

A small cohort of Island Southeast Asian women founded Madagascar

Murray P. Cox^{1,*}, Michael G. Nelson¹, Meryanne K. Tumonggor²,
François-X. Ricaut³ and Herawati Sudoyo⁴

¹*Institute of Molecular BioSciences, Massey University, Private Bag 11 222, Palmerston North, New Zealand*

²*Department of Anthropology, University of Arizona, Tucson, AZ 85721, USA*

³*Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse, Centre National de la Recherche Scientifique, Université de Toulouse (Paul Sabatier), Toulouse, France*

⁴*Eijkman Institute for Molecular Biology, Jakarta, Indonesia*

The settlement of Madagascar is one of the most unusual, and least understood, episodes in human pre-history. Madagascar was one of the last landmasses to be reached by people, and despite the island's location just off the east coast of Africa, evidence from genetics, language and culture all attests that it was settled jointly by Africans, and more surprisingly, Indonesians. Nevertheless, extremely little is known about the settlement process itself. Here, we report broad geographical screening of Malagasy and Indonesian genetic variation, from which we infer a statistically robust coalescent model of the island's initial settlement. Maximum-likelihood estimates favour a scenario in which Madagascar was settled approximately 1200 years ago by a very small group of women (approx. 30), most of Indonesian descent (approx. 93%). This highly restricted founding population raises the possibility that Madagascar was settled not as a large-scale planned colonization event from Indonesia, but rather through a small, perhaps even unintended, transoceanic crossing.

Keywords: Madagascar; Indonesia; settlement; coalescent; mitochondrial DNA

1. INTRODUCTION

The colonization of Madagascar, an island nation located off the east coast of Africa, is one of the more unusual events in human prehistory. Malagasy, a term that denotes the island's people and their language, shows both biological [1] and linguistic [2] connections to east African populations. More surprisingly, there is unequivocal evidence of biological [3] and linguistic [4] links to Island Southeast Asia, specifically Indonesia. Although these dual African–Indonesian associations have been confirmed by numerous genetic, linguistic, ethnographic and archaeological studies, the process by which Madagascar was initially settled remains poorly understood [5]. Here, we tackle this question by simulating a genetic model of the initial settlement process, with a specific focus on the long-range Indonesian female contribution to modern Malagasy.

Archaeological and paleoenvironmental evidence shows that people settled Madagascar only recently. There may have been sporadic early arrivals from Africa with unknown connections to today's Malagasy [6,7], but a burst of continuous settlement activity is clearly in evidence around the middle of the first millennium AD [8–11]. Despite detailed historical documents for other Indian Ocean regions from at least the Roman era onwards [12], the first several centuries of Malagasy history passed by completely unrecorded. Madagascar and its people are not mentioned in any written source until the Arab geographer al-Idrīsī in AD 1165 [13]. Madagascar's extreme peripheral location on Indian Ocean trade routes may partly explain this historical silence; as late as the fifteenth century, the

Arab cartographer ibn Mājid described Madagascar as 'bordering on the inhabited regions of the world' [13]. Just decades later, the Portuguese explorer Vasco da Gama sailed past Madagascar in 1497 on the first European voyage to India [14], thereby ushering in an era of detailed European record keeping on later Malagasy history.

In the absence of early written records, attempts to reconstruct the history of Madagascar necessarily rely on indirect evidence. All Malagasy today speak dialects of an Austronesian language that traces its origin back to Indonesia [4], a fact traditionally attributed to the Jesuit priest Luis Mariano in 1614 [15], but noted as early as the twelfth century by al-Idrīsī [13]. Most of the Malagasy lexicon is ultimately sourced from Ma'anyan, a language spoken along the Barito River valley of southeast Borneo [16–18]. However, only a small landlocked community of ethnic Dayak currently speaks Ma'anyan, and this inland region of southeast Borneo offers little ethnographic or historical evidence for any sort of maritime tradition. It is perhaps unsurprising then that Malagasy harbours far broader linguistic contributions drawn from across the Indonesian archipelago [2]. Loan words from Sanskrit, all with local linguistic modifications via Javanese or Malay [2,19,20], hint that Madagascar may have been colonized by settlers from the Srivijaya Empire, a major regional power in western Indonesia (modern Java, Sumatra and Malaysia) from the sixth to thirteenth centuries AD. Consistent with this hypothesis, Malagasy borrows large numbers of words from Javanese; the regional lingua franca, Malay; and even languages from southern Sulawesi, an island near the centre of the Indonesian archipelago. Cultural evidence—including iron working techniques, outrigger boats, musical instruments such as

* Author for correspondence (m.p.cox@massey.ac.nz).

the xylophone, and the cultivation of rice, bananas, yams and taro (i.e. a ‘tropical food kit’)—all supports a strong Southeast Asian connection from at least the eighth century onwards [21–23].

Genetic evidence paints much the same picture. Early blood protein studies identified dual African and Asian contributions [24,25]. Later studies of mitochondrial DNA (mtDNA) and Y chromosome variation improved this geographical resolution. Soodyall *et al.* [26,27] firmly established an Island Southeast Asian connection when they located the Polynesian motif in Madagascar. Apart from Madagascar, this mtDNA haplotype is restricted to remote Oceania, with low frequencies in Melanesia and eastern Indonesia, and only sporadic occurrences as far west as Bali and Borneo [28]. All sampled Malagasy are now known to carry a variant of the Polynesian motif, termed the Malagasy motif (characterized by polymorphisms 1473 and 3423A), whose global geographical range remains unclear [29]. In terms of male history, Y chromosome haplogroup O—otherwise restricted to Southeast Asia and Oceania—has been identified in Madagascar [3,5] and the Comoros Islands, a small archipelago just off Madagascar’s northern tip [30]. Where genetic dating has been performed, these genetic connections to Indonesia are consistent with the temporal bounds suggested by archaeology and linguistics [5], although invariably with much larger CI.

In combination, a broad outline of Malagasy history—especially later periods—is beginning to emerge. However, we still lack any real understanding of the island’s earliest history, particularly its initial colonization. What was the settlement process like? Was Madagascar colonized via large-scale population movements, or through a smaller translocation? How many people originally founded the Malagasy? What proportion of these settlers was Indonesian? And when did they first arrive? Here, we use new genetic data from Indonesian source populations, and a coalescent model specifically designed for the settlement of Madagascar to address these unresolved questions.

2. METHODS

(a) Samples

The Indonesian samples analysed in this study have been described extensively elsewhere [31–36]. In brief, 2745 individuals were screened from 12 island groups spanning the Indonesian archipelago: Sumatra ($n = 40$), Nias ($n = 62$), Mentawai ($n = 126$), Java ($n = 49$), Bali ($n = 517$), Sulawesi ($n = 188$), Sumba ($n = 639$), Flores ($n = 453$), Lembata ($n = 92$), Alor ($n = 23$), Pantar ($n = 27$) and Timor ($n = 529$). Samples were obtained with written informed consent by H.S., M.T., Golfiani Malik, Wuryantari Setiadi and Loa Helena Suryadi of the Eijkman Institute for Molecular Biology (Jakarta, Indonesia), and J. Stephen Lansing of the University of Arizona (Tucson, AZ, USA), with the assistance of Indonesian Public Health clinic staff. Sample collection followed protocols for the protection of human subjects established by both the Eijkman Institute and the University of Arizona institutional review boards. Permission to conduct research in Indonesia was granted by the Indonesian Institute of Sciences. All genetic screening was performed by M.T. at the University of Arizona.

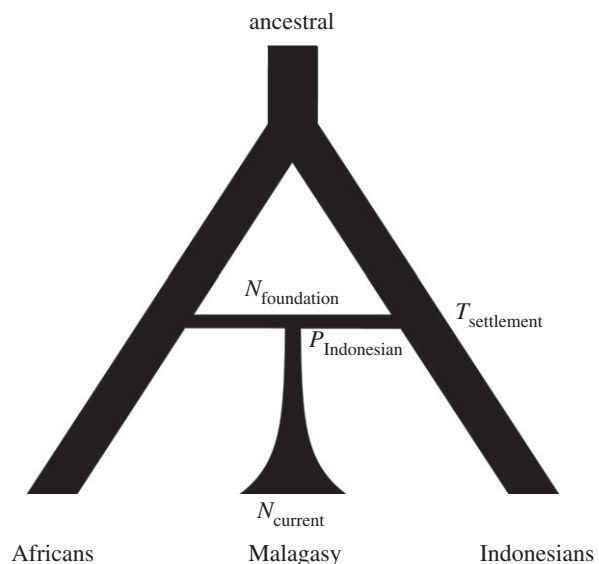


Figure 1. A demographic model of the settlement of Madagascar. An ancestral population splits into Africans and Indonesians. At some time, $T_{\text{settlement}}$, a proportion of the Indonesian population, $P_{\text{Indonesian}}$, and a proportion of the African population, $1 - P_{\text{Indonesian}}$, merge to form the founding population of Madagascar, $N_{\text{foundation}}$. This initial Malagasy population grows exponentially until it reaches its current effective population size, N_{current} .

(b) Genetic markers

Individuals carrying the Polynesian motif (haplogroup B4a1a1a) were identified by sequencing a 530 bp fragment of the mtDNA first hypervariable segment in all 2745 samples. All Polynesian motif carriers were screened for polymorphisms specific to the Malagasy motif, that is, restriction fragment length polymorphism (RFLP) losses at nucleotides 1473 (–1473 *HhaI*) and 3423A (–3423 *AciI*). Single nucleotide polymorphism (SNP) genotyping was performed using the method of Razafindrazaka *et al.* [29].

(c) Published data

Genetic diversity among Malagasy populations was drawn from previously published research [29]. In brief, 266 individuals were screened from three Malagasy ethnic groups: Mikea hunter–gatherers ($n = 127$), semi-nomadic Vezo fishermen ($n = 101$) and the dominant Andriana Merina ethnic group ($n = 38$). A detailed description of this dataset and associated research findings is available in the study of Razafindrazaka *et al.* [29].

(d) Demographic models

A hierarchical modelling approach was applied to determine the demographic parameters of Malagasy settlement. First, the historical relationships between African and Indonesian populations were determined. We adopted the simulation framework of Wollstein *et al.* [37], in which an ancestral African population ($N_{\text{ancestral}}$) split at some time (T_{split}) to form modern Africans (N_{African}) and Indonesians ($N_{\text{Indonesian}}$). We adjusted the Wollstein model to accommodate an Indonesian sample drawn from across the Indonesian archipelago, rather than just the island of Borneo, and accepted the African component of their model as it stands. Second, this African–Indonesian framework was extended to incorporate the Malagasy (figure 1). At some time ($T_{\text{settlement}}$), the Malagasy population ($N_{\text{foundation}}$) is formed from some proportion of Indonesian ($P_{\text{Indonesian}}$) and African ($1 - P_{\text{Indonesian}}$) settlers.

This Malagasy deme subsequently grows exponentially until it reaches its current effective population size (N_{current}).

(e) Coalescent simulation

Coalescent datasets were generated using the software *ms* [38]. The entire mtDNA genome was simulated using two-phase mutational scaling. First, the 530 bp fragment of the control region was simulated using a mutation rate of 8.0×10^{-6} mutations per site per generation [39] corrected for a generation interval of 25 years [40]. Second, the coding region was simulated by applying a mutation rate scalar of 1.57 coding region mutations to every control region mutation [41]. In the coalescent framework, the mutation rate and physical length of a locus are interrelated. We assume that the mitochondrial genome is not subject to recombination—the dominant view held by the research community today [42]. In this situation, the same coalescent outcome can be obtained by linearly scaling the mutation rate and physical length (e.g. doubling the mutation rate, while halving the length). We employ this scaling rule to simulate the entire mitochondrial genome because existing coalescent software cannot simulate multiple fully linked genetic regions with different mutation rates, but are instead flexible with regard to locus length. The control region rate was applied to the entire mitochondrial genome, and the coding region scaled by 1.57 times the control region length to reconstruct the correct overall ratio of coding to control region mutations. Soares *et al.* [41] provide a detailed explanation of this scaling factor.

(f) Data summaries

Summary statistics was developed based on unusual aspects of the Indonesian and Malagasy mtDNA datasets. First, the number of segregating sites— S summarizes the total length of the genealogy [43] and is an unbiased estimator of the population mutation rate ($\theta = N_e\mu$). The number of segregating sites in the Indonesian population was set as an acceptance criterion to control for the population mutation rate (i.e. it prevents acceptances of unreasonably large or small effective sizes) [44]. Simulations were accepted if the control region portion of the simulated dataset contained (i) the same number of segregating sites as in the observed Indonesian sample ($S = 186$) or (ii) segregating sites within 5 per cent of this value.

Second, the Malagasy data exhibit an unusual and distinctive pattern of mtDNA polymorphism. The Malagasy motif is a close relative of the Polynesian motif, but differs from it by two polymorphisms. Because mtDNA mutation rates only predict two mutation substitutions every 5216 years (using the corrected mutation rate of Soares *et al.* [41]) and archaeology constrains the settlement history of Madagascar to less than 2000 years, finding the derived Malagasy motif with two substitutions relative to the ancestral Polynesian motif is statistically unusual. Simulations were therefore accepted only if they contained a haplotype in the Malagasy sample that differed from an ancestral Indonesian haplotype by two changes in the coding region, but no changes in the control region.

Third, simulations were accepted only if simulated allele frequencies mirrored observed allele frequencies. The ‘ancestral’ haplotype in the simulated Indonesian sample was required to have a frequency less than or equal to the Polynesian motif frequency in the observed Indonesian data (less than or equal to 0.02). Similarly, the ‘derived’ haplotype in the simulated Malagasy sample was required to have a

frequency greater than or equal to the Malagasy motif frequency in the observed Malagasy data (greater than or equal to 0.22). Averaged across the Mikea, Vezo and Merina ethnic groups, this value is slightly lower than the B4a1a1a frequency observed for the Antandroy, Antanosy, Antaisaka and Merina (0.34) [5]. However, as Tofanelli *et al.* did not explicitly screen for the Malagasy motif, we chose to perform simulations using our more conservative estimate of the actual Malagasy motif frequency.

C++ code to calculate these summary statistics on *ms* coalescent simulations is available from the authors on request.

(g) Inferential statistics

Forty million coalescent simulations were run across a four-dimensional parameter space in the Malagasy settlement model. Values were drawn randomly from uniform distributions for $N_{\text{foundation}}$ (0, 2000), N_{current} (0, 8041), $T_{\text{settlement}}$ (0, 5000 years) and $P_{\text{Indonesian}}$ (0, 1). Only simulations matching all summary criteria (described above) were retained. Because simulated parameters were drawn from random uniform distributions, accepted simulations form an unbiased estimate of the likelihood. The peaks (modes) of reported probability densities are maximum-likelihood estimates (MLEs), and were calculated using kernel density estimation via the density function in the base package of R [45]. For each demographic parameter, 95% CI were calculated as asymmetric [0.025, 0.975] quantiles of the observed probability density.

3. RESULTS

(a) Malagasy motif in Madagascar

The Polynesian motif (mtDNA haplotype B4a1a1a) occurs at variable frequency among the ethnic groups (*foko*) of Madagascar and has been shown to be a direct translocation from Island Southeast Asia [3,29]. In a 2010 study, the Polynesian motif was identified in 58 of 266 (22%) Malagasy individuals, although its frequency varied among three ethnic groups: 50 per cent in Merina, 22 per cent in Vezo and 13 per cent in Mikea [29]. Whole mtDNA genome sequencing found that the Polynesian motif in Madagascar is characterized by two additional mutations (1473 and 3423A), which have been identified in all Polynesian motif carriers screened in Madagascar so far [29]. For this reason, this new haplotype has been termed the Malagasy motif and occurs in Madagascar at an average frequency of 22 per cent.

(b) Malagasy motif in Indonesia

To identify the distribution of the Malagasy motif in Indonesia, the diagnostic markers of the Polynesian and Malagasy motifs were screened in a large number of individuals ($n = 2745$) from across the Indonesian archipelago (table 1). Only 45 of 2745 sampled individuals (2%) carried the Polynesian motif, specifically on the islands of Bali ($n = 6$), Sumba ($n = 4$), Alor ($n = 1$) and Timor ($n = 34$). This is consistent with previous results, which show the Polynesian motif reaching greatest frequency in remote Oceania, with much reduced frequencies in eastern Indonesia and only sporadic occurrences in western Indonesia [46].

The 45 Polynesian motif carriers were screened for the diagnostic markers of the Malagasy motif. None carried either of the two additional mutations (1473 and 3423A). This is consistent with published results of full mtDNA

Table 1. Allele frequencies of the Polynesian and Malagasy motifs across the Indonesian archipelago.

population	sample size	Polynesian motif (n, %)	Malagasy motif n, %
Sumatra	40	0 —	0 —
Nias	62	0 —	0 —
Mentawai	126	0 —	0 —
Java	49	0 —	0 —
Bali	517	6 1.2	0 —
Sulawesi	188	0 —	0 —
Sumba	639	4 0.6	0 —
Flores	453	0 —	0 —
Lembata	92	0 —	0 —
Alor	23	1 4.4	0 —
Pantar	27	0 —	0 —
Timor	529	34 6.4	0 —
Total	2745	45 1.6	0 —

genome sequences; no Polynesian motif genomes outside Madagascar have been shown to carry the Malagasy motif [29,46]. We applied a Monte Carlo sampling error algorithm [31] to infer a possible upper bound on the allele frequency of the Malagasy motif across Indonesia as a whole. Even if present in Indonesia, the Malagasy motif is statistically unlikely to exceed a frequency of 0.1 per cent there. As this is only an inference, we do not use this value in our simulations. In its place, we focus on the concretely known frequency of the ancestral Polynesian motif in Indonesia. However, we conclude that (i) either the Malagasy motif arose in Madagascar and is therefore not present in Indonesia; or (ii) the Malagasy motif arose in Indonesia, but is either absent there or occurs only at an extremely low frequency there today.

(c) Demographic inference

A hierarchical series of coalescent models was developed to infer the demographic processes underpinning the initial settlement of Madagascar. Because this analysis is based on mtDNA evidence, we effectively report the history of Malagasy women. The inferential model was built around three especially unusual characteristics of the Malagasy and Indonesian mtDNA data: (i) the ancestral Polynesian motif occurs at very low frequency in Indonesia (this study); (ii) the derived Malagasy motif occurs at moderate-to-high frequency in Madagascar [29]; and (iii) the Malagasy motif differs from the Polynesian motif by two SNPs within the coding region [29]. Even given the relatively high rate of mtDNA mutations, two new polymorphisms would be expected to arise only once every approximately 5216 years [41]. Although this time frame is consistent with the estimated age of the ancestral Polynesian motif [46], it still far exceeds the colonization horizon of approximately 1500 years attested by archaeological and linguistic records for the settlement of Madagascar. Therefore, we set out to determine how often this unusual pattern of mtDNA diversity might occur under various demographic parametrizations of the settlement of Madagascar.

(d) African–Indonesian framework

To determine the demographic parameters of a Malagasy settlement model, we first had to formulate a basic

demographic framework relating Africans and Indonesians, the parental groups of modern Malagasy. This relationship was recently inferred using coalescent simulations, and is described in great detail by Wollstein *et al.* [37]. Rather than duplicate their efforts, we accepted this inferred demographic model with one key exception. The Wollstein model was specifically inferred for a single population sample from Borneo, whereas our Indonesian sample was drawn from across the archipelago. We modified the Wollstein model to accommodate this difference. Coalescent datasets were simulated under a range of effective population sizes for Indonesia (as opposed to just Borneo). The same number of segregating sites as observed in the real Indonesian dataset ($S = 186$) was chosen as the acceptance criterion. The MLE for the effective population size of Indonesians was 4637 (95% CI: 2435–6823), which slightly exceeds the estimate of Wollstein *et al.* for the island of Borneo (4034; 95% CI: 1993–6497). Because we are sampling a larger geographical range of genetic diversity, the effective size of Indonesians is expected to be larger than the effective size of the single Borneo population by itself. This pan-Indonesian effective size estimate was applied in all subsequent modelling.

(e) Malagasy settlement model

The African–Indonesian framework was extended to accommodate the key aspects of Malagasy settlement (figure 1). In this more complex model, the Malagasy population is formed through admixture between African and Indonesian sources. At some time ($T_{\text{settlement}}$), the Malagasy population ($N_{\text{foundation}}$) is formed from some proportion of Indonesian ($P_{\text{Indonesian}}$) and African ($1 - P_{\text{Indonesian}}$) settlers. The founding population subsequently expands until it reaches its current effective population size (N_{current}). Under this inferential model, we simulated coalescent datasets representing entire mtDNA genomes, which were accepted only if they possessed the key distinguishing characteristics of the Malagasy/Indonesian mtDNA dataset: (i) an ancestral haplotype ('Polynesian motif') in Indonesia with frequency less than or equal to 2 per cent; (ii) a derived haplotype ('Malagasy motif') in Madagascar with frequency greater than or equal to 22 per cent; (iii) ancestral and derived haplotypes that show no variation in the control region; (iv) ancestral and derived haplotypes that differ by two mutations in the coding region; and (v) 186 segregating sites in just the control region portion of the simulated Indonesian dataset.

Coalescent simulations were run on an eight-core UNIX cluster for 11 520 CPU hours (i.e. equivalent to one computer running for approx. 1.3 years). Only 671 simulations (0.0017%) were returned under all five acceptance criteria. To expand the acceptance rate, the segregating sites' criterion was extended to accommodate 5 per cent variance in the observed value ($S = 186 \pm 5\% = 177 - 195$). Under this more relaxed standard, 12 250 simulations (0.03%) were returned under all five acceptance criteria. As the two sets of results showed no substantive differences, demographic parameters are reported for the larger dataset.

Likelihood curves were plotted for all four demographic parameters in the Malagasy settlement model: foundation effective population size, current effective population size, time of settlement and proportion contribution from Indonesia (figure 2). MLEs and 95% CI are

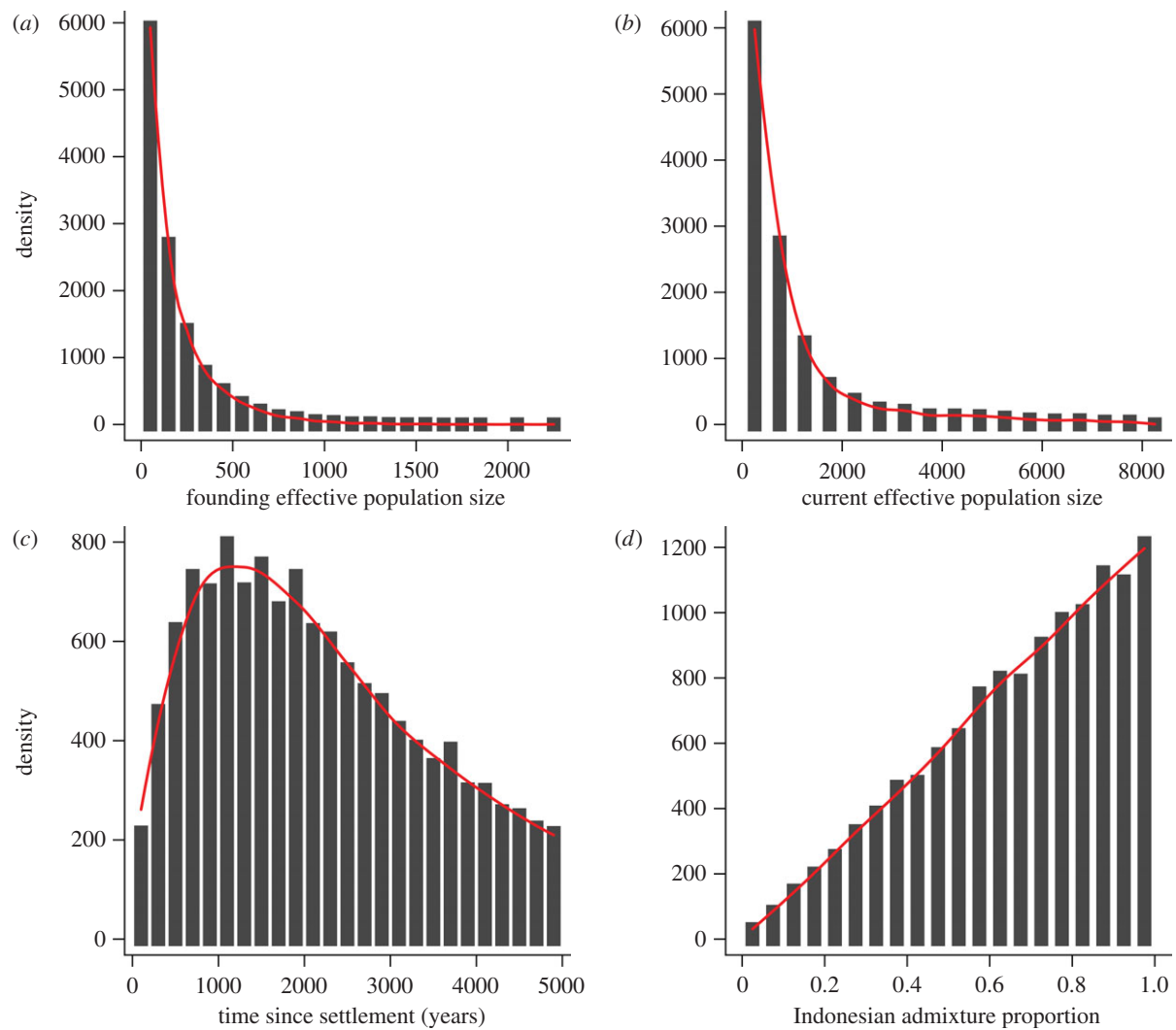


Figure 2. Likelihood curves inferred for demographic parameters in the Malagasy settlement model: (a) founding effective population size; (b) current effective population size; (c) time since settlement (years); (d) proportion of the founding population with Indonesian ancestry. Grey bars indicate the distribution of accepted Monte Carlo data points; red lines plot local regressions to the probability densities.

Table 2. Maximum-likelihood estimates (MLEs) and 95% CI of demographic parameters inferred for the Malagasy settlement model.

demographic parameter	label	MLE	95% CI
founding effective population size	$N_{\text{foundation}}$	30	3–746
current effective population size	N_{current}	193	29–5452
time since settlement (years)	$T_{\text{settlement}}$	1170	255–4721
Indonesian admixture proportion	$P_{\text{Indonesian}}$	0.93	0.16–0.99

reported in table 2. The most probable set of parameter values indicates that Madagascar was settled 1170 years ago (95% CI: 255–4721 years ago). The coalescent analysis infers that a very small number of women founded the Malagasy population—the best estimate is 30 (95% CI: 3–746). The proportion of this founding population having Indonesian ancestry was very high—0.93 (95% CI: 0.16–0.99). The Malagasy population subsequently grew approximately sixfold to reach its current effective population size (193; 95% CI: 29–5452). This effective size is very small compared with other globally distributed populations [47], but fits well with the fact that many Malagasy carry an identical mtDNA sequence (i.e. the Malagasy motif).

4. DISCUSSION

An extensive body of genetic, linguistic and anthropological research has now painted a picture of Malagasy history in broad brushstrokes. Humans reached Madagascar only relatively recently. In fact, Madagascar was among the last places on Earth to be settled—archaeological evidence suggests that the first sustained colonization did not occur until the middle of the first millennium AD. Today, Malagasy show clear evidence of genetic and linguistic contributions from Africa. Loanwords suggest contact with the Bantu-speaking peoples of east Africa [2], but the exact origin of these African settlers remains largely unknown. Contact with populations north of the Zambezi River is typically favoured [5,26]. Complementing this African connection, evidence

from multiple fields of study shows clear and substantial contributions from the Indonesian archipelago. The exact geographical provenance of this contribution is again unclear—arguments have been advanced for both west Indonesia (Ma'anyan language; Sanskrit, Malay and Javanese loanwords) and east Indonesia (South Sulawesi loanwords; prevalence of the Polynesian motif). The presence of Indonesian mtDNA and Y chromosomes in Madagascar argues for a mixed-sex founder population, and this Indonesian genetic component reaches substantial frequencies among Malagasy even today. Similarly, because African and Indonesian lineages differ little across all Malagasy ethnic groups screened to date, the main episode of African–Indonesian admixture likely occurred at the very beginning of Malagasy history.

Key detailing was recently added to this picture. The Polynesian motif in Madagascar carries two derived polymorphisms (1473 and 3423A), and this new haplotype—termed the Malagasy motif—is present in all Polynesian motif carriers screened in Madagascar so far [29]. However, the distribution of the Malagasy motif outside Madagascar has been unclear. Here, we show that this lineage is not widely distributed in Indonesia. Indeed, the Malagasy motif was not detected in 2745 individuals drawn from across the Indonesian archipelago—the largest Indonesian population dataset analysed to date. Even the Polynesian motif, the ancestral lineage of the Malagasy motif, is rare in Indonesia (2%), and with sporadic exceptions, this haplotype is restricted to eastern parts of the archipelago. It remains possible that the Malagasy motif may yet be found in a few small Indonesian communities, perhaps on the relatively understudied islands of Borneo and Sulawesi. Broader sampling in these regions will eventually answer this question. Alternately, the Malagasy motif may not be present in Indonesia at all. Perhaps it arose among the earliest Indonesian colonists to Madagascar, subsequently sweeping to high frequency either in Madagascar or somewhere along the path of their travel (e.g. putative stopping points along the east African coast). In either case, these scenarios hint at a substantial Indonesian founder event during the main settlement period in Madagascar, with strong genetic drift raising the frequency of the Malagasy motif to the elevated levels observed in Madagascar today.

To infer the parameters of this settlement model in more detail, a statistical simulation model was built around the unusual genetic patterns observed in the Malagasy/Indonesian mtDNA dataset: (i) a low frequency of the ancestral Polynesian motif lineage in Indonesia; (ii) two polymorphisms distinguishing the Polynesian and Malagasy motifs; and (iii) a high frequency of the derived Malagasy motif in Madagascar. We set out to ask: what range of settlement model parameters produce genetic patterns like the ones we observe?

Our coalescent modelling suggests several answers to this question. First, although the observed pattern of genetic diversity is relatively uncommon, it occurs more frequently under certain demographic parameters. Coalescent simulations best support settlement of Madagascar beginning around AD 830. This date is consistent with evidence from linguistics, which links the colonization of Madagascar to the expansion of Indonesian trading networks during the Srivijaya Empire [2]. Srivijaya reached its peak in the ninth century, but remained a major maritime

power in the Indian Ocean until well into the thirteenth century, consistent with the time frame suggested by our simulations. Our estimates also support a model in which Madagascar was settled by a small effective founding population—estimated at only approximately 30 women, most of whom had Indonesian ancestry (93%). Although this number of founding women might seem surprisingly small, it fits well with estimates of the small number of women (approx. 70) who founded New Zealand, another island nation settled by related Austronesian speakers at around the same time period [48]. In contrast with the simulations of Tofanelli *et al.* [5], our more complex inferential modelling shows that founder events and drift dynamics are sufficient to explain the frequencies of the Polynesian and Malagasy motifs in Indonesia and Madagascar. Tofanelli *et al.* employed a one-deme model, simulated only a small range of founding population sizes and fixed (rather than inferred) the Malagasy growth rate. Although they simulated genetic drift, it is not clear that they explicitly modelled the Malagasy founder event. In comparison, our multi-deme simulation framework is far more exact about the colonization event itself. Our highest likelihood model describes a settlement process in which relatively few women, most travelling from Indonesia, founded the Malagasy population—with a much smaller, but just as important, biological contribution from Africa.

This is the first genetic analysis where statistical bounds have been placed on the demographic parameters of Malagasy settlement. The most likely model favours a small founding population, which brings into question the broader context in which the settlement of Madagascar took place. A recurring hypothesis is that Indonesian maritime traders initially settled Madagascar, either as a single colonization event or via repeated settlement waves from the same source population (a process known to the Malagasy as *ranto*) [2,49,50]. Indeed, merchants have plied coastal Indian Ocean trade routes between east Africa and northern China at least since the Roman era [12]. However, early written records imply that these trading voyages were dominated by men; there is no mention of women on board long-distance trading vessels [13]. There is no clear evidence suggesting that Madagascar was settled in multiple waves, but because *ranto* is mentioned in traditional Malagasy narratives, this may be a profitable direction for future simulations.

An alternative hypothesis is that Madagascar was settled as a formal trading colony, perhaps under the auspices of the Srivijaya Empire (although Malagasy are not Hindu today), and possibly resembling later Arab trading centres in the region. Or perhaps Madagascar was settled as an ad hoc centre for refugees, drawn from those who lost land and power during the rapid expansion of Srivijayan influence. Such colonies would be established to be self-sufficient, and therefore, might be expected to include Indonesian women. However, there is little evidence—historical, archaeological or biological—of other Indonesian bases around the Indian Ocean, including the east African coast, although there may have been a Malay trading post in what is now Sri Lanka [50]. Founding such a centre in any official capacity on Madagascar—at the far extreme of Indonesia's trading reach—therefore seems out of keeping with contemporary Indonesian trading practices. Along related lines, historical documents leave no record of refugees fleeing the Srivijaya Empire, although early

chroniclers seldom paid attention to powerless groups as these refugees would likely have been.

A third hypothesis is that Madagascar was settled via a direct sailing route across the Indian Ocean, perhaps even as the result of an unintended transoceanic voyage. This view is traditionally considered unlikely, but has recently been revived based on seafaring simulations using ocean currents and monsoon weather patterns [51]. Indeed, during the Second World War, wreckage from ships bombed in the vicinity of Sumatra and Java later washed up in Madagascar, including—in one instance—a survivor in a lifeboat [52]. Cargo ships were substantial vessels during the first millennium AD—up to 500 tonnes and manned by over 100 sailors [53,54]. It is therefore not beyond the realms of possibility that a single wayward vessel might have effected the settlement of Madagascar. This would certainly be consistent with the extremely small initial size of the Malagasy population, although perhaps not with the mixed-sex founding group suggested by Malagasy genetics.

Without further evidence, the exact nature of Madagascar's settlement must necessarily remain unresolved. Nevertheless, we can now place the settlement process on a much firmer statistical footing. Malagasy are the children of both east and west, with clear Indonesian and African antecedents. Madagascar was settled approximately 1200 years ago, primarily by a small cohort of Indonesian women, and this Indonesian contribution—of language, culture and genes—continues to dominate the nation of Madagascar even today.

We thank Golfiani Malik, Wuryantari Setiadi and Loa Helena Suryadi of the Eijkman Institute for Molecular Biology (Jakarta, Indonesia), and J. Stephen Lansing of the University of Arizona (Tucson, AZ, USA), together with Indonesian Public Health clinic staff, for sample collection. We also thank Harilanto Razafindrakaza (University of Toulouse) for helpful discussion. Research was supported by an Early Career Medal from Massey University and a Rutherford Fellowship from the Royal Society of New Zealand (RDF-10-MAU-001) awarded to M.P.C.

REFERENCES

- Ferrand, G. 1908 L'origine africaine des Malgaches [The African origin of the Malagasy]. *J. Asiat.* **10**, 353–500.
- Adelaar, A. 2009 Loanwords in Malagasy. In *Loanwords in the world's languages: a comparative handbook* (eds M. Haspelmath & U. Tadmor), pp. 717–746. Berlin, Germany: De Gruyter Mouton.
- Hurles, M. E., Sykes, B. C., Jobling, M. A. & Forster, P. 2005 The dual origin of the Malagasy in Island Southeast Asia and East Africa: evidence from maternal and paternal lineages. *Am. J. Hum. Genet.* **76**, 894–901. (doi:10.1086/430051)
- Serva, M., Petroni, F., Volchenkov, D. & Wichmann, S. 2011 Malagasy dialects and the peopling of Madagascar. *J. R. Soc. Interface* **9**, 54–67. (doi:10.1098/rsif.2011.0228)
- Tofanelli, S., Bertoni, S., Castri, L., Luiselli, D., Calafell, F., Donati, G. & Paoli, G. 2009 On the origins and admixture of Malagasy: new evidence from high-resolution analyses of paternal and maternal lineages. *Mol. Biol. Evol.* **26**, 2109–2124. (doi:10.1093/molbev/msp120)
- Burney, D. A., Burney, L. P., Godfrey, L. R., Jungers, W. L., Goodman, S. M., Wright, H. T. & Jull, A. J. T. 2004 A chronology for late prehistoric Madagascar. *J. Hum. Evol.* **47**, 25–63. (doi:10.1016/j.jhevol.2004.05.005)
- Gommery, D., Ramanivosoa, B., Faure, M., Guérin, C., Kerloc'h, P., Sénégas, F. & Randrianantenaina, H. 2011 Les plus anciennes traces d'activités anthropiques de Madagascar sur des ossements d'hippopotames subfossiles d'Anjohibe (Province de Mahajanga) [The oldest traces of human activity on Madagascar in the subfossil bones of hippopotamus in Anjohibe (Mahajanga Province)]. *C.R. Palevol.* **10**, 271–278. (doi:10.1016/j.crpv.2011.01.006)
- Burney, D. A. 1987 Late Holocene vegetational change in central Madagascar. *Quat. Res.* **28**, 130–143. (doi:10.1016/0033-5894(87)90038-X)
- Burney, D. A. 1993 Late Holocene environmental changes in arid southwestern Madagascar. *Quat. Res.* **40**, 98–106. (doi:10.1006/qres.1993.1060)
- MacPhee, R. D. E. & Burney, D. A. 1991 Dating of modified femora of extinct dwarf Hippopotamus from southern Madagascar: implications for constraining human colonization and vertebrate extinction events. *J. Archaeol. Sci.* **18**, 695–706. (doi:10.1016/0305-4403(91)90030-S)
- Wright, H. T., Rakotoarisoa, J. A., Heurtebize, G. & Vérin, P. 1993 The evolution of settlement system in the Efaho River Valley: a preliminary report on archaeological reconnaissances of 1983–86. *Bull. Indo-Pacific Prehistory Assoc.* **12**, 2–20.
- Casson, L. 1989 *The periplus Maris Erythraei*. Princeton, NJ: Princeton University Press.
- Tibbetts, G. 1979 *A study of the Arabic texts containing material on South-East Asia*. Leiden, The Netherlands: E.J. Brill.
- Ravenstein, E. G. 1898 *The journal of the first voyage of Vasco da Gama, 1497–1499*. London, UK: Hakluyt Society.
- Mariano, L. 1904 Relation du voyage de découverte fait à l'île Saint Laurent dans les années 1613–1614 [Account of the voyage of discovery made on the island of Saint Laurent in the years 1613–1614]. In *Collection des ouvrages anciens concernant Madagascar* (eds A. Grandidier & G. Grandidier), pp. 1–64. Paris, France: Comité de Madagascar.
- Dahl, O. C. 1951 *Malgache et Maanyan: Une comparaison linguistique [Malagasy and Ma'anyan: A linguistic comparison]*. Oslo, Norway: Egede Instituttet.
- Dahl, O. C. 1977 La subdivision de la famille Barito et la place du Malgache [The subdivision of the Barito family and the place of Malagasy]. *Acta Orient.* **38**, 77–134.
- Dahl, O. C. 1991 *Migration from Kalimantan to Madagascar*. Oslo, Norway: The Institute for Comparative Research in Human Culture, Norwegian University Press.
- Adelaar, K. A. 1989 Malay influence on Malagasy: linguistic and culture-historical implications. *Ocean Linguist.* **28**, 1–46. (doi:10.2307/3622973)
- Adelaar, K. A. 1995 Asian roots of the Malagasy: a linguistic perspective. *Bijdragen tot de Taal-, Land- en Volkenkunde.* **151**, 325–356.
- Wright, H. T. & Rakotoarisoa, J.-A. 1997 Cultural transformations and their impacts on the environments of Madagascar. In *Natural change and human impact in Madagascar* (eds S. M. Goodman & B. D. Patterson), pp. 309–330. Washington DC: Smithsonian Institution Press.
- Blench, R. M. 2010 Evidence for the Austronesian voyages in the Indian Ocean. In *The global origins and development of seafaring* (eds A. J. Anderson, J. H. Barrett & K. V. Boyle), pp. 239–248. Cambridge, UK: MacDonal Institute.
- Perrier, X. et al. 2011 Multidisciplinary perspectives on banana (*Musa spp.*) domestication. *Proc. Natl Acad. Sci. USA* **108**, 11311–11318. (doi:10.1073/pnas.1102001108)
- Fourquet, R., Sarthou, J., Roux, J. & Aori, K. 1974 Hémoglobine S et origines du peuplement de Madagascar:

- nouvelle hypothèse sur son introduction en Afrique [Hemoglobin S and origins for the settlement of Madagascar: new hypothesis on its introduction to Africa]. *Arch. Inst. Pasteur Madagascar* **43**, 185–220.
- 25 Hewitt, R., Krause, A., Goldman, A., Campbell, G. & Jenkins, T. 1996 β -globin haplotype analysis suggests that a major source of Malagasy ancestry is derived from Bantu-speaking Negroids. *Am. J. Hum. Genet.* **58**, 1303–1308.
- 26 Soodyall, H., Jenkins, T., Hewitt, R., Krause, A. & Stoneking, M. 1996 The peopling of Madagascar. In *Molecular biology and human diversity* (eds A. Boyce & C. Mascie-Taylor), pp. 156–170. Cambridge, UK: Cambridge University Press.
- 27 Soodyall, H., Jenkins, T. & Stoneking, M. 1995 'Polynesian' mtDNA in the Malagasy. *Nat. Genet.* **10**, 377–378. (doi:10.1038/ng0895-377)
- 28 Cox, M. P. 2005 Indonesian mitochondrial DNA and its opposition to a Pleistocene era origin of proto-Polynesians in Island Southeast Asia. *Hum. Biol.* **77**, 179–188. (doi:10.1353/hub.2005.0037)
- 29 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**, 575–581. (doi:10.1038/ejhg.2009.222)
- 30 Msaidie, S., Ducourneau, A., Boetsch, G., Longepied, G., Papa, K., Allibert, C., Yahaya, A. A., Chiaroni, J. & Mitchell, M. J. 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**, 89–94. (doi:10.1038/ejhg.2010.128)
- 31 Cox, M. P., Karafet, T. M., Lansing, J. S., Sudoyo, H. & Hammer, M. F. 2010 Autosomal and X-linked single nucleotide polymorphisms reveal a steep Asian-Melanesian ancestry cline in eastern Indonesia and a sex bias in admixture rates. *Proc. R. Soc. B* **277**, 1589–1596. (doi:10.1098/rspb.2009.2041)
- 32 Karafet, T. M., Hallmark, B., Cox, M. P., Sudoyo, H., Downey, S. S., Lansing, J. S. & Hammer, M. F. 2010 Major east-west division underlies Y chromosome stratification across Indonesia. *Mol. Biol. Evol.* **27**, 1833–1844. (doi:10.1093/molbev/msq063)
- 33 Karafet, T. M. *et al.* 2005 Balinese Y-chromosome perspective on the peopling of Indonesia: genetic contributions from pre-Neolithic hunter-gatherers, Austronesian farmers, and Indian traders. *Hum. Biol.* **77**, 93–114. (doi:10.1353/hub.2005.0030)
- 34 Lansing, J. S. *et al.* 2007 Coevolution of languages and genes on the island of Sumba, eastern Indonesia. *Proc. Natl Acad. Sci. USA* **104**, 16 022–16 026. (doi:10.1073/pnas.0704451104)
- 35 Lansing, J. S., Watkins, J. C., Hallmark, B., Cox, M. P., Karafet, T. M., Sudoyo, H. & Hammer, M. F. 2008 Male dominance rarely skews the frequency distribution of Y chromosome haplotypes in human populations. *Proc. Natl Acad. Sci. USA* **105**, 11 645–11 650. (doi:10.1073/pnas.0710158105)
- 36 Lansing, S. J., Cox, M. P., de Vet, T. A., Downey, S. S., Hallmark, B. & Sudoyo, H. 2011 An ongoing Austronesian expansion in Island Southeast Asia. *J. Anthropol. Archaeol.* **30**, 262–272. (doi:10.1016/j.jaa.2011.06.004)
- 37 Wollstein, A., Lao, O., Becker, C., Brauer, S., Trent, R. J., Nürnberg, P., Stoneking, M. & Kayser, M. 2010 Demographic history of Oceania inferred from genome-wide data. *Curr. Biol.* **20**, 1983–1992. (doi:10.1016/j.cub.2010.10.040)
- 38 Hudson, R. R. 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics.* **18**, 337–338. (doi:10.1093/bioinformatics/18.2.337)
- 39 Sigurðardóttir, S., Helgason, A., Gulcher, J. R., Stefansson, K. & Donnelly, P. 2000 The mutation rate in the human mtDNA control region. *Am. J. Hum. Genet.* **66**, 1599–1609. (doi:10.1086/302902)
- 40 Fenner, J. N. 2005 Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423. (doi:10.1002/ajpa.20188)
- 41 Soares, P. *et al.* 2009 Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**, 740–759. (doi:10.1016/j.ajhg.2009.05.001)
- 42 Galtier, N., Nabholz, B., Glémin, S. & Hurst, G. D. D. 2009 Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* **18**, 4541–4550. (doi:10.1111/j.1365-294X.2009.04380.x)
- 43 Watterson, G. A. 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276. (doi:10.1016/0040-5809(75)90020-9)
- 44 Cox, M. P., Morales, D. A., Woerner, A. E., Sozanski, J., Wall, J. D. & Hammer, M. F. 2009 Autosomal resequencing data reveal Late Stone Age signals of population expansion in sub-Saharan African foraging and farming populations. *PLoS ONE* **4**, e6366. (doi:10.1371/journal.pone.0006366)
- 45 R Development Core Team. 2012 R: a language and environment for statistical computing. See <http://www.r-project.org>.
- 46 Soares, P. *et al.* 2011 Ancient voyaging and Polynesian origins. *Am. J. Hum. Genet.* **88**, 239–247. (doi:10.1016/j.ajhg.2011.01.009)
- 47 Cox, M. P., Woerner, A. E., Wall, J. D. & Hammer, M. F. 2008 Intergenic DNA sequences from the human X chromosome reveal high rates of global gene flow. *BMC Genet.* **9**, e76. (doi:10.1186/1471-2156-9-76)
- 48 Penny, D., Murray-McIntosh, R. & Harrison, G. L. 2002 Estimating the number of females in the founding population of New Zealand: analysis of mtDNA variation. *J. Polyn. Soc.* **111**, 207–221.
- 49 Beaujard, P. 2003 Les arrivées austronésiennes à Madagascar: vagues ou continuum? [Austronesian arrivals in Madagascar: Waves or continuum?]. *Étud. Océan Ind.* **35–36**, 59–147.
- 50 Ottino, P. 1974 Le moyen-âge de l'Océan Indien et le peuplement de Madagascar [The Indian Ocean in the Middle Ages and the settlement of Madagascar]. *Annu. Pays l'Océan Ind.* **1**, 197–221.
- 51 Fitzpatrick, S. M. & Callaghan, R. 2008 Seafaring simulations and the origin of prehistoric settlers to Madagascar. In *Islands of inquiry: colonization, seafaring and the archaeology of maritime landscapes* (eds G. Clark, F. Leach & S. O'Connor), pp. 55–66. Canberra, Australia: Australian National University Press.
- 52 Faublée, J. 1970 Les manuscrits arabico-malgaches du Sud-Est [Arab-Malagasy manuscripts of the Southeast]. *Revue Fr. d'histoire d'Outre-Mer* **57**, 268–287.
- 53 Adelaar, K. A. 2006 The Indonesian migrations to Madagascar: making sense of the multidisciplinary evidence. In *Austronesian diaspora and the ethnogenesis of people in Indonesian archipelago* (eds T. Simanjuntak, I. H. E. Pojoh & M. Hisyam), pp. 205–232. Jakarta, Indonesia: LIPI Press.
- 54 Chihara, D., Namikawa, R. & Hikata, R. 1971 *Borobudur*. Tokyo: Heibonsha.