the Malagasy motif (B4a1a1b haplogroup), its precursor haplogroup B4a1a1 (the Polynesian motif), and related subclades (supplementary table S1, Supplementary Material online) shows a clear phylogeographic distribution pattern (fig. 1). On the one hand, the Polynesian motif and all of its subclades, with the exception of the Malagasy motif subclade (B4a1a1b), are only found east of Wallace's line (with minor exceptions). On the other hand, the Malagasy motif is only detected in the WIO (Madagascar, and now Somalia and Yemen).

In Island Southeast Asia, the Banjar, who have been shown to be linguistically (Adelaar 1989, 2017) and genetically (Brucato et al. 2017) the descendants of the ancestors of the Asian background found in Madagascar and the Comoros, carry no maternal lineage affiliated with the Malagasy motif, and only one individual (out of 99) carries a maternal lineage affiliated to the Polynesian motif (supplementary table S3, Supplementary Material online). The other Banjar mitochondrial haplogroups are similar to the observed haplogroup diversity in other populations in Indonesia (Tumonggor et al. 2013; Kusuma et al. 2015).

The phylogenetic tree based on the B4a1a1b high-quality mitogenomes available (fig. 1 and supplementary fig. S1, Supplementary Material online) shows the absence of geographic structuring among these sequences, together with their very low diversity. The two sequences from Yemen and Madagascar are identical to the basal B4a1a1b lineage and the other sequences carry only one additional mutation each, mostly from mutational hotspots in the hypervariable noncoding region (16291T, 16357C, 16368C, and 16222T; Lott et al. 2013). This low diversity is in agreement with the recent coalescence age estimated for haplogroup B4a1a1b of

Downloaded from https://academic.oup.com/gbe/article-abstract/11/3/748/5306180 by scd - universit Toulouse III user on 20 March 2019

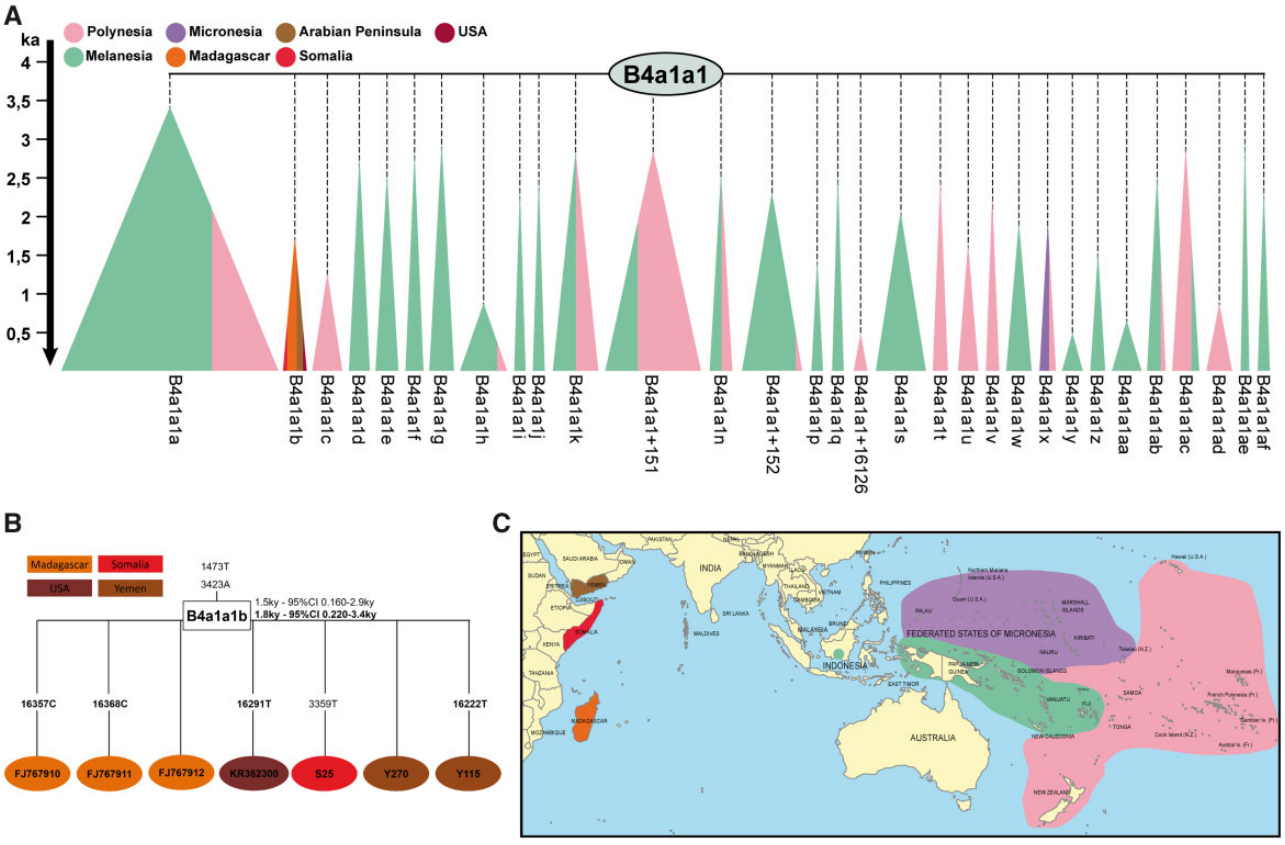

Fig. 1.—Schematic representations of the B4a1a1 phylogeny based on complete mtDNA sequences. Subclades are represented by triangles. Subclades are colored according to their geographic origin, as shown on the map. (*A*) B4a1a1 tree. (*B*) B4a1a1b tree. (*C*) Geographic distribution of the subclades. KA, thousand years ago.

Table 1

TMRCA Age Estimates Using ML and Rho (ρ) for B4a1a1 and B4a1a1b

		TMRCA Age Estimate (Years)					
		ML		Rho Complete		Rho Synonymous	
Haplogroup	*n*	Age	95% CI	Age	95% CI	Age	95% CI
B4a1a1[a]	7	3,796	2,961–4,636	6,191	4,121–8,289	6,648	1,129–12,166
B4a1a1b[b]	162	1,535	167–2,916	1,842	226–3,476	0	79–7,805

[a]Polynesian motif.
[b]Malagasy motif.

1,500–1,800 years BP (table 1 and supplementary table S4, Supplementary Material online), a narrower window than previously estimated (1,200–16,800 years BP; Razafindrazaka et al. 2010). This new age estimate predates, but is consistent with, the earliest Austronesian settlement of the East African offshore islands (Madagascar and Comoros), which is dated to 800–1,200 years BP from recent archaeological (Crowther et al. 2016) and genetic studies (Brucato et al. 2016, 2018).

Genome-Wide Analysis

To determine whether genome-wide genetic data can shed further light on the history of individuals carrying the Malagasy motif, the genetic ancestries present in the genome-wide data set were decomposed with ADMIXTURE v.1.3 (fig. 2). The lowest cross-validation values were obtained with 26 ancestries (supplementary fig. S2, Supplementary Material online). For clarity, using a reduced admixture data set including representative populations from each region and

Downloaded from https://academic.oup.com/gbe/article-abstract/11/3/748/5306180 by scd - universit Toulouse III user on 20 March 2019

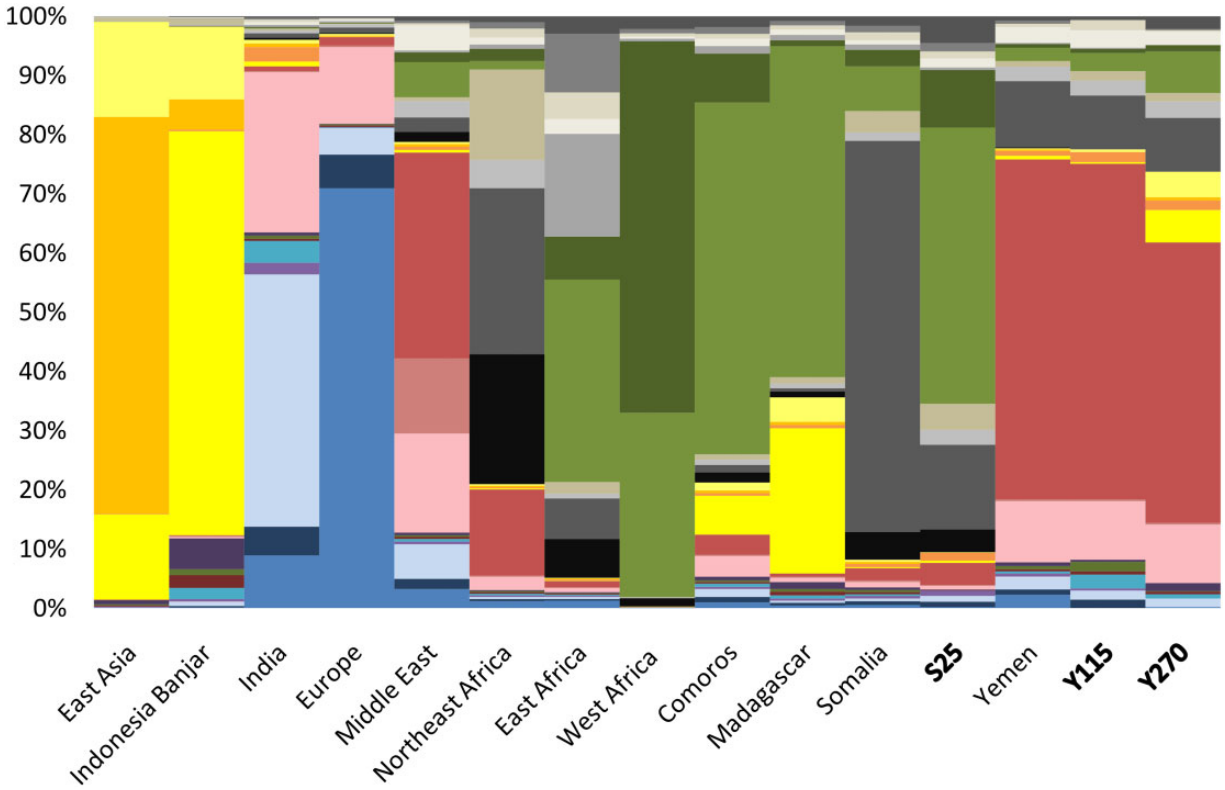

Fig. 2.—ADMIXTURE analysis ($K = 26$) for 15 population groups and the two Yemeni and one Somali individuals carrying the Malagasy motif B4a1a1b maternal lineage. Each colored line represents a sampled population whose genetic background can be decomposed into 26 genetic components.

the population of interest, $K = 4$ also displays the main African, European, South Asia, and Southeast Asian ancestries (supplementary fig. S3, Supplementary Material online). We observed that the Somali (S25) and two Yemeni (Y270 and Y115) individuals have relatively similar ancestry patterns to their local populations, including a dual Asian ancestry reflected by a South Asian (India) and Southeast Asian (Indonesia) components (fig. 2 and supplementary fig. S3, Supplementary Material online). Moreover, Yemenese and Somali do not have significantly more Asian ancestry (0.87% and 1.18%, respectively) than paired more inland populations [Student's *t*-test: Yemen vs. Saudi Arabia $t(54) = -0.67$, $P = 0.51$; Somalia vs. Ethiopia $t(30) = 1.9$, $P = 0.067$; Somalia vs. Egypt $t(23) = -0.43$; $P = 0.68$; Student 1908] (supplementary fig. S4, Supplementary Material online). The two Yemeni individuals have one major component (pinky red gradient: 65–80%) shared with other Middle East populations. Other minor ancestries are shared with populations from the horn of Africa (gray gradient: 13%), sub-Saharan Bantu speakers (green gradient: 7–10%), and Europeans (blue gradient 2–5%). The Asian component (yellow gradient) represents just a small component: from 1% in Y115—similar to those in Yemeni populations, up to 10% in Y270. The Somali individual has one major component (green gradient: 60%) shared with Bantu speakers, and a smaller component shared with populations from the horn of Africa (gray gradient: 25%), the Middle East (pink red gradient: 5%), and very minor components from Asia (yellow gradient <1%).

The admixture scenario from *f*3 statistics (supplementary tables S5–S7, Supplementary Material online) shows that the two Yemeni (Y115 and Y270) and the Somali (S25) samples cannot be modeled as any kind of simple mixture, using any Comoros, Madagascar, or Banjar groups as one population source and one of their respective parental or neighboring populations as the other source. To explore their admixture further, the admixture history and Asian component of these individuals was subjected to PCAdmix analysis (table 2). This indicated that two individuals (Y115 and S25) have significantly more Malagasy fragments than are observed in their respective parental populations (*Z*-score = 1,572 and 5,556, respectively), thus suggesting a likely Malagasy origin of their Asian ancestry. The other Yemeni individual (Y270) carries more Indonesian fragments than observed in its parental population (i.e., Asian and African Bantu fragments are not associated) (*Z*-score = 8,368), suggesting an origin of its Asian component from a population with limited African Bantu admixture, and thus excluding the Malagasy as a source (in average, Malagasy populations

Downloaded from https://academic.oup.com/gbe/article-abstract/11/3/748/5306180 by scd - universit Toulouse III user on 20 March 2019

Table 2

Malagasy (African Bantu and Indonesian Banjar Associated Fragments) and Indonesian Fragments Observed in the Yemeni and Somali Individuals, Compared with Their Respective Parental Population, Based on PCAmix Ancestry Results

ID	Observed Malagasy Fragments	Averaged Observed Malagasy Fragments in Yemeni or Somali Population	*Z*-Score	Observed Indonesian Fragments	Averaged Observed Asian Fragments in Yemeni or Somali Population	*Z*-Score
Yemeni_Y115	0.028	0.008	1.572	0.039	0.036	0.158
Yemeni_Y270	0.019	0.008	0.875	0.216	0.036	8.368
Somali_S25	0.068	0.004	5.556	0.002	0.029	−1.776

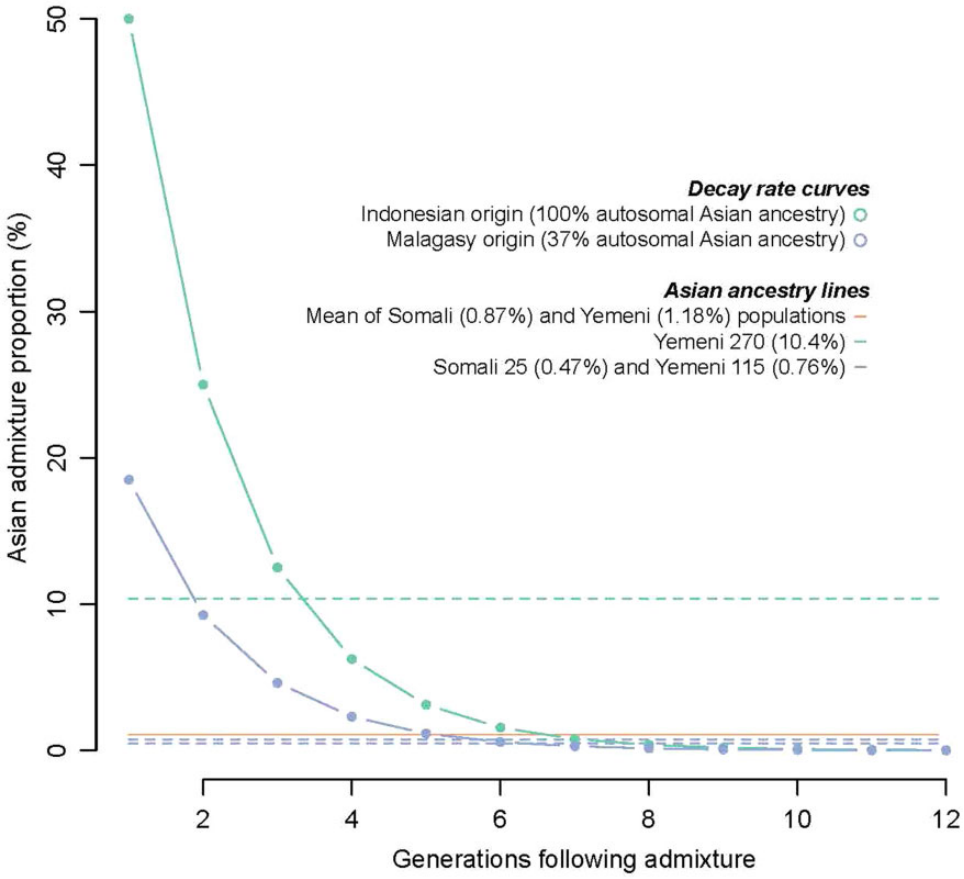

Fig. 3.—Decay curves showing expected autosomal Asian ancestry at each generation of intermarriage to partners with no Asian ancestry. Decay rates for individuals with Indonesian (green) and Malagasy (blue) origins. Horizontal lines mark observed Asian ancestry for the Somali and Yemeni populations (orange), Indonesian descendant Yemeni 270 (green), and Malagasy descendants Somali 25 and Yemeni 115 (blue).

analyzed so far have around 60% African Bantu ancestry and 40% Asian ancestry; Pierron et al. 2014, 2017; Brucato et al. 2016, 2018).

The date of inheritance of the Asian component found in the Yemeni and Somali Malagasy motif carriers can be estimated using an exponential decay function to explore the decline in total genome-wide Asian ancestry. Figure 3 shows that the two individuals with <1% Asian SNPs (Y115 and S25) likely had an Asian ancestor before 5–7 generations ago (lower bound prior to 250–190 years BP) and probably much older. As the Asian ancestry in these individuals is similar to other individuals in the local population who do not carry the Malagasy motif, the inheritance is possibly far earlier. In contrast, the individual with 10% Asian SNPs (Y270) likely had an Asian ancestor 3–4 generations ago (160–130 years BP), and because his total Asian ancestry is much higher than others in his local population (which has similar Asian ancestry to neighboring populations, supplementary fig. S4, Supplementary Material online), the inheritance date is probably relatively close to this date estimate. There may therefore

Downloaded from https://academic.oup.com/gbe/article-abstract/11/3/748/5306180 by scd - universit Toulouse III user on 20 March 2019

be multiple reasons why the Malagasy motif was inherited by these three individuals.

Surprisingly, these results suggest that despite the sharing of a clear Asian component, from their maternal lineage and autosomal DNA, the geographic origin and timing of admixture of this Asian component may be different among the three Yemeni and Somali individuals, revealing a more complex dynamic of Austronesian gene flow in the Western Indian Ocean region.

Discussion

Our results bring new information on the Austronesian dispersal westward across the Indian Ocean. We show that the key maternal lineage B4a1a1b, the so-called Malagasy motif, has a coalescence age (1,500–1,800 years BP) either predating or very early during the period of Austronesian arrivals to the East African offshore islands of Madagascar and the Comoros (table 1 and supplementary table S4, Supplementary Material online and fig. 1). This age estimate gives some temporal support for the emergence of the Malagasy motif in Island Southeast Asia (also in agreement with the fact that demographic expansion predates the geographic expansion). Although it remains undetected in the Banjar population, leaving its population source in ISEA unknown (supplementary table S3, Supplementary Material online), its recent contribution detected in this study (nineteenth century) from a population without African admixture implies that it should probably still exist in some location in ISEA.

The absence of B4a1a1b in the Banjar is surprising as it represents the most frequent Asian maternal lineage in Madagascar (22%, Razafindrazaka et al. 2010; Cox et al. 2012), and considering the cultural, linguistic and genetic links between the Banjar and Madagascar/Comoros populations (Adelaar 1989, 2017; Beaujard 2012a, 2012b; Brucato et al. 2016, 2018), this lineage was expected to be present in the Asian parental population (the Banjar). Its absence may suggests that the main Austronesian maternal lineage (B4a1a1b) found in the WIO may not have originated from the Banjar population. However, an origin of the Malagasy motif in situ in Madagascar or the Comoros after the arrival of the Polynesian motif carriers is also unlikely (Razafindrazaka et al. 2010), as it requires that 1) two mutations (1,473 and 3,423) arose in Madagascar in the last 1,300–1,000 years (Brucato et al. 2018), 2) diffused into population groups across the entire island, and 3) the Malagasy motif precursor (the Polynesian motif), later disappeared from these populations.

This "missing" lineage in Southeast Borneo 1) may have been lost in Indonesia due to genetic drift in small island populations or 2) may remain undetected due to the lower sampling coverage in Borneo compared with other studied regions. As suggested previously (Razafindrazaka et al. 2010; Cox et al. 2012; Kusuma et al. 2015), the motif may have emerged in eastern Indonesia—where the Polynesian motif, its immediate precursor, is geographically restricted—and later brought into the Banjar gene pool. This may have occurred during long distance trading activities and admixture among ISEA regional populations favored by the emergence Hindu Malay Kingdoms, such as Śrīvijaya (sixth to thirteenth centuries), which developed the Banjarmasin trading post in Southeast Borneo (Ras 1968; Beaujard 2012a; Brucato et al. 2017). This eastern Indonesian input into the Southeast Borneo Banjar finds some support in the fact that the Banjar are one of the rare populations in western Indonesia, along with the sea nomad Bajo from Borneo, to carry the Polynesian motif (supplementary table S3, Supplementary Material online, Tumonggor et al. 2013; Kusuma et al. 2015, 2017), demonstrating their previous connection with eastern Indonesians.

In the Indian Ocean rim west of Island Southeast Asia (Mainland Southeast Asia and the Indian subcontinent), we note the absence of the Polynesian motif or any of its subclades (fig. 1). This suggests that the Austronesian population that brought the Polynesian Motif subclade B4a1a1b to the WIO (e.g., to Madagascar) did not leave detectable genetic traces on the northern Indian Ocean rim, in agreement with an absence of Indonesian gene flow to this region as inferred from recent genome-wide data (Brucato et al. 2017). It is still unclear whether this is due to Indonesian populations restricting their interactions in these regions to trading and/or cultural activities, or because Indonesian traders used a more direct route across the Indian Ocean to cover the 7,500 km between Indonesia and Madagascar and the Comoros, which is possible based on ocean current and monsoon weather patterns (Fitzpatrick and Callaghan 2008). Both scenarios merit further study.

For the first time, we observe the Malagasy motif outside of Madagascar, in one individual from East Africa (Somalia) and two from the South Arabian Peninsula (Yemen). To date, a long-term Austronesian presence in the Western Indian Ocean region has only been supported on the island territories of Madagascar and the Comoros and is very tenuous on the African continent and Arabian Peninsula, from genetic (Brucato et al. 2017, 2018), historical (Beaujard 2012a, 2012b), and archaeological (Crowther et al. 2016) perspectives. Our results suggest clear, but limited impact, of Austronesian genetic input into these East African and South Arabian regions (0.02%, 3 out of 14,461 individuals from mtDNA Austronesian key marker, up to around 1% when considering the entire Asian autosomal signal), but leaves open the question of their origin.

When considering the Asian component in the autosomal DNA, beyond the fact that no more than 1% of Asian substrate is detected in East Africa/Arabia (fig. 2 and supplementary fig. S3, Supplementary Material online), the three new Malagasy motif carriers displayed a more complex pattern. For two individuals (Y115 and S25), an inheritance from Malagasy individuals before the late eighteenth century and potentially

Downloaded from https://academic.oup.com/gbe/article-abstract/11/3/748/5306180 by scd - universit Toulouse III user on 20 March 2019

much earlier is likely. For the other individual (Y270), more recent gene flow in the late nineteenth century from a different population source (likely with primary Asian ancestry) seems to be the most parsimonious explanation (figs. 2 and 3 and table 2).

This complex pattern of Asian genetic input suggests that these three individuals obtained their Asian genetic ancestry from two different processes or events, via both a primary and secondary Asian input to the East African coast and South Arabian Peninsula. This involved at least two different population sources: an African–Asian admixed population (e.g., potentially from Madagascar) for the oldest admixture event; and a population without African admixture as the source of the most recent admixture event. These two events likely took place during the last millennium with the intensification of the Indian Ocean trading network among Africa, the Middle East and Asia, leading to the exchange of ideas, goods, and people (Beaujard 2012a, 2012b; Brucato et al. 2017), the arrival of the Austronesians in the Western Indian Ocean region around the eighth to thirteenth centuries AD, and the development of the Swahili corridor (from southern Somalia in the north to the Comoros archipelago, Madagascar, and Central Mozambique in the south; Horton 1987; Brucato et al. 2018).

Although the Asian inheritance for two individuals (Y115 and S25) may reflect the development of trading posts by Arab merchants in the East African region, and deportation of African slaves northward to Arabia and South Asia (Hellenthal et al. 2014; Blench 2014; Brucato et al. 2017), we cannot reject an older origin from the early Austronesian period. The Asian ancestry of the other individual (Y270) represents a different pattern with a likely origin in Island Southeast Asia, which might be explained by the emergence of several sultanates in the fifteenth to sixteenth centuries AD in East Africa and in ISEA (Beaujard 2012b). Diaspora communities spread around the Indian Ocean rim may have played an important role to develop trading activities and also to maintain contact with their native land (Beaujard 2012b). The Hadrami community from Hadramawt in Southeast Yemen—to which the Yemeni individual (Y270) belongs—emerge as one plausible cause to reconcile the different geographic and temporal origins of the Asian inheritance found in the three Yemeni and Somali individuals. The Hadrami diaspora, established in the Comoros at the end of the first millennium (Beaujard 2012b) in South Asia (Gujarat in India) and in Island Southeast Asia (Malacca in Malaysia) from the beginning of the second millennium, was involved in trading activities between Africa, Middle East, and Asia. Despite strict intra community marriage rules, admixture with local population was frequent, as the men could choose their bride from among the local African and Asian populations (Boxberger 2002). In the nineteenth century, some economically successful Hadramis (i.e., from Singapore and Hyderabad, and other places of the diaspora) returned to their native land in the Yemen bringing back their admixed family (Manger 2010).

The pattern emerging is that trading activities and diaspora communities may have favored small scale genetic admixture through direct contact between distant regions and populations from the Arabian Peninsula, Africa, and Island Southeast Asia, and that these involved Indonesian populations carrying this peculiar maternal lineage, the Malagasy motif B4a1a1b.

Conclusions

Beyond confirming the usefulness of uniparental genetic markers to identify past contact events, this study provides the first clear evidence for the presence of Austronesian genetic input in Eastern Africa and South Arabia, beyond the African offshore islands of Madagascar and the Comoros. This input may have spread firstly during an early stage (prior to the late eighteenth century) and secondly at a later stage (end of the nineteenth century) from the Austronesian presence in the Western Indian Ocean region, from Madagascar and directly from an unknown location in Island Southeast Asia. Moreover, this suggests that beyond the main and early well identified Austronesian dispersal event westward into the Indian Ocean, later and more specific small scale events (e.g., Arab trading activities, diaspora communities) favor limited genetic exchanges and spread of the Austronesian genetic component within the Western Indian Ocean rim populations. More genetic data from the Indian Ocean rim populations are necessary to better understand and model these multiple episodes of Austronesian genetic admixture within the Indian Ocean, including their date of admixture, routes taken, and the population and geographic scale of these contact events.

Acknowledgments

We thank Alexander Hübner, Enrico Macholdt, and Roland Schröder for assistance with generating the mtDNA sequences. We acknowledge support from the GenoToul bioinformatics facility of Genopole Toulouse Midi-Pyrénées, France. We thank the CNRGH production team for genotyping data generation. This work was supported by French Ministry of Research grant ANR-14-CE31-0013-01 (OCEOADAPTO) to F.-X.R, the French Ministry of Foreign and European Affairs (French Archaeological Mission in Borneo (MAFBO) to F.-X.R.), and the French Embassy in Indonesia through its Cultural and Cooperation Services (Institut Français en Indonésie); a fellowship from the Alexander von Humboldt Foundation to M.P.C.; funding from the Max Planck Society to M.S.; and grants from COMPETE 2020 and a Fundação para a Ciência e a Tecnologia-funded project (POCI-01-0145-FEDER-016609) to V.F.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Downloaded from https://academic.oup.com/gbe/article-abstract/11/3/748/5306180 by scd - universit Toulouse III user on 20 March 2019

Literature Cited

Adelaar KA. 1989. Malay influence on Malagasy: linguistic and culture-historical implications. Oceanic Linguistics 28(1):1–46.

Adelaar KA. 2017. Who were the first Malagasy, and what did they speak? In: Acri A, Blench R, Landmann A, editors. Spirit and ships: cultural transfer in early monsoon Asia. Singapore: Institute of Southeast Asian Studies. p. 441–469.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19(9):1655–1664.

Andrews RM, et al. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 23(2):147.

Arias L, Barbieri C, Barreto G, Stoneking M, Pakendorf B. 2018. High resolution mitochondrial DNA analysis sheds light on human diversity, cultural interactions, and population mobility in Northwest Amazonia. Am J Phys Anthropol. 165(2):238–255.

Beaujard P. 2012a. Les mondes de l'océan Indien. Vol. 1: De la formation de l'Etat au premier système-monde afro-eurasien (4e millénaire av. J.-C.-6e siècle apr. J.-C.). Paris: Armand Collin Press.

Beaujard P. 2012b. Les mondes de l'océan Indien. Vol. 2: L'océan Indien, au cœur des globalisations de l'Ancien Monde (7e-15e siècles). Paris: Armand Collin Press.

Bellwood P. 2017. First islanders: prehistory and human migration in Island Southeast Asia. Hoboken (NJ): Wiley Blackwell Press.

Blench R. 2014. Tracking the origins of African slaves in the Indian Ocean through personal names: the evidence of Sumatra records. In: Vernet T, Beaujard P, editors. Afriques on East Africa and the Indian Ocean. p. 2–18. http://www.rogerblench.info/Archaeology/Indian%20Ocean/Tracking%20the%20origins%20of%20African%20slaves%20in%20the%20Indian%20Ocean%20through%20personal%20names.pdf.

Boxberger L. 2002. On the edge of empire: hadhramawt, emigration, and the Indian Ocean, 1880s–1930s. New York: State University of New York Press.

Brandão A, et al. 2016. Quantifying the legacy of the Chinese Neolithic on the maternal genetic heritage of Taiwan and Island Southeast Asia. Hum Genet. 135(4):363–376.

Brucato N, et al. 2016. Malagasy genetic ancestry comes from an historical Malay trading post in Southeast Borneo. Mol Biol Evol. 33(9):2396–2400.

Brucato N, et al. 2017. Genomic admixture tracks pulses of economic activity over 2,000 years in the Indian Ocean trading network. Sci Rep. 7(1):2919.

Brucato N, et al. 2018. The Comoros shows the earliest Austronesian gene flow in East Africa. Am J Hum Genet. 102(1):58–68.

Černý V, Čížková M, Poloni ES, Al-Meeri A, Mulligan CJ. 2016. Comprehensive view of the population history of Arabia as inferred by mtDNA variation. Am J Phys Anthropol. 159(4):607–616.

Černý V, et al. 2008. Regional differences in the distribution of the sub-Saharan, West Eurasian, and South Asian mtDNA lineages in Yemen. Am J Phys Anthropol. 136(2):128–137.

Chang CC, et al. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4:7.

Cox MP, Nelson MG, Tumonggor MK, Ricaut F-X, Sudoyo H. 2012. A small cohort of Island Southeast Asian women founded Madagascar. Proc R Soc B Biol Sci. 279(1739):2761–2768.

Crowther A, et al. 2016. Ancient crops provide first archaeological signature of the westward Austronesian expansion. Proc Natl Acad Sci U S A. 113(24):6635–6640.

Delaneau O, Marchini J; 1000 Genomes Project Consortium. 2014. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat Commun. 5:3934.

Delaneau O, Marchini J, Zagury J-F. 2011. A linear complexity phasing method for thousands of genomes. Nat Methods. 9(2):179–181.

Duggan AT, et al. 2014. Maternal history of Oceania from complete mtDNA genomes: contrasting ancient diversity with recent homogenization due to the Austronesian expansion. Am J Hum Genet. 94(5):721–733.

Fenner JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Am J Phys Anthropol. 128(2):415–423.

Fitzpatrick SM, Callaghan R. 2008. Seafaring simulations and the origin of prehistoric settlers to Madagascar. In: Clark G, Leach F, O'Conno S, editors. Islands of inquiry: colonisation, seafaring and the archaeology of maritime landscapes. Australia: ANU Press. p. 47–58.

Hellenthal G, et al. 2014. A genetic atlas of human admixture history. Science 343(6172):747–751.

Horton M. 1987. The Swahili Corridor. Sci Am. 255:86–93.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Kircher M, Sawyer S, Meyer M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 40(1):e3.

Kloss-Brandstätter A, et al. 2011. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum Mutat. 32(1):25–32.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour. 15(5):1179–1191.

Kusuma P, et al. 2015. Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations. BMC Genomics. 16:191.

Kusuma P, et al. 2016. Contrasting linguistic and genetic origins of the Asian source populations of Malagasy. Sci Rep. 6:26066.

Kusuma P, et al. 2017. The last sea nomads of the Indonesian archipelago: genomic origins and dispersal. Eur J Hum Genet. 25(8):1004–1010.

Lott MT, et al. 2013. mtDNA variation and analysis using Mitomap and Mitomaster. Curr Protoc Bioinformatics 1(123):1.23.1–1.23.26.

Manger LO. 2010. The Hadrami diaspora: community-building on the Indian Ocean rim. New York: Berghahn Books.

Maricic T, Whitten M, Paabo S. 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS One 5(11):e14004.

Mazières S, et al. 2018. Genes flow by the channels of culture: the genetic imprint of matrilocality in Ngazidja, Comoros Islands. Eur J Hum Genet. 26(8):1222–1226.

Msaidie S, et al. 2011. Genetic diversity on the Comoros Islands shows early seafaring as major determinant of human biocultural evolution in the Western Indian Ocean. Eur J Hum Genet. 19(1):89–94.

Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8(11):e1002967.

Pierron D, et al. 2014. Genome-wide evidence of Austronesian-Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar. Proc Natl Acad Sci U S A. 111(3):936–941.

Pierron D, et al. 2017. Genomic landscape of human diversity across Madagascar. Proc Natl Acad Sci U S A. 114(32):E6498–E6506.

Ras JJ. 1968. Hikajat Banjar: a study in Malay historiography. The Hague (The Netherlands): Martinus Nijhoff.

Razafindrazaka H, et al. 2010. Complete mitochondrial DNA sequences provide new insights into the Polynesian motif and the peopling of Madagascar. Eur J Hum Genet. 18(5):575–581.

Saillard J, Forster P, Lynnerup N, Bandelt HJ, Nørby S. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet. 67(3):718–726.

Downloaded from https://academic.oup.com/gbe/article-abstract/11/3/748/5306180 by scd - universit Toulouse III user on 20 March 2019

Skoglund P, et al. 2016. Genomic insights into the peopling of the Southwest Pacific. Nature 538(7626):510–513.
Soares P, et al. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. Am J Hum Genet. 84(6):740–759.
Soodyall H, Jenkins T, Stoneking M. 1995. 'Polynesian' mtDNA in the Malagasy. Nat Genet. 10(4):377–378.
Student. 1908. The probable error of a mean. Biometrika 6:1–25.
Tofanelli S, et al. 2009. On the origins and admixture of Malagasy: new evidence from high-resolution analyses of paternal and maternal lineages. Mol Biol Evol. 6:2109–2124.
Tumonggor MK, et al. 2013. The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific. J Hum Genet. 58(3):165–173.
van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat. 30(2):E386–E394.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13(5):555–556.

Associate editor: Naruya Saitou

Downloaded from https://academic.oup.com/gbe/article-abstract/11/3/748/5306180 by scd - universit Toulouse III user on 20 March 2019