A genetic bottleneck in the founder population may explain the enigmatic genetic replacement in Vanuatu

Yotam Ben-Oren[1], Oana Carja[2], Erella Hovers[3], Oren Kolodny[1]

[1]Department of Ecology, Evolution & Behavior, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem.
[2]Computational Biology Department, School of Computer Science, Carnegie Mellon University
[3] Institute of Archaeology, The Hebrew University of Jerusalem, Jerusalem, Israel

Abstract

The prehistory of Vanuatu presents an enigma: within several centuries from its initial settlement by Austronesians, the population went through a genetic turnover - switching from almost pure Austronesian genetic ancestry to approximately 90% Papuan ancestry – without replacement of its Austronesian languages. This purportedly unprecedented finding has been a subject of extensive debate in recent years, with no general agreement on the processes that led to it. We suggest that the existing hypotheses, suggesting either a single or multiple selectively-neutral introgression events, are insufficient to explain the turnover. As an alternative, we propose that an important contributing factor may have been a genetic bottleneck experienced by the founding Austronesian colonizers prior to the arrival of the Papuan population. Using a computational model, we demonstrate how a period of population reduction or recurring bottlenecks could have led to increased homozygosity and accumulation of deleterious mutations in the founder population. In this state, individuals joining the population from the outside, as well as admixed offspring, would have a significant fitness advantage. Therefore, even a small number of individuals joining from outside – which would have been likely to accept the predominant language and be culturally assimilated – could eventually lead to the replacement of most of the population's original genetics. We further suggest that this general scenario, in which bottlenecked populations are genetically replaced in a cryptic manner via rare migration events, may be more common than appreciated in human and non-human populations.

Introduction

The prehistory of the human population of Vanuatu has remained enigmatic despite extensive study and debate in recent years. Vanuatu, an archipelago in the South Pacific, was first colonized circa 3000 years BP by people associated with the Lapita culture (Bellwood, 1991). Ancient DNA and archeological data point to an Austronesian descent of these colonizers (Skoglund et al., 2016; Lipson et al., 2018; Posth et al., 2018). Austronesian people originated in Southeast Asia, presumably in Taiwan, between 6000 and 4500 years BP, and spread throughout the Pacific (Bellwood, 1991). They migrated from island to island, using outrigger canoes, ultimately colonizing regions as far as Hawaii and Madagascar. While some of the islands through which Lapita spread were already inhabited by other peoples, genetic data from the first inhabitants of Vanuatu shows that they were almost entirely of East Asian descent. However, ancient DNA samples show a drastic change in the genetic composition of Vanuatu 500-1000 years after the initial colonization, leading to the present, with modern populations composed of as little as 10% East-Asian ancestry and 90% Papuan ancestry (Skoglund et al., 2016; Lipson et al., 2018; Posth et al., 2018). Even more surprisingly, and possibly unprecedentedly (Posth et al., 2018), although the genetic signature of the Austronesians was almost erased, their languages were conserved in modern populations, and all indigenous Vanuatuan languages belong to the Austronesian clade of languages (François et al., 2015).

One generally expects to find a correlation between a population's genetics and language (Barbujani & Sokal, 1990; Cavalli-Sforza et al., 1988; Chen et al., 1995; Nettle & Harriss, 2003; Sokal, 1988). For example, this was found to be the case in the Indonesian island Sumba, which also experienced admixture between Papuans and Austronesians (although in the opposite order, with the first colonizers being Papuan, followed by Austronesians). Lansing et al. 2007 (Lansing et al., 2007) showed that the frequency of Austronesian vocabulary in different languages throughout Sumba correlates with the frequency of Austronesian Y lineages. In other cases, however, such correlation is not found (e.g., Rosser et al. 2000). Cases of mismatch between language and genetic ancestry are usually thought to be a result of cultural change in which a population largely retained its genetic makeup and adopted a new language (Renfew, 1987). Instances where a population was genetically replaced but the

originally occurring language was retained, as in Vanuatu, are virtually unknown (Posth et al., 2018) and present a puzzle: which dynamics could have led to this phenomenon?

To explain the genetic replacement in Vanuatu, researchers offered two competing hypotheses. The first hypothesis, hereafter referred to as the "single wave hypothesis", is that 500-1000 years after the arrival of the Lapita people to Vanuatu, a second significant wave of Papuan migrants arrived and replaced the population's genetic composition (Lipson et al. 2018, 2020). The second hypothesis, hereafter referred to as the "multiple wave hypothesis", is that the genetic replacement was gradual and resulted from a series of migration events over an extended period (Posth et al., 2018), a process in which the Austronesian genetics could be gradually diluted, with each wave being small enough to be linguistically assimilated.

Several putative lines of evidence support each of these hypotheses. In support of the single wave hypothesis, it has been argued that the genetic composition of present-day people indicates a limited time frame of introgression and a specific source for Papuan genetic ancestry (Lipson et al., 2020). Others, in contrast – supporting the multiple wave hypothesis - interpret ancient DNA findings from approximately 2500 years BP as supporting a gradual reduction in the Austronesian ancestry (Posth et al., 2018), a trend that seems inconsistent with a single migration event. Additionally, they assert that there is no evidence in the archeological record to support a significant migration event (Posth et al., 2019). In further support of the multiple waves hypothesis, it has been asserted that when a population is replaced by a wave of migrants, language is usually replaced as well (Posth et al., 2018).

We begin by demonstrating analytically that the observed turnover is unlikely to have resulted from regular, selectively-neutral admixture. We show that under neutral conditions, neither one wave of migration nor multiple waves is expected to have led to the observed genetic turnover, as both scenarios would require an unrealistically large number of Papuan migrants. Next, using a computational model, we simulate the prehistory of Vanuatu: the demographic bottleneck experienced by the Lapita people, its genetic consequences in terms of inbreeding and accumulation of deleterious alleles, and the possible influence of a small number of Papuan migrants

on the genetic composition of the population when considering fitness differentials among individuals in the population, which were admixed to different extents.

Genetic bottlenecks can have three negative consequences on population fitness (Templeton, 2021). First, they reduce heterozygosity, making deleterious recessive mutations more likely to be expressed as inbreeding depression (Charlesworth & Charlesworth, 1999; Stebbins, Ledyard Stebbins, and Wright, 1977; Peischl & Excoffier, 2015). Second, when populations are small, genetic drift is stronger, making selection less efficient; this increases the probability of fixation of deleterious alleles (M. Lynch, Conery, and Burger 1995; Peischl, Kirkpatrick, and Excoffier 2015). Finally, bottlenecks reduce populations' genetic diversity, limiting their ability to adapt to environmental challenges (James, 1970).

Considering two negative consequences of genetic bottlenecks—loss of heterozygosity and accumulation of deleterious alleles—we demonstrate how the arrival of even a small number of outbred immigrants could have led to the almost complete replacement of the Austronesian component in the population's genetics.

Importantly, while Vanuatu's prehistory provides an interesting test case for exploring such dynamics, we view it as merely an example of a potentially much broader ecological phenomenon, in which large bottlenecked populations can be cryptically replaced by a small number of outbred migrants. Thus, this work does not aim to prove that the genetic turnover in Vanuatu was a product of the bottleneck experienced by the Lapita, but rather to use this debated historical case as an example of the proposed phenomenon of cryptic genetic replacement in bottlenecked populations.

Results

In what follows, we explore how different demographic and genetic processes are expected to have influenced the population's genetic composition. We begin by modeling the selectively-neutral dynamics of population admixture that have so far been proposed in the literature. We show that the number of migrants required to dilute the Austronesian ancestry to the observed levels (from ~100% to ~10%) under neutral conditions is unrealistically large. Next, we present a computational model that simulates the three phases of Vanuatu's prehistory: the genetic bottleneck during the Austronesian spread, the establishment of the Austronesian population in Vanuatu, and the arrival of Papuans. Using this model, we examine the influence of two effects of genetic bottlenecks – the loss of heterozygosity and the accumulation of deleterious mutations – coupled with an exploration of the impact of the population size and its variance during the bottleneck phase on the admixture and the ensuing dynamics.

Null model for the relationship between the number of migrants and the observed admixture

We create a simple mathematical model to assess the expected levels of admixture for different demographic scenarios, under neutral conditions (i.e., without selection in favor of either group). The expected genetic composition of the present population is calculated as the average of the genetic compositions of its ancestors at the admixture event. For simplicity, our model does not consider the effect of genetic drift on the ancestry composition (which *is* included in the following computer simulations), nor does it explicitly include the interactions between the populations (e.g., wars, assimilation etc.). Hence, for example, if a population of 100 individuals whose ancestry was 50% Austronesian and 50% Papuan, was joined by a group of 100 'pure' Papuans, we estimate the expected ancestry composition of their descendants is expected to be 25% Austronesian and 75% Papuan. The model assumes that by the arrival of the first Papuans, the Austronesian population size had already reached its carrying capacity. This assumption is supported by empirical data showing that the Lapita population had grown significantly and spread throughout the archipelago long before the arrival of Papuans (Spriggs et al., 2019) and by the observation that island

populations often reach their carrying capacity shortly after colonization (Armstrong et al., 2005). For multiple migration events, we assume that the population returned to its carrying capacity between events (as by definition, even after a large number of migrants join a population, it cannot sustain more individuals than its carrying capacity for extended periods of time), and that the proportions of different ancestries were maintained.

We begin by examining the simpler scenario, in which the genetic replacement is assumed to be a product of a single migration event. If we assume a single admixture event between two "pure" populations, the observed fraction of Austronesian ancestry after the introgression, A, should reflect the proportions of individuals from each population at the time of introgression. Because we assume that the Lapita population reached carrying capacity, K, before the introgression event, for a given number of Papuan migrants the expected fraction of Austronesian ancestry can be calculated as follows:

$$A = \frac{K}{M + K}$$

Meaning that, for a single introgression event, the fraction of Austronesian ancestry in the admixed population is expected to be the same as the fraction of Austronesian ancestors at the introgression event.

Thus, to reach the observed admixture composition (approx. 10% Austronesian and 90% Papuan), the Papuan wave of migration had to be ~9 times larger than the size of the preexisting Austronesian population. Hence, if, for example, the Lapita population had reached a carrying capacity of 50,000 individuals before the arrival of Papuans, a single migration wave would need to be as large as 450,000 in order to "dilute" the Austronesian ancestry to the observed levels, under neutral conditions.

One might ask whether a gradual but steady trickle of migrants, or several waves of migration, might require less migrating individuals in order to reach 90% Papuan ancestry in the population. Let us take, once again, 50,000 individuals as the carrying capacity, and explore several such scenarios. If, for example, we introduce 50,000 migrants in 1 wave of migration, the expected proportion of Austronesian ancestry in

the descendants would be 50%. However, if we divide this wave into 10 smaller waves of 5000 migrants, we can get different results. After the arrival of the first wave, the population will be 90% Austronesian (50,000/55,000, or 9/10 of pure Austronesian ancestry and 1/1 of pure Papuan ancestry). After the second wave, it will be 81% Austronesian (9/10 of 9/10 Austronesian ancestry and 1/10 of pure Papuan ancestry), and by the 10th wave, the population will become only 34.9% Austronesian. This process can be formulated as follows:

$$A = \left(\frac{K}{\frac{M}{W} + K} \right)^W$$

Where M signifies the **total** number of Papuan migrants and W signifies the number of migration waves. As demonstrated above, the more migration waves ("W"), the smaller the number of migrants ("M") it would take to dilute the Austronesian ancestry. If we consider the extreme case, in which a migration event happened every single generation, W will be ~40 (assuming an introgression period of 1000 years and a generation time of 25 years). In this scenario, the number of migrants would have to be 2.37 times bigger than the habitat's carrying capacity. Thus, even under a conservative scenario, the number of Papuan immigrants required to dilute the population's Austronesian component to 10% is very large, much larger than Vanuatu's overall carrying capacity. Going back to the example, or a carrying capacity of 50,000, the number of migrants required would be 118,500.

We, therefore, suggest that the genetic replacement of the Austronesian ancestry is unlikely to have been a result of only neutral processes. Instead, we propose that natural selection offers a reasonable explanation for this replacement.

The effects of loss of heterozygosity and inbreeding depression on genetic admixture

Austronesians are known to have spread throughout the Pacific relatively quickly ("the express train sequence", (Bellwood, 1991) and in small groups, leading to a series of genetic bottlenecks (Duggan et al., 2014; Tätte et al., 2022). These bottlenecks, as discussed above, were likely to have caused inbreeding depression and accumulation

of deleterious mutations. We set up a simple computational framework to explore the separate effects of inbreeding depression and mutational load (Templeton, 2021), and their combination. We simulated the three stages of the human colonization of Vanuatu – (1) A 60-generation (~1500 years) genetic bottleneck experienced by the Austronesian founder population on its way to Vanuatu, (2) A 40-generation (~1000 years) period where the population established in Vanuatu in isolation and (3) a 1000-generation (~2500 years) period following the arrival of Papuans.

We modeled two effects of genetic bottlenecks on fitness – Loss of heterozygosity and accumulation of deleterious mutations. The effects were examined both separately and simultaneously. Each individual in the model was assigned a diploid genome, consisting of 100 loci, with each locus leading to reduction in fitness (i.e., the probability to be chosen for reproduction) if it was in a homozygote form, or if it accumulated deleterious mutations. At the beginning of each simulation, populations were generated with maximal genetic variation (i.e., with each gamete in the population being unique) and without deleterious load. Over generations, new deleterious mutations were accumulated by each offspring, drawing from a Poisson distribution with a default Λ 2.2 per individual per generation (see Keightley, 2012), and homozygosity increased when individuals inherited the same allele from both parents (descending, ultimately, from the same ancestor).

We explore this dynamic for a default set of parameters (see Figure 1); carrying capacity was set to 50,000 (Notably, while this was an arbitrary choice, we found carrying capacity had little effect on the results, see supplementary figure S1). Austronesian and Papuan waves of migration were set to 50 individuals, representing 0.1% of the carrying capacity, and in accordance with migration in outrigger canoes, which have limited capacity (e.g., see Sheppard, 2019, who estimated the Lapita founder population at Sumba to be a few dozen). Conservatively, by default, we examined the effect of a single Papuan migration wave. This is because multiple migration waves would only increase the Papuan ancestry fraction, meaning that if we find that one migration wave would be enough to replace the population's genetics, then multiple waves would be as well (see previous section). For further information

about the model and the choices of parameter values, see Methods. Examples of broader parameter ranges are found below and in the Supplementary Materials.

We begin by showing that under the null hypothesis (i.e., when the proportion of Papuan ancestry is subject only to neutral drift) a genetic replacement by a small number of migrants is highly unlikely, even considering genetic drift. As shown in the upper-left panel of Figure 1, and as expected from our estimations in the previous section, under neutral conditions, the fraction of Papuan ancestry in the admixed population is expected to roughly reflect the fraction of Papuan individuals at the introgression event (i.e., 0.1% in our model). Depending on population size and the number of generations after admixture, drift is even likely to remove the Papuan component over time. Next, we examine the effects of inbreeding depression or mutational load on the expected ancestry composition, under the same demographic scenario. We simulate the expected ancestry composition for different fitness costs for single-locus homozygosity (between 0 and 2.5%) and for deleterious alleles (between 0 and 2.5%). We find that each effect can, in itself, reduce Austronesian ancestry significantly (>50%) and that when the effects are combined, their impact is additive. Notably, and as expected by theory, we found (see supplementary figure S3) that in a population with two alleles (e.g., "Austronesian" and "Papuan"), selection against homozygotes would lead to allele frequencies of 50% each, regardless of their initial frequency. This means that even regardless of the effect of deleterious load, or the differences in genetic diversity between the populations, selection for heterozygosity could lead to a substantial increase in Papuan ancestry.
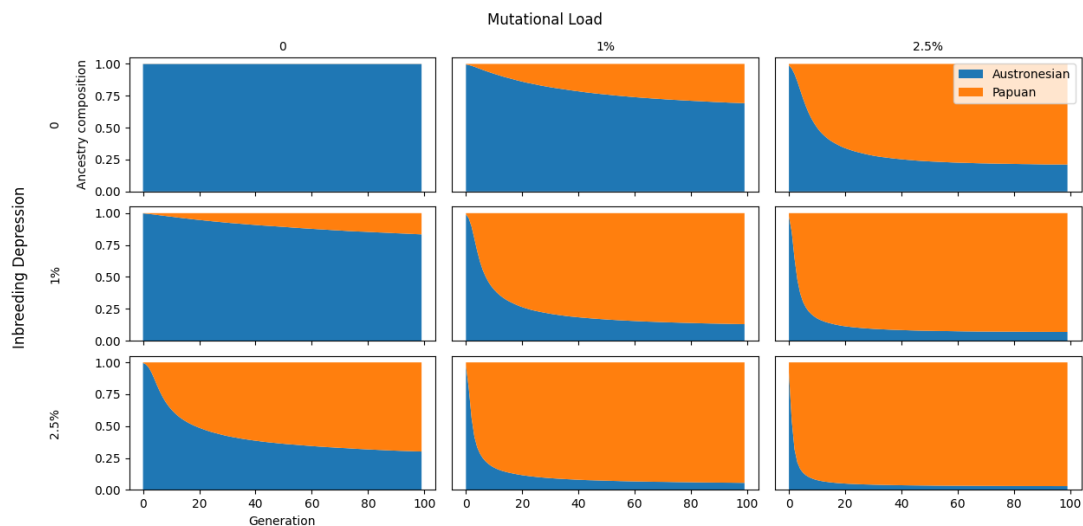
Figure 1 The effects of inbreeding depression and mutational load on genetic replacement. The graphs represent the proportion of Austronesian ancestry in a population of 50,000 over 100 generations (~2500 yrs), starting from 99.9%, following the introduction of 50 Papuan immigrants in a single wave. Each graph is based on a single simulation run. Rows represent different inbreeding depression values (from no reduction in fitness to a reduction of 2.5% per homozygotic locus), and columns represent different mutational load values (again, from no reduction in fitness to a reduction of 2.5% per homozygotic locus).

The effects of size and variance of the bottleneck on genetic admixture

To calculate the extent of inbreeding depression and mutation load, we simulated the Austronesian wave that originally colonized Vanuatu as having stayed as small as 50 individuals for 60 generations. This is, of course, a simplistic assumption. In reality, population size during the spread to Vanuatu would have probably grown at times, upon arriving on new islands and before expanding further from them. On the other hand, it would have reduced in size substantially at each subsequent migration step to the next island and during extreme environmental events throughout this process. To explore the consequences of a more complex scenario, we simulated a larger, more variable Austronesian wave of migration across the Pacific, with different average sizes (ranging from 50 to 400) and standard deviations (ranging from 25 to 75 percent of the average size). Conservatively, the second, Papuan, wave was kept at 50. We find (see Figure 2) that even for Austronesian migration waves as large as 400 individuals on average, a genetic replacement by the Papuan ancestry can occur, stemming from this single 50-individual Papuan migration wave. We find that the replacement rate is more dependent on extreme events that cause severe bottlenecks along the Austronesian expansion than it is on the mean size of this population throughout its spread phase (see in Figure 2, how graphs are more similar across columns, representing the variation in population size for different average sizes, than across rows, representing the same average sizes with different variation). This result is

coherent with the theoretical expectation in population genetics, that effective population size, both in terms of drift (predicting the accumulation of deleterious load) and in terms of accumulation of inbreeding (predicting the loss of heterozygosity), will be affected more by extreme events than by average population sizes (Templeton, 2021).
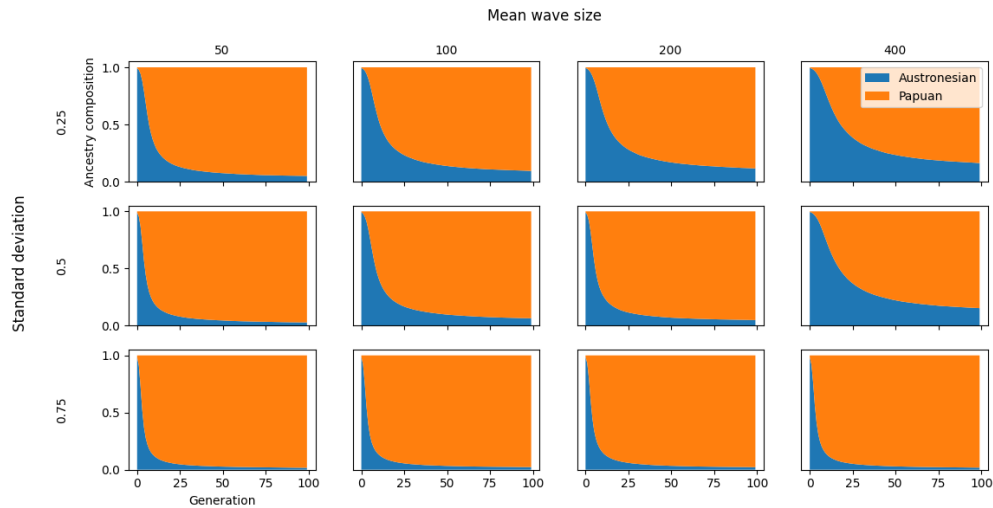
Figure 2 The effects of average size and standard deviation (as a proportion of the average) of the population size in the bottleneck period. The shown simulations were otherwise conducted with default parameters. We find that demographic fluctuations in the founding population promotes genetic turnover even if average population size is large.

In our simulations, we set the carrying capacity of the archipelago at a default value of 50,000. However, as mentioned above, to demonstrate that carrying capacity does not qualitatively change the results, we ran a control simulation with a carrying capacity of 500,000, which yielded similar results (see supplementary figure S1)

Discussion

The genetic turnover in Vanuatu has been subject to extensive debate in recent years (Skoglund et al., 2016; Lipson et al., 2018, 2020; Spriggs et al., 2019; Posth et al., 2018; Posth et al., 2019), with some evidence seeming to support a limited timeframe of admixture and others to support gradual dilution by multiple waves of migration (see Chart 1). Several important lines of evidence have been put forth in support of each of the two hypotheses regarding genetic turnover. However, we suggest that both suffer a major drawback of being incompatible with a selectively neutral scenario, as has been implicitly assumed, leaving the puzzle of population replacement unresolved. We propose that considering the likely impact of population bottlenecks, which occurred during the settlement of Remote Oceania, offers a solution to this puzzle, reconciling seemingly contradictory lines of evidence that have been laid out. Our results suggest that even a small group of migrants, which does not have a substantial cultural impact nor leaves a detectable archeological record, may cryptically replace the genetics of a population, and that even a single migration event (let alone multiple events) can trigger a gradual genetic replacement that may span dozens of generations, in line with the gradual replacement reported in (Posth et al., 2018).

	"One wave" hypothesis	"Multiple waves" hypothesis	"One wave" + genetic bottleneck	"Multiple waves" + genetic bottleneck
Genetic data suggesting a specific genetic source for Papuan admixture (Lipson et al., 2020)	Supported	Contradicting	Supported	Contradicting
Genetic data suggesting a specific timeframe of admixture (Lipson et al., 2020)	Supported	Contradicting	Supported	Contradicting
Preservation of the Austronesian languages (François et al., 2015; J. Lynch, Crowley, and Others 2001; Posth et al., 2018)	Contradicting	Supported	Supported	Supported
Lack of Archeological evidence for a big migration wave (Posth et al., 2019)	Contradicting	Supported	Supported	Supported
Sporadic genetic evidence for gradual decay in Austronesian ancestry (Posth et al., 2018)	Contradicting	Supported	Supported	Supported
Neutral dynamics require an unrealistically large number of migrants to dilute the Austronesian ancestry (see Results)	Contradicting	Contradicting	Supported	Supported

Chart 1 The existing data does not seem to align with either the "one wave" hypothesis or the "multiple waves" hypothesis. When a genetic bottleneck experienced by the Lapita population is considered, the existing data align with the one-wave hypothesis. Notably, while in our model even a single small

migration wave can lead to genetic turnover, multiple waves would enhance the replacement. Therefore, our model in itself does not rule out either of the hypotheses.

We start our exploration by suggesting that the genetic replacement in Vanuatu is unlikely to have resulted from neutral dynamics, as those would require an extremely large number of Papuan migrants. We argue that the arrival of so many Papuan migrants is unlikely, given their mode of transportation, and that if such a large number of migrants did arrive, it would likely have resulted in a cultural replacement. Several non-neutral scenarios come to mind. For example, Papuans could have had cultural or genetic adaptations that assisted the spread of their genetic ancestry. Alternatively, there could be sexual biases in reproduction, allowing few Papuan males to leave a substantial genetic mark. However, these explanations have their own limitations. For example, assuming the Papuans had beneficial cultural adaptations, these could either be copied by the others, which would stop the spread of Papuan ancestry, or stay strongly associated with a "pure" Papuan population, likely leading to language replacement. Similarly, if, for example, following a single Papuan wave of migration, reproduction was extremely biased, and women reproduced exclusively with Papuans, a single wave would lead to ~50% Papuan ancestry in the offspring. To get to the observed ancestry composition, a series of such events is required. Even assuming such a series of events could happen, such bias should be explained. We suggest that the historical and genetic evidence for bottlenecks experienced by Austronesians during their spread to Remote Oceania (Duggan et al. 2014) make it a more parsimonious explanation.

We continue by simulating the effects of inbreeding depression and mutational load on the reduction in Austronesian ancestry, showing that both factors can contribute to it, and that their combined effect can be additive. Further, we found that even if both populations are equally bottlenecked, selection in favor of heterozygosity can in itself lead to an increase in Papuan ancestry up to 0.5, regardless of its initial frequency (see supplementary Figure S3).

One could ask, if such genetic bottleneck was indeed experienced by the Lapita, how could they colonize Vanuatu in the first place? Interestingly, there are numerous cases where populations undergo a genetic bottleneck and seem to survive and even thrive.

For example, the channel island fox (*Urocyon littoralis*), a dwarfed species derived from the gray fox (*Urocyon cinereoargenteus*), has survived in isolation in the California Channel Islands for thousands of years in extremely small numbers, while reaching exceptionally low genetic diversity and accumulating deleterious mutations (Robinson et al. 2016). Similarly, invasive populations are commonly a product of a small number of individuals that establish in a novel habitat (e.g., famously, Froggat 1936). Such observations demonstrate that inbred populations can survive for extended periods as long as the ecological niche is vacant, i.e. as long as they don't face significant competition for resources from other species or from less inbred groups of their own species. When non-bottlenecked conspecifics do arrive, they may take over the population genetically. In the case of the Isle Royale wolves, an isolated population in Lake Superior, USA, a single individual from a non-bottlenecked lineage that migrated to the island using an ice bridge in 1997, was an ancestor of 59.4% of the individuals just 2.5 generations later (Hedrick et al. 2014).

Naturally, our model includes a number of simplifications regarding the demography and structure of the populations. For example, for simplicity we set the population size during the bottleneck phase to be 25 individuals. However, we show that similar results are reached for much higher mean population sizes, if the variance in population size is large (meaning, if there are only rare events in which the population size is small). Furthermore, Individuals were modelled as monoecious and the population was panmictic, meaning effects like reproduction skew were not considered. These phenomena should be further explored; we suggest they may further reinforce the replacement dynamics, making our model conservative. For example, reproductive biases during the expansion phase could further reduce the Austronesian population's effective population size compared to the modeled population size, increasing the reduction in fitness and respectively the likelihood that a small non-bottlenecked group of Papuans could lead to large-scale genetic replacement. Moreover, after the arrival of Papuans, assortative mating could help their population grow before admixing with the Austronesians, and a sexual skew could allow even fewer Papuan males to leave a substantial mark on the population's genetics.

Furthermore, while our model does not simulate any additional gene-flow from external populations other than the migration event(s), previous studies have shown that even when under gene-flow expansion load can be accumulated making genetic replacement possible (Peischl, Kirkpatrick, and Excoffier, 2015).

Future accumulation of genetic and skeletal data may help test our hypothesis. We would expect, for example, to find evidence for low heterozygosity in the Lapita people and an increase in heterozygosity after the introduction of Papuans. We might further expect to find genetic or morphological examples of genetic diseases that would have been alleviated after the admixture process.

Human prehistory is studied much more extensively than the evolutionary history of other species and can therefore shed light on ecological processes that can otherwise remain unrecognized. Specifically, in the case of Vanuatu, the combination of genetic, linguistic, and archeological data over an extended time frame can provide evidence for a cryptic genetic replacement, such as may have gone unnoticed in other prehistoric human populations and in other species. In other words, a population, even one that is tracked continuously, might seem stable and viable while in reality, the local genetic diversity goes extinct and is replaced by a different within-species genotypic composition.

There may be different ecological scenarios where cryptic replacement may occur:

1. Similarly to the case of Vanuatu, in expanding populations, where the invasion front constantly accumulates expansion load (Peischl & Excoffier, 2015; Peischl, Kirkpatrick, and Excoffier, 2015), slower-moving migration waves that undergo less extreme genetic bottlenecks can replace the genetics of the first colonizers.

2. In small isolated populations that may seem stable, genetic load could actually accumulate over time, allowing rare migrants to occasionally replace their genetics in a cryptic manner (i.e., without significant changes in demography, evidence of conflict, etc.)

3. In Human-facilitated cases of genetic rescue, driven by facilitated migration of individuals between populations (Adams et al., 2011; Johnson et al., 2010;

Whiteley et al., 2015; Harris, Zhang, and Nielsen, 2019; Hedrick et al., 2014), while these can increase the population fitness, they may also erase much of the local diversity it is aimed to preserve.

Methods

We create an agent-based simulation (using Python 3.8) to replicate three stages of the human colonization of Vanuatu – (1) The progress of Austronesians from Southeast Asia to Remote Oceania, (2) The time they spent in Vanuatu before the arrival of Papuans and (3) The arrival of Papuans and the genetic trends following it.

Individuals were simulated as diploid and monoecious. The first individuals of each lineage were created with maximal heterogeneity (meaning that each allele was unique) and with no mutational load. However, because selection was soft (meaning that fitness was only relative to other individuals in the population), these conditions are computationally equivalent to assuming the same non-zero rate of heterozygosity and mutation load for all new individuals. Generations were discrete and non-overlapping. Reproduction was simulated by randomly drawing mating pairs from the population according to their relative fitness – meaning that for N offspring N pairs are drawn, with replacement.

We modeled two effects of genetic bottlenecks on fitness – Loss of heterozygosity and accumulation of deleterious mutations. The effects were examined individually to demonstrate their possible contribution to genetic replacement and together to explore their potential extent. Both Austronesian and Papuan populations start the simulation with maximal heterozygosity (with each allele in the population being different) and without accumulated mutations. Each individual's genome consisted of 100 loci. In each locus, individuals had 2 alleles, which could be homozygous (=descending from the same ancestral allele) or heterozygous, with each homozygous locus leading to a default reduction of 1% in fitness. In addition, over generations, alleles could accumulate additional deleterious mutations, with each mutation leading to a reduction of 1% in fitness. Λ of deleterious mutations per individual per generation was defined by default as 2.2, a value based on assessment by Keightley 2012 (Keightley, 2012). Fitness was calculated in a multiplicative manner, meaning that, for example, an individual with two homozygous alleles and one accumulated deleterious mutation would have a fitness of 0.99·0.99·0.99 = 0.9703. Each generation, N new individuals were generated by drawing two random parents from the population in proportion to their fitness. Since each parent has two alleles in each

locus, in the first locus, one allele was randomly chosen to be inherited to the offspring. Then, for each consecutive locus, in 99% od cases, the following linked allele was inherited, and in 1% of cases, recombination occurred and the *trans* allele was inherited. Therefore, for a genome of 100 loci, on average, a single recombination event occurred.

During the first stage of the simulation (the migration of Austronesians from Southeast Asia to Vanuatu), a migration wave is simulated for 60 generations (equivalent to approximately 1500 years). The default population size in the bottleneck period was 25 individuals, but other population sizes (including variable sizes) were examined as well. This assumption seems reasonable, knowing that Austronesians migrated for generations from island to island using outrigger canoes with limited capacity (Bellwood 1991). Later, we simulated the colonization of Vanuatu by the Lapita people. The population was allowed to grow to its carrying capacity (50,000 individuals) in that size for additional 40 generations (equivalent to ~1000 years of Austronesian domination). The carrying capacity was chosen for computational time complexity reasons, but a carrying capacity of 500,000 was also simulated, using the default parameters, and produced similar results (see supplementary figure S1). For simplicity and conservatism, we assumed the population achieved carrying capacity within 1 generation. Slower, logistical growth was examined as well and yielded qualitatively similar results (if anything, slow growth is expected to prolong the bottleneck effects and promote genetic replacement). Afterward, by default, a second wave of 50 Papuan individuals (i.e., 1% of the recipient population) was introduced into the population, and the simulation was run for an additional 100 generations. Different, less conservative scenarios with multiple waves of migration were also explored and yielded similar results.

Bibliography

1. Adams, Jennifer R., Leah M. Vucetich, Philip W. Hedrick, Rolf O. Peterson, and John A. Vucetich. 2011. "Genomic Sweep and Potential Genetic Rescue during Limiting Environmental Conditions in an Isolated Wolf Population." *Proceedings. Biological Sciences / The Royal Society* 278 (1723): 3336–44.
2. Armstrong, Doug P., R. Scott Davidson, John K. Perrott, Jon Roygard, and Len Buchanan. 2005. "Density-Dependent Population Growth in a Reintroduced Population of North Island Saddlebacks." *Journal of Animal Ecology*. https://doi.org/10.1111/j.1365-2656.2004.00908.x.
3. Bellwood, Peter. 1991. "The Austronesian Dispersal and the Origin of Languages." *Scientific American* 265 (1): 88–93.
4. Charlesworth, B., and D. Charlesworth. 1999. "The Genetic Basis of Inbreeding Depression." *Genetical Research* 74 (3): 329–40.
5. Duggan, Ana T., Bethwyn Evans, Françoise R. Friedlaender, Jonathan S. Friedlaender, George Koki, D. Andrew Merriwether, Manfred Kayser, and Mark Stoneking. 2014. "Maternal History of Oceania from Complete mtDNA Genomes: Contrasting Ancient Diversity with Recent Homogenization due to the Austronesian Expansion." *American Journal of Human Genetics* 94 (5): 721–33.
6. François, Alexandre, Sebastien Lacrampe, Michael Franjieh, and Stefan Schnell. 2015. "The Languages of Vanuatu: Unity and Diversity." In *The Australian National University*. Asia-Pacific Linguistics, School of Culture, History and Language, College of Asia and the Pacific, The Australian National University.
7. Froggatt, W.W., 1936. "The introduction of the great Mexican toad Bufo marinus into Australia." *Australian Naturalist*, *9*, pp.163-164.
8. Harris, Kelley, Yulin Zhang, and Rasmus Nielsen. 2019. "Genetic Rescue and the Maintenance of Native Ancestry." *Conservation Genetics* 20: 59–64.
9. Hedrick, Philip W., Rolf O. Peterson, Leah M. Vucetich, Jennifer R. Adams, and John A. Vucetich. 2014. "Genetic Rescue in Isle Royale Wolves: Genetic Analysis and the Collapse of the Population." *Conservation Genetics* 15 (5): 1111–21.
10. James, J. W. 1970. "The Founder Effect and Response to Artificial Selection." *Genetical Research* 16 (3): 241–50.
11. Johnson, Warren E., David P. Onorato, Melody E. Roelke, E. Darrell Land, Mark Cunningham, Robert C. Belden, Roy McBride, et al. 2010. "Genetic Restoration of the Florida Panther." *Science* 329 (5999): 1641–45.
12. Keightley, Peter D. 2012. "Rates and Fitness Consequences of New Mutations in Humans." *Genetics* 190 (2): 295–304.
13. Lansing, J. Stephen, Murray P. Cox, Sean S. Downey, Brandon M. Gabler, Brian Hallmark, Tatiana M. Karafet, Peter Norquest, et al. 2007. "Coevolution of Languages and Genes on the Island of Sumba, Eastern Indonesia." *Proceedings of the National Academy of Sciences of the United States of America* 104 (41):

16022–26.

14. Lipson, Mark, Pontus Skoglund, Matthew Spriggs, Frederique Valentin, Stuart Bedford, Richard Shing, Hallie Buckley, et al. 2018. "Population Turnover in Remote Oceania Shortly after Initial Settlement." *Current Biology: CB* 28 (7): 1157–65.e7.

15. Lipson, Mark, Matthew Spriggs, Frederique Valentin, Stuart Bedford, Richard Shing, Wanda Zinger, Hallie Buckley, et al. 2020. "Three Phases of Ancient Migration Shaped the Ancestry of Human Populations in Vanuatu." *Current Biology: CB* 30 (24): 4846–56.e6.

16. Lynch, John, Terry Crowley, and Others. 2001. *Languages of Vanuatu: A New Survey and Bibliography*. Pacific Linguistics, Research School of Pacific and Asian Studies, The ….

17. Lynch, Michael, John Conery, and Reinhard Burger. 1995. "Mutation Accumulation and the Extinction of Small Populations." *The American Naturalist* 146 (4): 489–518.

18. Peischl, Stephan, and Laurent Excoffier. 2015. "Expansion Load: Recessive Mutations and the Role of Standing Genetic Variation." *Molecular Ecology* 24 (9): 2084–94.

19. Peischl, Stephan, Mark Kirkpatrick, and Laurent Excoffier. 2015. "Expansion Load and the Evolutionary Dynamics of a Species Range." *The American Naturalist* 185 (4): E81–93.

20. Posth, C., K. Nägele, H. Colleran, and F. Valentin. 2019. "Response to 'Ancient DNA and Its Contribution to Understanding the Human History of the Pacific Islands'(Bedford et Al. 2018)." *Archaeology in*. https://pure.mpg.de/rest/items/item_3255221/component/file_3255224/content.

21. Posth, Cosimo, Kathrin Nägele, Heidi Colleran, Frédérique Valentin, Stuart Bedford, Kaitip W. Kami, Richard Shing, et al. 2018. "Language Continuity despite Population Replacement in Remote Oceania." *Nature Ecology & Evolution* 2 (4): 731–40.

22. Robinson, Jacqueline A., Diego Ortega-Del Vecchyo, Zhenxin Fan, Bernard Y. Kim, Bridgett M. vonHoldt, Clare D. Marsden, Kirk E. Lohmueller, and Robert K. Wayne. 2016. "Genomic Flatlining in the Endangered Island Fox." *Current Biology: CB* 26 (9): 1183–89.

23. Rosser, Z. H., T. Zerjal, M. E. Hurles, M. Adojaan, D. Alavantic, A. Amorim, W. Amos, et al. 2000. "Y-Chromosomal Diversity in Europe Is Clinal and Influenced Primarily by Geography, rather than by Language." *American Journal of Human Genetics* 67 (6): 1526–43.

24. Sheppard, Peter. 2019 "Early Lapita colonisation of Remote Oceania: An update on the leapfrog hypothesis." *Debating Lapita: Distribution, chronology, society and subsistence* 52 : 135.

25. Skoglund, Pontus, Cosimo Posth, Kendra Sirak, Matthew Spriggs, Frederique Valentin, Stuart Bedford, Geoffrey R. Clark, et al. 2016. "Genomic Insights into the Peopling of the Southwest Pacific." *Nature* 538 (7626): 510–13.

26. Spriggs, Matthew, Frederique Valentin, Stuart Bedford, Ron Pinhasi, Pontus Skoglund, David Reich, and Mark Lipson. 2019. "Revisiting Ancient DNA Insights into the Human History of the Pacific Islands." *Archaeology in Oceania* 54 (1): 53–56.

27. Stebbins, G. Ledyard, G. Ledyard Stebbins, and Sewall Wright. 1977. "Evolution and the Genetics of Populations. Volume 3. Experimental Results and Evolutionary Deductions." *Systematic Botany*. https://doi.org/10.2307/2418471.

28. Tätte, Kai, Ene Metspalu, Helen Post, Leire Palencia-Madrid, Javier Rodríguez Luis, Maere Reidla, Erika Tamm, et al. 2022. "Genetic Characterization of Populations in the Marquesas Archipelago in the Context of the Austronesian Expansion." *Scientific Reports* 12 (1): 5312.

29. Templeton, Alan R. 2021. *Population Genetics and Microevolutionary Theory*. John Wiley & Sons.

30. Whiteley, Andrew R., Sarah W. Fitzpatrick, W. Chris Funk, and David A. Tallmon. 2015. "Genetic Rescue to the Rescue." *Trends in Ecology & Evolution* 30 (1): 42–49.