Sex-biased admixture and assortative mating shape genetic variation and influence demographic inference in admixed Cabo Verdeans

Katharine L Korunes,[1] Giordano Bruno Soares-Souza,[2] Katherine Bobrek,[3] Hua Tang,[4] Isabel Inês Araújo,[5] Amy Goldberg,*[1,6] Sandra Beleza**[2,6]

1) Evolutionary Anthropology, Duke University, Durham, North Carolina, 27705, USA
2) Department of Genetics and Genome Biology, University of Leicester, Leicester, LE1 7RH, United Kingdom
3) Department of Anthropology, Emory University, Atlanta, Georgia, 30322, USA
4) Department of Genetics, Stanford University School of Medicine, Stanford, California, 94305 USA
5) Faculdade de Ciências e Tecnologia, Universidade de Cabo Verde (Uni-CV), Praia, Ilha de Santiago, Cabo Verde
6) These authors contributed equally to this work
*Correspondence: amy.goldberg@duke.edu
**Correspondence: sdsb1@leicester.ac.uk

Abstract

Genetic data can provide insights into population history, but first we must understand the patterns that complex histories leave in genomes. Here, we consider the admixed human population of Cabo Verde to understand the patterns of genetic variation left by social and demographic processes. First settled in the late 1400s, Cabo Verdeans are admixed descendants of Portuguese colonizers and enslaved West African people. We consider Cabo Verde's well-studied historical record alongside genome-wide SNP data from 563 individuals from 4 regions within the archipelago. We use genetic ancestry to test for patterns of nonrandom mating and sex-specific gene flow, and we examine the consequences of these processes for common demographic inference methods and for genetic patterns. Notably, multiple population genetic tools that assume random mating underestimate the timing of admixture, but incorporating non-random mating produces estimates more consistent with historical records. We consider how admixture interrupts common summaries of genomic variation such as runs-of-homozygosity (ROH). While summaries of ROH may be difficult to interpret in admixed populations, differentiating ROH by length class shows that ROH reflect historical differences between the islands in their contributions from the source populations and post-admixture population dynamics. Finally, we find higher African ancestry on the X chromosome than on the autosomes, consistent with an excess of European males and African females contributing to the gene pool. Considering these genomic insights into population history in the context of Cabo Verde's historical record, we can identify how assumptions in genetic models impact inference of population history more broadly.

Author Summary

Patterns of genetic variation are often used to infer human population histories; however population-genetic models make a variety of simplifying assumptions that often neglect the demographic and social dynamics of human populations. Here, we use the population of Cabo Verde as a case study to understand the patterns of genetic variation left by social and demographic processes. The islands of Cabo Verde were first settled in the late 1400s by Portuguese colonizers and enslaved West African people. We consider genomic data from four regions within the archipelago alongside historical records that document settlement patterns, timing, and sociocultural dynamics within the islands. We use genetic ancestry to test for nonrandom mating and sex-specific demography. We show that these sociocultural processes may bias inference of population history parameters and distributions of shared ancestry in the region. Overall, by providing insights into the patterns of genetic variation social processes leave in human genomes for a population with a well-studied historical record, we highlight processes to consider in order to more accurately understand the history of populations without extensive records.

Introduction

Patterns of genetic variation are often used to infer human population histories, such as fluctuations in population size, the timing of gene flow, and changes in population structure (1–7). These studies have been informative about ancient and modern populations that may lack

historical records (8–12). However, by necessity, population-genetic models make a variety of simplifying assumptions that often neglect the demographic and social dynamics of human populations. For example, many classical models assume random mating, but within empirical populations mating is often nonrandom, known as assortative mating. Positive assortative mating in human populations can keep haplotype blocks together, affecting evolutionary processes and our ability to infer population history (13–15). Further, male and female demographic processes can differ and vary over time. Comparisons between the X chromosome and autosomes have shown that population sizes and migration rates differ between males and females of many human populations (16–20).

The challenges of inferring human demographic history are particularly apparent on short timescales of tens of generations, where changes in allele frequencies may be difficult to observe and biased by dynamics not captured by classic models. Admixed populations provide an opportunity to examine evolutionary processes on short timescales using admixture linkage-disequilibrium (LD) structure rather than potentially small changes in allele frequencies. Understanding the patterns generating the distribution of ancestry within an admixed population is also important as genomic data has shown how widespread admixture is throughout human evolution (8,11,21–26).

In admixed populations, nonrandom mating can be driven by factors that correlate with genetic ancestry. For example, spouse pairs within multiple Latino populations are correlated in their genomic ancestry (13,15). When mating patterns lead to a positive correlation in genomic

63 ancestry, this specific type of nonrandom mating, referred to as ancestry-assortative mating, can
64 increase the likelihood that individuals in mating pairs share recent common ancestors. Recent
65 work has begun to assess the implications of these patterns, including effects on common
66 demographic inference strategies and on genetic/phenotypic variation (7,14,24). One genetic
67 consequence of positive ancestry-assortative mating is an increase in long runs of homozygosity
68 (ROH), which arise when long tracts of homozygous genotypes occur due to inheritance of
69 identical by descent (IBD) haplotypes from recent common ancestors. Across modern human
70 populations, ROH are common and typically range in length from kilobases to megabases (27–
71 30). In the field of population genetics, we are just beginning to learn about ROH dynamics and
72 expectations in admixed populations, which do not necessarily fit expectations built from
73 homogenous populations (30,31). For example, we know that distinct length classes of ROH
74 reflect demographic factors from different timescales, but as we explore here, it is unclear what
75 these distributions of ROH mean in populations that have undergone major demographic
76 transitions such as recent admixture. Patterns of genetic variation can also be shaped by
77 differences in male and female demographic processes. In many human populations,
78 population sizes and migration rates differ between males and females, leading to differences in
79 ancestry patterns on the X chromosome vs the autosomes (16–20).
80
81 Here, we use the admixed population of Cabo Verde as a case study to understand the patterns
82 of genetic variation left by multifaceted social and demographic processes in humans. Admixed
83 populations of African ancestry, such as the population of Cabo Verde, are often excluded in the
84 context of medical genetics and human evolutionary genetics, despite their importance and

85 widespread global presence (32–34). The record of Cabo Verde population history is vast,
86 including royal charters, letters from Crown officials, captains and other sea explorers' journals,
87 records of the economic activities in Cabo Verde (e.g., landowners' rents and number of
88 enslaved individuals traded), detailed ship logs of trade activity, and church censuses from the
89 18th century on, which allows us to generate realistic assumptions about colonization and
90 demography (35–41). We consider patterns of genetic variation alongside historical records that
91 document many aspects of Cabo Verdean history, including settlement patterns, timing, and
92 sociocultural dynamics. Specifically, we use distributions of genetic ancestry to test for ancestry-
93 assortative mating and sex-specific gene flow. We examine the consequences of these processes
94 for genetic variation, such as patterns of homozygosity, and for common demographic inference
95 methods. In turn, by elucidating the patterns of genetic variation social processes leave in
96 human genomes for a population with a well-studied historical and ethnographic record, we
97 can better use population-genomic methods to explore the history of populations without
98 extensive records.

99 Cabo Verdean history

100 The recently admixed human population of Cabo Verde presents a particularly relevant
101 opportunity to examine how sex-biased admixture and nonrandom mating shape genetic
102 patterns and influence demographic inference. Cabo Verde is an archipelago off the coast of
103 Senegal, inhabited today by admixed descendants of Portuguese colonizers and enslaved West
104 African people who settled the unpopulated islands beginning in the mid-1400s (42–45). Since
105 the archipelago was unoccupied prior to the 1460s, we have historical knowledge of the start of

106 admixture and the contributing source populations, with written records clearly documenting
107 arrival times by island. Additionally, the island geography imposes population structure within
108 Cabo Verde, and historical records document differences in population sizes, mating patterns,
109 and social customs by island population (see Methods for historical data sources).
110
111 The settlement of the islands was influenced by island geography and ecology, and is often
112 divided into three different temporal stages, which are associated with changes in the economy
113 (35). The initial settlement stage (beginning in the 1460s) corresponds to the peopling of the
114 Southern island Santiago, followed by the nearby island of Fogo 20 years later (Fig 1). During
115 this time, the economy was centered around large cotton plantations and expanding trade with
116 the Senegambian coast. This trade included the exportation of great numbers of enslaved
117 Africans to Cabo Verde, the majority of which became part of the commercial trade to the
118 Americas (36,37). In the 17th century, following the decline of the slavery-based economy of
119 Santiago and Fogo, many free farmers seeking better farming conditions migrated to the
120 northwestern islands of Santo Antão and São Nicolau. These migrations comprised the second
121 settlement stage. In contrast to Santiago and Fogo, these northwestern islands did not become
122 populous economic centers (35). During the third and final settlement stage, São Vicente in the
123 northwestern group of islands was colonized by settlers from the neighboring islands in the
124 mid 19th century. With the opening of Mindelo's harbor at that time, São Vicente prospered
125 from the advent of commercial Atlantic shipping, becoming the second most important island
126 of the archipelago after Santiago in terms of population size. In the 19th century, the eastern

127 islands Sal, Boa Vista, and Maio were also open to English and North American ships, but these
128 islands never became densely populated (35,37).
129
130 Previous investigations of genetic ancestry in Cabo Verde (42,43,45) have shown extensive
131 admixture in the archipelago and differences in the mean African genetic ancestry across
132 islands, which may reflect the differences in the settlement history and suggest restricted gene
133 flow between islands. These past studies of Cabo Verdean genetic admixture have explored
134 how genetic ancestry connects to phenotypic variation such as skin and eye color (42,43), and to
135 cultural information such as linguistic variation (45). Some earlier analyses of post-admixture
136 population structure were based on Y-chromosome diversity (35), giving insight into male-
137 specific demographic processes. Here, we examine how historical and social processes influence
138 genetic variation and population-genetic inference of demography. We consider how the
139 distributions of segments of DNA shared identically by descent within and among individuals
140 (as measured by IBD, kinship, and ROH) relate to patterns of genetic ancestry in an admixed
141 population. We integrate several population genomic approaches using autosomal and X-
142 chromosome ancestry and genetic variation patterns to infer the sex-specific demographic
143 history of the last ~20 generations since founding, and we consider the potential biases of
144 ancestry-assortative mating on the inference of admixture timing. Overall, we provide insights
145 into the population history of Cabo Verde and demonstrate how admixed populations can
146 provide powerful test cases for understanding evolutionary processes on short timescales.

Results

Patterns of shared ancestry reflect the colonization history of the islands

Using genome-wide SNP data from 563 individuals, we examined four island regions of Cabo Verde (Fig 1; Supp Fig 1), which differ in their settlement histories. The partitioning of the considered island regions was also supported by genetic patterns. Namely, these regions showed quantitatively distinct distributions of IBD tracts, as we explore below, and were supported by clustering patterns in PCA (Supp Fig 1), in which the three northwestern islands grouped together (and with the eastern island Boa Vista) despite the different settlement time of São Vicente Island.

Global ancestry proportions reflect extensive European and West African admixture and concur with previous results (43) (Supp Fig 2), with the median West African autosomal ancestry in Cabo Verdeans varying by island from 50% in Fogo, 56% in the northwestern islands in aggregate (the Northwest Cluster), 63% in Boa Vista, and 75% in Santiago. These inter-island differences alongside historical knowledge provide a comparative dataset for understanding how ancestry patterns can be used to infer demographic processes. Further, the known bound on admixture timing and the relative simplicity of the admixture history on the islands allows genetic ancestry to be decomposed into two source populations, which makes Cabo Verde a case study closer to population genetic models of admixture than most worldwide human populations.

The process of admixture can influence measures such as IBD and ROH that are often used to inform inference of population history, and theoretical expectations for these measures are less clear for admixed populations compared to homogeneous populations. Thus, we use multiple methods to examine relatedness in Cabo Verde, and we use this case study to underscore the need for further empirical and theoretical work to understand the dynamics of IBD and ROH in admixed populations. Patterns of IBD within and between populations provide opportunities to examine common ancestry based on the number and sizes of segments of IBD. Using these summaries of IBD to examine relatedness between and within regions of Cabo Verde, we found that patterns of shared ancestry reflect the successive settlement history of the islands. Notably, Santiago has the lowest mean number and total length of IBD segments between individuals (Supp Fig 4-5). In contrast, we observe the highest levels of IBD within and between the Northwest Cluster and Boa Vista.

To summarize patterns of IBD within and between the four island regions, we built a network of relationships with pairs of individuals connected if they have IBD levels in the top 3% of the Cabo Verde IBD distribution (Fig 1; see Supp Fig 3 for lower levels of IBD and within-island networks). In a non-admixed population, these edges roughly correspond to 4^{th}-degree relatives or closer (e.g., great-great-grandparents, great-great-grandchildren, great-aunts/uncles, or grand-nieces/nephews). At this level of relatedness, we find that the presence of pairs of individuals within and between islands connected by an edge is common, suggestive of recent relatives both within and across islands. However, relatedness parameters estimated under

189 simple models of structure likely do not hold in many human populations, such as recently
190 admixed populations. Kinship estimated under a framework (46) specifically designed for
191 arbitrary population structures (Supp Fig 6A) is consistent with both the overall trends in IBD
192 tracts (Supp Fig 4-5), and supports the presence of a subset of very closely related individuals
193 within the data.
194
195 The lower levels of IBD (Fig 1, Supp Fig 3-5) and lower kinship estimates (Supp Fig 6) observed
196 in Santiago are consistent with expectations based on the island's history, as Santiago was the
197 first Cabo Verdean island to be founded and has the largest population size. Santiago is also
198 likely to have relatively high genetic diversity as a result of having the highest proportion of
199 West African ancestry, based on the expectation that kinship in human populations generally
200 reflects serial bottlenecks due to dispersal from Africa (47,48). In contrast, the high levels of IBD
201 within and between the Northwest Cluster and Boa Vista reflect well-documented serial
202 founding and migrations during the settlement of the islands. In general, kinship estimates also
203 reflect the colonization history of the islands, with Cabo Verdeans of greater European ancestry
204 sharing greater kinship compared to individuals with greater West African ancestry (Supp Fig
205 6). Exceptions to this trend were most noticeable in individuals from Fogo and the Northwest
206 Cluster, where it was most apparent that some observed relatedness patterns do not result in
207 smooth ordering of global ancestry proportions (Supp Fig 6B). This observation led us to
208 hypothesize that other sociocultural processes beyond proportions of ancestry, such as
209 nonrandom mating, may drive patterns of relatedness in Cabo Verde.

Nonrandom mating shapes patterns of ancestry and influences demographic inference

To test for ancestry assortative mating within Cabo Verde, we examined whether the genomic ancestries of individuals in mating pairs correlate with each other. To this end, we applied ANCESTOR to computationally infer the parental ancestry proportions that likely preceded the ancestry haplotypes we observe today. We repeated this analysis for each chromosome independently to prevent uncertainty introduced by matching interchromosomal haplotypes. As an example, the inferred ancestries of the parental haplotypes that likely preceded chromosome 7 are significantly positively correlated for all islands except for Boa Vista (Fig 2A). We observed similar results across the full set of autosomes (Fig 2B). We found that the inferred ancestries of mating pairs in the previous generation are positively correlated, and that these correlations differ significantly from expectations under random mating (random sampling empirical distributions of parental haplotypes into pairs, shown in Supp Fig 7).

We next considered how the observed genetic evidence of assortative mating can be leveraged in the context of inferring population history. To estimate the timing of the onset of admixture, we applied the method of Zaitlen et al. (2017) (7), which describes the decay of local ancestry disequilibrium (LAD) as a function of assortative mating strength, migration rate, recombination rate, and the number of generations since admixture began. Using this LAD-based method, we inferred admixture timing under random mating and under our empirical estimates of ancestry-assortative mating strength (Fig 3). The LAD-based method produced

231 older estimates of admixture timing under a model including both assortative mating and
232 constant migration (Fig 3, Supp Fig 8), with increasingly older timing estimates as assortative
233 mating strength increases (Supp Fig 8). Under the same set of migration rates (0 or 1% per
234 generation, to demonstrate estimates under a broad range of migration rates), when we varied
235 the strength of ancestry-assortative mating from a situation of random mating to a situation of
236 strong ancestry-assortative mating, the models that considered substantial ancestry-assortative
237 mating (parental correlation in ancestry > 0.3), yielded admixture timing estimates closest to the
238 historical estimates (Fig 3, Supp Fig 8).
239
240 We additionally applied two strategies, ALDER (4) and MultiWaver 2.0 (5), that are not
241 intended to account for assortative mating. ALDER uses the extent of LD decay among
242 neighboring loci to infer mixture proportions and dates. In contrast, MultiWaver uses the length
243 distribution of ancestral tracks. Assuming a generation time of 25 years, decay rates estimated
244 with ALDER suggest admixture timing in the mid to late 1700s (Fig 3; all timing estimates are
245 shown in Supp Table 1). MultiWaver chose a multi-wave admixture model for Santiago (Fig 3A)
246 and the Northwest Cluster (Fig 3C), but selected an isolation model for Fogo (Fig 3B) and Boa
247 Vista (Fig 3D).
248
249 The estimates of admixture timing that do not explicitly consider ancestry-assortative mating,
250 from ALDER and MultiWaver, are noticeably more recent than the known founding of Cabo
251 Verde in late 1400s. Results from the LAD-based method described in Zaitlen et al. (2017) (7)
252 align more closely with historical records, particularly when we allowed for both ancestry-

assortative mating and constant migration (Fig 3, Supp Fig 8). While multiple complex demographic factors likely impact these estimates of admixture timing (see "Patterns of genetic variation and admixture in Cabo Verde influence demographic inference" in the Discussion), the evidence of ancestry-assortative mating in Cabo Verde (Fig 2, Supp Fig 7) and the LAD-based inferences incorporating assortative mating (Fig 3, Supp Fig 8) suggest that explicitly accounting for ancestry-assortative mating improves estimates of admixture timing, while the assumption of random mating leads to underestimation of the age of admixture.

Runs-of-homozygosity (ROH) reflect the contributions of the source populations and patterns of nonrandom mating.

We identified ROH using the method of Pemberton et al. (2012) as implemented in the software GARLIC (30,49). This method classifies ROH into three categories (short, medium, and long) based on the modeling of the length distribution in each population as a mixture of Gaussian distributions. ROH arise when IBD haplotypes are inherited from a common ancestor. Thus, distinct length classes reflect inheritance from ancestors at different timescales in the population's history. We examined the distributions of total ROH per genome (sum of ROH segments of a specific length) within Cabo Verde, partitioning the distributions into length classes that are enriched for pre- and post-colonization processes. Under the ROH classification model of GARLIC, short and medium ROH formed due to processes that largely occurred within the source populations prior to the founding of Cabo Verde. Thus, while ROH in non-admixed populations are often modeled as three length classes, we emphasize the differences in

pre- and post-admixture dynamics in Cabo Verde by focusing on shorter ROH (comprising both short and medium as called with GARLIC) vs. long ROH (Fig 4A-B). To ensure our comparisons of shorter vs long ROH used reasonable classification boundaries, we repeated this analysis under four alternative length classification schemes (see Methods; Supp Fig 9). The results were consistent across the four classification schemes (Supp Fig 9).

Since methods for detecting and interpreting ROH have been developed primarily in non-admixed populations (30,50), we consider ROH in Cabo Verde as an example of ROH distributions in admixed populations where we can explore results in the context of ancestry and population history. For example, to test for genetic evidence that population bottlenecks contributed to observed patterns, we estimated ancestry-specific population sizes (Supp Fig 10). Ancestry-specific population sizes in Cabo Verde suggest bottlenecks of both source populations, followed by population expansion within the past 10 generations (1,2). If the length of ROH in an individual's genome is driven by distance from Africa, as expected based on previous work (29,30), we predict the total length of ROH per genome to negatively correlate with West African admixture proportion. Consistent with this expectation, we observe a negative correlation between the total ROH and West African ancestry proportions in Santiago, Fogo, and the Northwest Cluster (Supp Fig 11A; correlations and significance values for total ROH and by length class are provided in Supp Fig 11). Breaking total ROH into shorter and long ROH, we see that this negative correlation is driven primarily by shorter ROH (Fig 4; Supp Fig 11B-C, Supp Fig 9).

295 Comparing the Cabo Verdean islands to each other, we observed the lowest levels of total ROH
296 in all length classes of ROH in Santiago (pairwise Mann-Whitney U tests $p < 1 \times 10^{-8}$). The low
297 amounts of all length classes of ROH in Santiago are consistent with several historical attributes
298 of the islands, including serial founding beginning with the settlement of Santiago. Serial
299 founding resulting in bottlenecks is expected to increase shorter ROH by reducing the number
300 of ancestral lineages. But because shorter ROH reflects background relatedness from the source
301 populations, shorter ROH are expected to be more sensitive to the contributions of the source
302 populations.
303
304 Despite recent colonization bottlenecks, we observed that some admixed individuals in Cabo
305 Verde present lower levels of ROH per genome than either of the source population proxies.
306 Specifically, our classification of ROH into shorter and long ROH (Fig 4; Supp Fig 9) revealed
307 that Santiago has significantly lower levels of shorter ROH compared the West African
308 reference population (Mann-Whitney U $p = 5.87 \times 10^{-7}$), while long ROH are not depleted
309 (Mann-Whitney U $p = 0.841$). This observation that Santiago has significantly lower levels of
310 shorter ROH compared to even the West African reference is perhaps surprising given the
311 canonical view of African source populations as generically having the lowest ROH in
312 worldwide samples. This is consistent with the idea that shorter ROH are contributed to the
313 population from older shared ancestors in the source populations, and that these tracts can be
314 further interrupted with local ancestry from another source population upon admixture. In
315 contrast, long ROH are likely driven by post-admixture dynamics such as small population size
316 and mating preferences, and long ROH tracts may span multiple local ancestries. Indeed, we

observed that ROH often span multiple ancestries, with long ROH spanning more switches in local ancestry per megabase. Specifically, 1.21% of shorter ROH contained at least one ancestry switch and 18.99% of long ROH contained at least one ancestry switch. Dividing ancestry switches by tract length, shorter ROH contain an average of 0.072 ancestry switches/Mb (standard deviation = 0.89), while long ROH contain an average of 0.089 switches/Mb (standard deviation = 0.24) (pairwise Mann-Whitney U test $p < 1 \times 10^{-8}$).

While ROH patterns in Cabo Verde are consistent with several historical observations, we emphasize that caution is warranted in applying methods and expectations built using non-admixed populations. We also note that, though we used a model-based ROH detection approach that included steps to mitigate genotyping errors and biases in allele frequency estimates, sensitivity and specificity of ROH detection is generally lower for shorter ROH. Further, records suggest that enslaved African individuals in Cabo Verde came from the Senegambian region of Africa, and lack of genomic data from this vast region makes it difficult to assess how the diversity in contributing African ethnic groups might contribute to low shorter ROH levels in Cabo Verde. These sources of uncertainty necessitate caution in inferring population history from ROH distributions. However, the observed differences in ROH among the islands suggest that ROH are sensitive to population genetic processes even on the short timescale since Cabo Verde's settlement. Together, these observations and caveats underscore the need for future work testing the effects of admixture on ROH and on ROH detection, as we explore in the Discussion.

Ancestry patterns on the X chromosome vs the autosomes reflect sex-biased demographic processes

On all islands, autosomal versus X chromosome ancestry patterns suggest that male and female contributions differ significantly by source population (Fig 5A). Specifically, there is higher West African ancestry on the X chromosome than the autosomes (mean African ancestry proportion for the X chromosome = 0.76; mean African ancestry proportion for autosomes = 0.60; Wilcoxon Signed-Rank Test $p < 1 \times 10^{-8}$ for Santiago, Fogo, and the Northwest Cluster; $p = 0.0013$ for Boa Vista), consistent with sex-biased contributions of the founders. To quantitatively explore these differences in male vs female demographic history, we used a model-based approach to estimate differences between male and female contributions (3). A model of constant admixture is supported over a model of instantaneous admixture by historical records (see Methods for historical data sources) and by our comparison of admixture timing estimates (above). Using the method of Goldberg and Rosenberg for a model of constant admixture, we compared observed and expected X-chromosomal and autosomal ancestry over a grid of possible parameter values. Following the implementation from Goldberg et al. (2017) (51), Fig 5B presents results from the 0.1% of parameter sets closest to observed data based on a Euclidean distance between observed population mean ancestries on the X and autosomes and expected ancestries under the model (Supp Fig 12 shows the same data represented by plotting the male contributions). These estimated sex-specific contributions from West African and European source populations under a model of constant admixture suggest greater contributions from West African females and European males.

359

360 Given our observation that source populations make distinct contributions to ROH in Cabo
361 Verde, we next investigate differences in autosomal vs X chromosome ROH content. Differences
362 in autosomal vs X chromosome ROH may arise in part due to the source populations
363 contributing in a sex-biased manner. Notably, this effect would be seen most in shorter ROH,
364 since shorter ROH reflect the homozygosity of older haplotypes and background relatedness
365 from the source populations. However, it is challenging to disentangle processes shaping X vs
366 autosomal ROH, due to the smaller effective population size of the X chromosome. Our results
367 suggest that sex-biased admixture processes in Cabo Verde are reflected in ROH, with lower
368 levels of shorter ROH on the X chromosome than on autosomes (Fig 6A). European individuals
369 have higher levels of ROH than West African individuals, and the higher contributions of
370 African X chromosomes (vs European X chromosomes) may drive the lower levels of short
371 ROH in Cabo Verde on the X chromosome vs the autosomes (Fig 6A). Long ROH reflects
372 different dynamics, potentially including the reduced post-admixture population size and other
373 sex-specific processes.

374 Discussion

375 In this study, we leveraged patterns of genetic variation in Cabo Verde to infer the demographic
376 history of the past ~20 generations. We found that distinct genetic patterns of four island
377 regions within the archipelago reflect the colonization history of the islands, including island-
378 specific settlement timing, admixture dynamics, mating patterns, and sex-biased demography.

379 Together these results demonstrate how patterns of ancestry and genetic variation are shaped
380 by social and demographic forces on short timescales. By better understanding how complex
381 population histories generate genetic variation, we can improve interpretation of inference from
382 populations without historical records.

383 Patterns of genetic variation and admixture in Cabo Verde reflect
384 colonization history and subsequent sociocultural dynamics

385 The observed island-specific genetic patterns reflect the complex ecological and social factors
386 that shaped the settlement of Cabo Verde. Historical evidence suggests that the islands were
387 founded in a stepping-stone pattern, with settlement stages that were accompanied by
388 important economic and sociocultural shifts (35). Early written records of the population
389 (described below) suggest that admixture began early during the settlement of Cabo Verde,
390 despite the highly racially stratified, slavery-based system that characterized the first settlement
391 stage (38). The second and third stages were carried out by mostly already-admixed individuals
392 who had become a significant group within Cabo Verdean society and who migrated from the
393 southern to the northern islands (39). We found that the staggered settlement history and the
394 island-specific population dynamics shaped patterns of ancestry and genetic variation within
395 Cabo Verde.
396
397 Admixture in Cabo Verde is consistent with a model of continuous gene flow from Europe and
398 West Africa, with European males and African females contributing predominantly to the Cabo

Verdean gene pool, as seen in other admixed populations that result from the trans-Atlantic slave trade (52,53). PCA (Supp Fig 1), IBD (Fig 1, Supp Fig 3-5), and kinship (Supp Fig 6) are consistent with genetic drift occurring during the archipelago's settlement history of consecutive founder effects and subsequent relative isolation of the islands. Specifically, individuals are genetically more related within an island than among islands, and considerably more individuals from the northern islands and Fogo are connected to each other by high IBD pairwise connections (Supp Fig 4-5). We observed differences in IBD that reflect the contributions of the source populations, such as Santiago having both the highest West African ancestry proportions and the lowest overall levels of IBD sharing. The higher African genomic ancestry on Santiago is consistent with previous results (42,43), and indicates differences in settlement patterns despite the fact that colonization of the other islands involved migrations from Santiago and was based on the same slavery-based system.

Using a mechanistic model of sex-biased admixture (3), we used differences in ancestry on the X chromosome vs the autosomes to examine sex-biased migration in Cabo Verde. We found that admixture occurred primarily through the mating of European males and African females, consistent with historical work that documents sex bias during the settlement of Cabo Verde. First records of the demographic distribution in Cabo Verde are found in a letter by a judicial official Pero de Guimarães to the Portuguese King in 1513, which stated the presence of 118 European individuals in Santiago, of which only four were (single) women (54). The marginal presence of European females throughout the founding of Cabo Verde contributed to the extensive genetic admixture currently found in the archipelago. The distribution of ROH in

421 autosomes versus X chromosome is influenced by initial male versus female contributions to the
422 admixed population's gene pool. Throughout the islands, the higher contributions of African X
423 chromosomes (vs European X chromosomes) may drive the lower levels of short ROH in Cabo
424 Verde on the X chromosome vs the autosomes (Fig 6A). These results underscore the long-
425 lasting genetic impacts of the trans-Atlantic slave trade, as has been recently shown in admixed
426 populations across the Americas (53).
427
428 Despite only a few hundred years of unique population histories among the islands, we also
429 found that patterns of genetic variation reflect island-specific patterns in post-admixture
430 dynamics. Observed patterns of IBD, kinship, and long ROH are all consistent with overall high
431 levels of relatedness within Cabo Verde, particularly in Fogo and the northwestern islands. In
432 other worldwide populations, similarly high IBD levels have been attributed to founder effects
433 (55,56), though we note that pairwise IBD measures cannot be directly compared due to the
434 admixture process. In Cabo Verde, patterns of relatedness may be shaped by both founder
435 effects and by post-admixture dynamics such as nonrandom mating keeping similar haplotypes
436 together rather than distributing them randomly throughout mating pairs in the population.
437 Indeed, we inferred a positive correlation between mating pairs of the previous generation,
438 suggesting ancestry-assortative mating on the short timescale since the founding of the islands.
439 Mating patterns and sex-biased admixture are distinct—but interacting—processes in the
440 history of Cabo Verde. Sex-biased migration determined the gene pool within Cabo Verde,
441 while patterns of nonrandom mating within the islands shaped subsequent generations of
442 admixture. These mating patterns were tied to social and economic structure. European

443 enslavers sexually coerced enslaved African women during slavery. After the end of slavery,
444 racially unbalanced control of landownership and higher social status of Europeans resulted in
445 continued socioeconomic pressures for African women to have children with European men.
446 Such unions often resulted in improved socioeconomic status for both the women and their
447 children. For example, the first admixed individual to assume an important position within the
448 administration of Santiago was recorded in 1546 (57). Over generations, Cabo Verde became a
449 population of extensively admixed ancestry, with mating patterns shaped by the interconnected
450 effects of social position and skin color. This type of nonrandom mating may have been more
451 prominent in the second half of the 17th century, due to the large exodus of the white population
452 following the decay of the slave trade. With this population shift, historical sources say that the
453 socioeconomic elite of the islands became predominately the “white children of the islands"
454 (38), leading to further racial segregation. This historical evidence of nonrandom mating
455 patterns—together with our genetic observations—demonstrates how quickly changes in social
456 structure can impact population homozygosity and differentiation.
457
458 The island of Fogo stands out as having unique social and historical processes, even though it
459 was settled shortly after Santiago. Fogo’s society was, since its origins, a conservative rural
460 society whose main economic activity was to produce goods to trade in the African coast. The
461 socioeconomic elite was patriarchal and aristocratic (40), and composed primarily of related
462 families, which promoted first-cousin marriages (41). The sustenance of this class depended on
463 land ownership, which allowed the slavery-based system to be perpetuated longer in Fogo than
464 in Santiago (35,40). The unique attributes of Fogo are reflected in population genetic patterns.

For example, we found that African ancestry proportion align with kinship patterns overall, but exceptions were most obvious in individuals from Fogo and the Northwest Cluster, where some observed relatedness patterns are not as clearly explained by global ancestry proportions. In our estimation of admixture timing using local ancestry disequilibrium, we observed that Fogo has greater levels of local ancestry disequilibrium than the other islands. Higher local ancestry disequilibrium may be driven by ancestry-assortative mating. Additionally, it may be that Fogo has experienced stronger founder effects, which would decrease the number of ancestral lineages. Indeed, the higher levels of shorter ROH within Fogo (Fig 4; Supp Fig 9) are consistent with founder effects increasing background relatedness and thus increasing shorter ROH. Though we emphasize the need for further work to understand the dynamics of ROH in admixed populations, these patterns are in agreement with the lower Y-chromosome haplotype diversity observed in Fogo compared to the other islands (42). Long ROH and IBD patterns gives us further insights about the post-admixture demographic processes in the archipelago, including uncovering genetic consequences of consecutive founder effects.

Patterns of genetic variation and admixture in Cabo Verde influence demographic inference

The admixture dynamics, founder effects, and mating patterns within Cabo Verde shape summaries of genetic variation that are often used to inform demographic inference, such as local ancestry disequilibrium, IBD, and ROH. To investigate the effects of these population dynamics on demographic inference, we estimated admixture timing using multiple population

genetic tools. Estimates of admixture timing from genetic data were most concordant with historical records when using inference based on local ancestry disequilibrium (7), allowing for ancestry-assortative mating and migration after the founding of the admixed population. Under this method, we tested a range of ancestry-assortative mating strengths (0 vs values inferred using ANCESTOR in Fig 3 and a broader range of values in Supp Fig 8) and migration rates (0 vs 1% each generation). The migration levels are used to demonstrate the trends of timing estimates using constant migration and assortative mating, rather than to obtain a specific estimate of the migration parameters. Consistent with previous theoretical work on the impact of assortative mating on the timing of admixture, we find that estimates that do not account for ancestry assortment are more recent than historical records and than estimates that include ancestry assortment (7,58). Under random mating, haplotypes are distributed randomly in mating pairs in a population, allowing recombination to shuffle haplotypes over generations. In contrast, ancestry-assortative mating can keep ancestral haplotypes more intact, leading to underestimation of the number of generations of admixture when estimating admixture timing. Zaitlen et al. (2017) and Goldberg et al. (2020) theoretically demonstrate that certain models of non-random mating in admixed populations maintain variation in the ancestry proportion and linkage over time—summary statistics that will make admixture appear more recent than it is when mating patterns are not considered (7,58). Our evidence of ancestry-assortative mating in Cabo Verde underscores the importance of accounting for nonrandom mating in understanding admixture and inferring demographic history. Assortative mating has been documented with respect to genetic ancestry, socioeconomic factors, and phenotypic characteristics (7,13–15,59),

506 but remains an under-recognized force in human population genetics and demographic
507 inference.
508
509 Notably, assortative mating alone was not enough to obtain time of admixture estimates
510 concordant with historical records. Admixture timing estimates were most concordant with the
511 historical records when allowing both assortative mating and constant admixture. An
512 admixture scenario that allows multi-wave or constant admixture model is compatible with
513 historical work suggesting constant migration from both Europe and West African throughout
514 the settlement of the islands. Importantly, the authors of ALDER and MultiWaver point out that
515 methods for inferring admixture timing have varying sensitivities to different admixture
516 scenarios (4,5). MultiWaver may not be able to accurately infer demographic histories that
517 deviate from its pre-defined admixture models. ALDER notes that multi-wave or continuous
518 admixture can lead to more recent estimates of admixture timing, as the method inherently
519 assumes single-point admixture. Despite this caveat, ALDER is frequently applied in cases
520 where admixture is not strictly instantaneous.
521
522 All of the timing methods we used placed admixture timing for the different islands closer
523 together than historical dates of settlement, consistent with historical expectations that the
524 initial admixture in the southern islands was significant, and that many individuals that
525 occupied the northern and eastern islands during the second and third settlement stages of
526 Cabo Verde were already admixed. For example, the serial founding of the islands may explain
527 why estimates of admixture timing for Boa Vista were closer to historical records. While Boa

Vista was founded most recently, it was founded by already admixed individuals. This observation, together with IBD and kinship patterns, support the serial founding of the groups of islands as the main model of settlement of Cabo Verde, as opposed to their independent settling as some historical data suggest (35). This type of serial founding scenario is common throughout recent human migration, underscoring that settlement patterns, in addition to settlement timing, are critical components of accurately inferring human demographic history.

We found that Cabo Verde's island-specific demographic history and admixture dynamics have important genomic consequences, as observed with ROH. Despite the relatively recent colonization of the islands, some individuals presented even lower overall levels of ROH than African reference populations. We found that low overall levels of ROH in Cabo Verde are driven by shorter length ROH. This observation is consistent with the idea that shorter ROH can be attributed to older shared ancestors from the source populations, and these tracts can be interrupted with admixed ancestry upon admixture. However, many admixed individuals still present excess long ROH, likely reflecting post-admixture processes such as serial founding and ancestry-assortative mating. The observed island-specific distributions of ROH are consistent with the colonization history of the islands. For example, the cultural dynamics and higher rates of first-cousin marriages in Fogo (41), possibly driving high levels of long ROH (Fig 4B). In contrast, Santiago has both the oldest population and the largest population size, and has comparatively low levels of both shorter and long ROH. These observations suggest that more work, both empirical and theoretical, is needed to understand the interacting forces of local ancestry and ROH.

In sum, we provide insights into the population history of Cabo Verde and demonstrate how admixed populations can provide powerful test cases for understanding demographic processes and genomic consequences in recent human history. We show that patterns of shared ancestry between and within the islands (quantified with IBD and kinship estimates) reflect serial founder effects as well as settlement patterns such as post-admixture nonrandom mating. We find that accounting for nonrandom mating allows us to improve inference of admixture timing and better contextualize genomic consequences of admixture dynamics, such as ROH. We find that differences in ancestry on the X chromosome vs the autosomes reflect sex-biased demographic processes. Given the ubiquity of admixture throughout modern human population, these results provide important, generalizable considerations for the study of recent human evolution.

Methods

Study population and ancestry reference panels

We used genotype data from Beleza et al. (2013) (43), which included 563 Cabo Verdeans from the following regions of the archipelago: Santiago (n = 172), Fogo (n = 129), Boa Vista (n = 26), and the three northwestern islands in aggregate (Northwest Cluster; n = 236) (Figure 1; Supp Fig 1). The genotype data exclude cryptic related individuals (first-degree relatives) identified by kinship analyses at the time of sampling. In accordance with historical records and previous

work showing that Iberian and Senegambian populations are suitable proxies for the ancestry sources of Cabo Verde (45), we leveraged data from the 1000 Genomes Project to estimate admixture proportions and call local ancestry in the Cabo Verde individuals, as described below. We merged the Cabo Verde genotypes with genotypes from 107 GWD (Gambian in Western Division - Mandinka) and 107 IBS (Iberian Population in Spain) samples called from high-coverage resequencing data released through the International Genome Sample Resource (60,61). Our final merged dataset consisted of 884,656 autosomal and 20,967 X chromosome SNPs shared between the Cabo Verde samples and the reference samples, with average missingness rates of 0.0017 by SNP for autosomes, and 0.0024 for the X chromosome.

Historical records

Throughout our analyses, we draw comparisons between genetic results and historical records. We used primary historical documents, mainly historical letters to the Portuguese Crown, documenting demographic characteristics of the islands across time. Our analyses of the historical data took into consideration the generalized interpretations of historian scholars such as Correia e Silva (2001; 2002) (35,39), Cabral (2001; 2012) (38,40) and Baleno (2001) (36). These sources document the dates of the stages of settlement mentioned above, along with associated with changes in the economy. We refer to the interpretation of genealogical data by Cabral (2012) (40) and Barbosa (1997) (41) for evidence of the structure of the Fogo's society and for evidence of first-cousin marriages within Fogo (41).

Characterization of ancestry

To produce estimates of admixture proportions, we first performed unsupervised clustering of the samples using ADMIXTURE v1.3.0 (62) (see Supp Table 3 for a summary of all computational methods used in this study). ADMIXTURE was run separately for the autosomal and X chromosome datasets after pruning based on linkage disequilibrium (LD) using the indep-pairwise option of PLINK v1.9 with a 50-SNP sliding window incremented by 10 SNPs, and an LD threshold of r2 = 0.5 (63). This pruned dataset included 514,551 autosomal and 12,706 X chromosome SNPs. We estimated individual ancestries by averaging over ten independent unsupervised ADMIXTURE runs using K = 2, given the historical and genetic support for two source populations (43,45). Using the LD-pruned dataset, we also visualized the data using PLINK's principal component analysis (--pca) function, confirming that the samples cluster approximately based on their geographic memberships, and that West African vs European ancestry clearly separates on the first principal component (Supp Fig 1). We phased all samples using SHAPEIT2 (64). After running SHAPEIT -check to exclude sites not contained within the reference map, we ran SHAPEIT to yield phased genotypes at 881,279 autosomal SNPs and 20,793 X chromosome SNPs. We then called local ancestry with RFMix v1.5.4 (65) PopPhased under a two-way admixture model using the West African and European reference genotypes described above. These local ancestry calls are publicly available via Zenodo (66). Ancestry proportions estimated with ADMIXTURE and RFMix here are highly correlated with results from Beleza et al. (2013) (43), using the same genotype data but with different software (*frappe* (67) and SABER (68)) and an older reference dataset (Supp Fig 2). To test whether phasing and

609 local ancestry calls were robust against the choice of reference panels, we repeated the
610 SHAPEIT and RFMix steps using a reference dataset composed by all African (ACB, ASW, ESN,
611 GWD, LWK, MSL, and YRI) and European (CEU, FIN, GBR, IBS, and TSI) populations available
612 from the high-coverage resequencing data released by the 1000 Genomes Project (60,61). The
613 resulting local ancestry calls with all AFR and EUR reference panels correlated closely with calls
614 using GWD and IBS as reference panels (Supp Fig 2C). We also performed local ancestry calling
615 using ELAI, a method that performs both phasing and local ancestry assignment (69). Again
616 using the IBS and GWD reference genotypes described above, we ran ELAI under a two-way
617 admixture model using the following parameters: -mg (number of generations) 20, -s (EM steps)
618 30, -C (upper clusters) 2, and -c (lower clusters) 10. The resulting local ancestry calls from ELAI
619 correlated closely with calls from RFMix (Supp Fig 2D).

620 ## Inference of admixture timing

621 We applied three distinct strategies for estimating the timing of the onset of admixture in Cabo
622 Verde: ALDER (4), MultiWaver 2.0 (5), and a method based on patterns of linkage
623 disequilibrium between local ancestry tracts (7). We first converted the genotypes to
624 EIGENSTRAT format using *convert* (70,71). We then used ALDER to date admixture timing on
625 each island using default parameters with mindis = 0.005 and source populations GWD and
626 IBS. We ran MultiWaver using the ancestry tracts inferred by RFMix, default parameters, and
627 100 bootstraps. Finally, we followed the pipeline of Zaitlen et al. (2017) (7) and its
628 accompanying scripts to use the RFMix local ancestry calls to measure local ancestry
629 disequilibrium (LAD) decay in 10 Mb windows, overlapping by 1 Mb. Specifically, we started

with the first SNP on each autosome, used the 10 Mb window end point to identify the SNP closest to the inside of this boundary, and then used local ancestry calls at these positions to determine LAD. We repeated this process along each chromosome to obtain LAD in 279 autosomal windows. Using possible values of admixture generations ranging from 5-25, we determined the best fit using island-specific mean LAD decay over the 279 autosomal windows, assortative mating parameters estimated with ANCESTOR (see below), a starting autosomal admixture proportion of 0.65, and either no migration or with migration (migration rate = 0.01).

Tests for ancestry-assortative mating

We tested for ancestry-assortative mating over the last generation using ANCESTOR (15,72). ANCESTOR uses phased local ancestry tracts in the current generation to estimate the ancestral proportions of the two parents of each individual. We used the inferred parental ancestries to test for assortative mating seen as a positive correlation in ancestry between inferred mating pairs, and we used the strength of the observed positive correlation (Pearson's R for each population, using the median value from all chromosomes) as the assortative mating parameter in our application of the LAD-based method for inferring admixture timing. We also examined evidence of nonrandom mating by estimating genetic relatedness within Cabo Verde using both IBD (see below) and kinship coefficients (46).

Identification of IBD tracts and ROH

Following the analysis pipeline of S. R. Browning et al. (2018) (2), we inferred segments of IBD using the haplotype-based IBD detection method Refined IBD (73), and we estimated ancestry-

specific population size using IBDNe (1). After running Refined IBD, we used its gap-filling utility to remove gaps between segments less than 0.6 cM that had at most one discordant homozygote. We then filtered out IBD segments smaller than 5 cM, as short segments below this threshold are difficult to accurately detect (74). To visualize IBD sharing in Fig 1 and Supp Fig 3, the sum of IBD shared in each pairwise comparison of individuals was plotted by using Cytoscape 3.8 to group nodes by island, position nodes within each island according to a prefuse force-directed algorithm, and scale the color of edges based on log-transformed total IBD length (75).

We called ROH using GARLIC v.1.1.6 (49), which implements the ROH calling pipeline of Pemberton et al. (2012) (30). We performed this analysis separately for the autosomes and X chromosome. Based on the pipeline described in Szpiech et al. (2017), we used a single constant genotyping error rate of 0.001, we allowed GARLIC to automatically choose a window size for each population (--auto-winsize), and we used the resample flag to mitigate biases in allele frequency estimates caused by differing sample sizes. This resulted in GARLIC selecting a window size of 50 SNPs in each of the Cabo Verde regions and in GWD, and a window size of 40 for IBS. Using a three-component Gaussian mixture model, GARLIC classified ROH into three length groups: small/class A, medium/class B, and long/class C ROH. Across all populations, class A/B and class B/C size boundaries were inferred as approximately 300 kb and 1 Mb, respectively (see Supp Table 2 for population-specific parameters, including LOD cutoffs and size boundaries). Using only the females, we then classified ROH for the X chromosome.

672 To evaluate the influence of ROH size classification cutoffs in the results, we repeated size
673 classification using the --size-bounds flag in GARLIC to impose four additional sets of size class
674 boundaries. We first collected the population-specific length classification cut-offs from the 64
675 worldwide populations provided by Pemberton et al. (2012), which represents an extensive
676 range of classification schemes. We repeated ROH-calling with GARLIC using the minimum,
677 mean, and maximum of the 64 population-specific boundaries reported in Pemberton et al.
678 (2012) (Supp Fig 9A-C). Finally, we classified ROH based on the commonly used rule of thumb
679 that a time depth of *m* meioses is expected to give a mean IBD segment length of 100/*m* cM (76).
680 In this case, common ancestors within the past 20 generations (post-admixture) would give an
681 expected IBD segment length of approximately 2.5 Mb or longer, so we applied 2.5 Mb as the
682 shorter/long ROH boundary in Supp Fig 9D.

683 Estimation of sex bias

684 To examine sex bias, we applied a mechanistic model of sex-biased admixture (3) to infer
685 admixture parameters under a model of constant migration. We performed this analysis for all
686 563 Cabo Verdean individuals pooled together, and then repeated the analysis for each island
687 independently. We provided the model with the following parameters: grid increment size
688 (0.02), the observed autosomal and X admixture proportions (based on ADMIXTURE estimates
689 described above), the proportion of females in the sample, and the percentile cutoff for
690 parameters to keep (0.1%).

Acknowledgements

We thank Jorge Rocha for his contribution during project conception and sampling, and Greg Barsh for his support in later stages of data collection. We thank Noah Zaitlen for sharing scripts to assist in our application of the Zaitlen et al. (2017) local ancestry disequilibrium method. This work was supported by MRC-UK MR/M01987X/1 and FCT-Portugal PTDC/BIA-BDE/64044/2006 awarded to Sandra Beleza; NIH NIGMS grant R35 GM133481 awarded to Amy Goldberg; NIH NIGMS grant F32 GM139313 awarded to Katharine Korunes; and by CNPq (CSF/PDE 207651/2014-0) awarded to Giordano Bruno Soares-Souza. We acknowledge the source of the reference datasets: 1000G 30x data generated at the New York Genome Center with funds provided by NHGRI Grant 3UM1HG008901-03S1. We thank Dr. Kirk Lohmueller and three anonymous reviewers for their helpful feedback on this work. Finally, we would like to thank the Cabo Verdean participants for their invaluable contributions, and the University of Cabo Verde administration for their support.

Declaration of interests

The authors declare no competing interests.

Data and Code Availability

Sampling consent forms do not allow for public release of genotype data. Inferred local ancestry information can be found at https://doi.org/10.5281/zenodo.4021277. Code generated for this study can be found at https://github.com/agoldberglab/CaboVerde_Demographic_Analyses.

References

1. Browning SR, Browning BL. Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. Am J Hum Genet. 2015 Sep 3;97(3):404–18.

2. Browning SR, Browning BL, Daviglus ML, Durazo-Arvizu RA, Schneiderman N, Kaplan RC, et al. Ancestry-specific recent effective population size in the Americas. PLOS Genet. 2018 May 24;14(5):e1007385.

3. Goldberg A, Rosenberg NA. Beyond 2/3 and 1/3: The Complex Signatures of Sex-Biased Admixture on the X Chromosome. Genetics. 2015 Sep;201(1):263–79.

4. Loh P-R, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, et al. Inferring admixture histories of human populations using linkage disequilibrium. Genetics. 2013 Apr;193(4):1233–54.

5. Ni X, Yuan K, Liu C, Feng Q, Tian L, Ma Z, et al. MultiWaver 2.0 : modeling discrete and continuous gene flow to reconstruct complex population admixtures. Eur J Hum Genet. 2019 Jan;27(1):133–9.

6. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. Nat Genet. 2014 Aug;46(8):919–25.

7. Zaitlen N, Huntsman S, Hu D, Spear M, Eng C, Oh SS, et al. The Effects of Migration and Assortative Mating on Admixture Linkage Disequilibrium. Genetics. 2017 Jan;205(1):375–83.

8. Busby GBJ, Hellenthal G, Montinaro F, Tofanelli S, Bulayeva K, Rudan I, et al. The Role of Recent Admixture in Forming the Contemporary West Eurasian Genomic Landscape. Curr Biol. 2015 Oct 5;25(19):2518–26.

9. Fernandes V, Brucato N, Ferreira JC, Pedro N, Cavadas B, Ricaut F-X, et al. Genome-Wide Characterization of Arabian Peninsula Populations: Shedding Light on the History of a Fundamental Bridge between Continents. Mulligan C, editor. Mol Biol Evol. 2019 Mar;36(3):575–86.

10. Font-Porterias N, Arauna LR, Poveda A, Bianco E, Rebato E, Prata MJ, et al. European Roma groups show complex West Eurasian admixture footprints and a common South Asian genetic origin. PLOS Genet. 2019 Sep 23;15(9):e1008417.

11. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature. 2014 Sep 18;513(7518):409–13.

12. Skoglund P, Malmström H, Raghavan M, Storå J, Hall P, Willerslev E, et al. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. Science. 2012 Apr 27;336(6080):466–9.

13. Risch N, Choudhry S, Via M, Basu A, Sebro R, Eng C, et al. Ancestry-related assortative mating in Latino populations. Genome Biol. 2009 Nov 20;10(11):R132.

14. Sebro R, Peloso GM, Dupuis J, Risch NJ. Structured mating: Patterns and implications. PLOS Genet. 2017 Apr 6;13(4):e1006655.

15. Zou JY, Park DS, Burchard EG, Torgerson DG, Pino-Yanes M, Song YS, et al. Genetic and socioeconomic study of mate choice in Latinos reveals novel assortment patterns. Proc Natl Acad Sci U S A. 2015 Nov;112(44):13621–6.

16. Arbiza L, Gottipati S, Siepel A, Keinan A. Contrasting X-Linked and Autosomal Diversity across 14 Human Populations. Am J Hum Genet. 2014 Jun 5;94(6):827–44.

17. Bustamante CD, Ramachandran S. Evaluating signatures of sex-specific processes in the human genome. Nat Genet. 2009 Jan;41(1):8–10.

18. Keinan A, Mullikin JC, Patterson N, Reich D. Accelerated genetic drift on chromosome X during the human dispersal out of Africa. Nat Genet. 2009 Jan;41(1):66–70.

19. Ramachandran S, Rosenberg NA, Feldman MW, Wakeley J. Population differentiation and migration: Coalescence times in a two-sex island model for autosomal and X-linked loci. Theor Popul Biol. 2008 Dec 1;74(4):291–301.

20. Wilkins JF, Marlowe FW. Sex-biased migration in humans: what should we expect from genetic data? BioEssays. 2006;28(3):290–300.

21. Busby GB, Band G, Si Le Q, Jallow M, Bougama E, Mangano VD, et al. Admixture into and within sub-Saharan Africa. Pickrell JK, editor. eLife. 2016 Jun 21;5:e15266.

22. Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, et al. A Genetic Atlas of Human Admixture History. Science. 2014 Feb 14;343(6172):747–51.

23. Laso-Jadart R, Harmant C, Quach H, Zidane N, Tyler-Smith C, Mehdi Q, et al. The Genetic Legacy of the Indian Ocean Slave Trade: Recent Admixture and Post-admixture Selection in the Makranis of Pakistan. Am J Hum Genet. 2017 Dec;101(6):977–84.

24. Ruiz-Linares A, Adhikari K, Acuña-Alonzo V, Quinto-Sanchez M, Jaramillo C, Arias W, et al. Admixture in Latin America: Geographic Structure, Phenotypic Diversity and Self-Perception of Ancestry Based on 7,342 Individuals. PLOS Genet. 2014 Sep 25;10(9):e1004572.

775 25. Triska P, Soares P, Patin E, Fernandes V, Cerny V, Pereira L. Extensive Admixture and
776 Selective Pressure Across the Sahel Belt. Genome Biol Evol. 2015 Dec 1;7(12):3484–95.

777 26. Korunes KL, Goldberg A. Human genetic admixture. PLOS Genet. 2021 Mar
778 11;17(3):e1009374.

779 27. Blant A, Kwong M, Szpiech ZA, Pemberton TJ. Weighted likelihood inference of genomic
780 autozygosity patterns in dense genotype data. BMC Genomics. 2017 Dec 1;18(1):928.

781 28. Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. Genomic Runs
782 of Homozygosity Record Population History and Consanguinity. PLOS ONE. 2010 Nov
783 15;5(11):e13996.

784 29. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et
785 al. Runs of Homozygosity in European Populations. Am J Hum Genet. 2008 Sep
786 12;83(3):359–72.

787 30. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic
788 Patterns of Homozygosity in Worldwide Human Populations. Am J Hum Genet. 2012
789 Aug;91(2):275–92.

790 31. Mooney JA, Huber CD, Service S, Sul JH, Marsden CD, Zhang Z, et al. Understanding the
791 Hidden Complexity of Latin American Population Isolates. Am J Hum Genet. 2018 Nov
792 1;103(5):707–26.

793 32. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016 13;538(7624):161–
794 4.

795 33. Popejoy AB, Ritter DI, Crooks K, Currey E, Fullerton SM, Hindorff LA, et al. The clinical
796 imperative for inclusivity: Race, ethnicity, and ancestry (REA) in genomics. Hum Mutat.
797 2018;39(11):1713–20.

798 34. Landry LG, Ali N, Williams DR, Rehm HL, Bonham VL. Lack Of Diversity In Genomic
799 Databases Is A Barrier To Translating Precision Medicine Research Into Practice. Health
800 Aff Proj Hope. 2018;37(5):780–5.

801 35. Correia e Silva AL. Dinâmicas de decomposição e recomposição de espaços e sociedades.
802 In: Santos, MEM, editor. História geral de Cabo Verde. Lisbon, Portugal; Praia, Cabo
803 Verde: Instituto de Investigação Científica Tropical, Instituto Nacional de Investigação,
804 Promoção e Património Culturais de Cabo Verde; 2002. p. 1–66.

805 36. Baleno IC. Povoamento e Formação da Sociedade. In: Albuquerque, L and Santos, MEM,
806 editor. História geral de Cabo Verde. Lisbon, Portugal; Praia, Cabo Verde: Instituto de
807 Investigação Científica Tropical, Instituto Nacional de Investigação, Promoção e
808 Património Culturais de Cabo Verde; 2001. p. 125–77.

809 37. Russell-Wood A. The Portuguese Empire, 1415-1808: A World on the Move. Baltimore,
810 Maryland: Johns Hopkins University Press; 1998. 289 p.

811 38. Cabral I. Ribeira Grande: vida urbana, gente, mercancia, estagnação. In: Santos, MEM,
812 editor. História geral de Cabo Verde. Lisbon, Portugal; Praia, Cabo Verde: Instituto de
813 Investigação Científica Tropical, Instituto Nacional de Investigação, Promoção e
814 Património Culturais de Cabo Verde; 2001. p. 225–73.

815 39. Correia e Silva AL. Espaço, ecologia e economia interna. In: de Albuquerque L, Santos,
816 MEM, editor. História geral de Cabo Verde. Lisbon, Portugal; Praia, Cabo Verde: Instituto
817 de Investigação Científica Tropical, Instituto Nacional de Investigação, Promoção e
818 Património Culturais de Cabo Verde; 2001.

819 40. Cabral I. A evolução da sociedade foguense: Através de um estudo prosopográfico da elite
820 da ilha (sec. XVI-XVIII). 2012.

821 41. Barbosa, Gilda. Casamento entre primos. J Terra Nova No 246. 1997;

822 42. Beleza S, Campos J, Lopes J, Araújo II, Hoppfer Almada A, e Silva AC, et al. The
823 Admixture Structure and Genetic Variation of the Archipelago of Cape Verde and Its
824 Implications for Admixture Mapping Studies. Relethford JH, editor. PLoS ONE. 2012
825 Nov;7(11):e51103.

826 43. Beleza S, Johnson NA, Candille SI, Absher DM, Coram MA, Lopes J, et al. Genetic
827 Architecture of Skin and Eye Color in an African-European Admixed Population. Spritz
828 RA, editor. PLoS Genet. 2013 Mar;9(3):e1003372.

829 44. Fernandes AT, Velosa R, Jesus J, Carracedo A, Brehm A. Genetic differentiation of the
830 Cabo Verde archipelago population analysed by STR polymorphisms. Ann Hum Genet.
831 2003 Jul;67(Pt 4):340–7.

832 45. Verdu P, Jewett EM, Pemberton TJ, Rosenberg NA, Baptista M. Parallel Trajectories of
833 Genetic and Linguistic Admixture in a Genetically Admixed Creole Population. Curr Biol.
834 2017;27(16):2529-2535.e3.

835 46. Ochoa A, Storey JD. New kinship and FST estimates reveal higher levels of differentiation
836 in the global human population. bioRxiv. 2019 May;653279.

837 47. Reed FA, Tishkoff SA. African human diversity, origins and migrations. Curr Opin Genet
838 Dev. 2006 Dec 1;16(6):597–605.

839 48. Tishkoff SA, Williams SM. Genetic analysis of African populations: human evolution and
840 complex disease. Nat Rev Genet. 2002 Aug;3(8):611–21.

841 49. Szpiech ZA, Blant A, Pemberton TJ. GARLIC: Genomic Autozygosity Regions Likelihood-
842 based Inference and Classification. Berger B, editor. Bioinformatics. 2017 Jul;33(13):2059–
843 62.

844 50. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows
845 into population history and trait architecture. Nat Rev Genet. 2018 Apr;19(4):220–34.

846 51. Goldberg A, Günther T, Rosenberg NA, Jakobsson M. Ancient X chromosomes reveal
847 contrasting sex bias in Neolithic and Bronze Age Eurasian migrations. Proc Natl Acad Sci.
848 2017 Mar 7;114(10):2657–62.

849 52. Eltis D, Richardson D, Blight DW. Atlas of the Transatlantic Slave Trade [Internet]. Yale
850 University Press; 2010 [cited 2020 Oct 9]. Available from:
851 https://www.jstor.org/stable/j.ctt5vm1s4

852 53. Micheletti SJ, Bryc K, Esselmann SGA, Freyman WA, Moreno ME, Poznik GD, et al.
853 Genetic Consequences of the Transatlantic Slave Trade in the Americas. Am J Hum Genet.
854 2020 Aug 6;107(2):265–77.

855 54. Pero de Guimarães. Corregedor de Cabo Verde, carta de 22 de Maio de 1513. In: de
856 Albuquerque L, Santos, MEM, editor. História geral de Cabo Verde. Lisbon, Portugal;
857 Praia, Cabo Verde: Instituto de Investigação Científica Tropical, Instituto Nacional de
858 Investigação, Promoção e Património Culturais de Cabo Verde; 1991. p. 219–23.

859 55. Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR, et al.
860 Reconstructing the Population Genetic History of the Caribbean. PLOS Genet. 2013 Nov
861 14;9(11):e1003925.

862 56. Campbell CL, Palamara PF, Dubrovsky M, Botigué LR, Fellous M, Atzmon G, et al. North
863 African Jewish and non-Jewish populations form distinctive, orthogonal clusters. Proc Natl
864 Acad Sci [Internet]. 2012 Jul 31 [cited 2021 Aug 2]; Available from:
865 https://www.pnas.org/content/early/2012/07/31/1204840109

866 57. Seibert G. Crioulização em Cabo Verde e São Tomé e Príncipe: divergências históricas e
867 identitárias. Afro-Ásia. 2014 Jun 1;41–70.

868 58. Goldberg A, Rastogi A, Rosenberg NA. Assortative mating by population of origin in a
869 mechanistic model of admixture. Theor Popul Biol. 2020 Aug 1;134:129–46.

870 59. Robinson MR, Kleinman A, Graff M, Vinkhuyzen AAE, Couper D, Miller MB, et al.
871 Genetic evidence of assortative mating in humans. Nat Hum Behav. 2017 Jan 9;1(1):1–13.

872 60. Clarke L, Fairley S, Zheng-Bradley X, Streeter I, Perry E, Lowy E, et al. The international
873 Genome sample resource (IGSR): A worldwide collection of genome variation
874 incorporating the 1000 Genomes Project data. Nucleic Acids Res. 2017 Jan 4;45(D1):D854–9.

875 61. Fairley S, Lowy-Gallego E, Perry E, Flicek P. The International Genome Sample Resource
876 (IGSR) collection of open human genomic variation resources. Nucleic Acids Res. 2020 Jan
877 8;48(D1):D941–7.

878 62. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in
879 unrelated individuals. Genome Res. 2009 Sep;19(9):1655–64.

880 63. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool
881 set for whole-genome association and population-based linkage analyses. Am J Hum
882 Genet. 2007 Sep;81(3):559–75.

883 64. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease
884 and population genetic studies. Nat Methods. 2013 Jan;10(1):5–6.

885 65. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling
886 approach for rapid and robust local-ancestry inference. Am J Hum Genet. 2013
887 Aug;93(2):278–88.

888 66. Hamid I, Korunes K, Beleza S, Goldberg A. Rapid adaptation to malaria facilitated by
889 admixture in the human population of Cabo Verde [Internet]. Zenodo; 2020 [cited 2020 Sep
890 21]. Available from: https://zenodo.org/record/4021277#.X2kMuPxKhTY

891 67. Tang H, Peng J, Wang P, Risch NJ. Estimation of individual admixture: analytical and
892 study design considerations. Genet Epidemiol. 2005 May;28(4):289–301.

893 68. Tang H, Coram M, Wang P, Zhu X, Risch N. Reconstructing Genetic Ancestry Blocks in
894 Admixed Individuals. Am J Hum Genet. 2006 Jul;79(1):1–12.

895 69. Guan Y. Detecting structure of haplotypes and local ancestry. Genetics. 2014
896 Mar;196(3):625–42.

897 70. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. PLOS Genet. 2006
898 Dec 22;2(12):e190.

899 71. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal
900 components analysis corrects for stratification in genome-wide association studies. Nat
901 Genet. 2006 Aug;38(8):904–9.

902 72. Zou JY, Halperin E, Burchard E, Sankararaman S. Inferring parental genomic ancestries
903 using pooled semi-Markov processes. Bioinformatics. 2015 Jun 15;31(12):i190–6.

904 73. Browning BL, Browning SR. Improving the accuracy and efficiency of identity-by-descent
905 detection in population data. Genetics. 2013 Jun;194(2):459–71.

74. Nelson D, Kelleher J, Ragsdale AP, Moreau C, McVean G, Gravel S. Accounting for long-range correlations in genome-wide simulations of large cohorts. PLOS Genet. 2020 May 5;16(5):e1008619.

75. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003 Nov;13(11):2498–504.

76. Thompson EA. Identity by Descent: Variation in Meiosis, Across Genomes, and in Populations. Genetics. 2013 Jun;194(2):301–26.

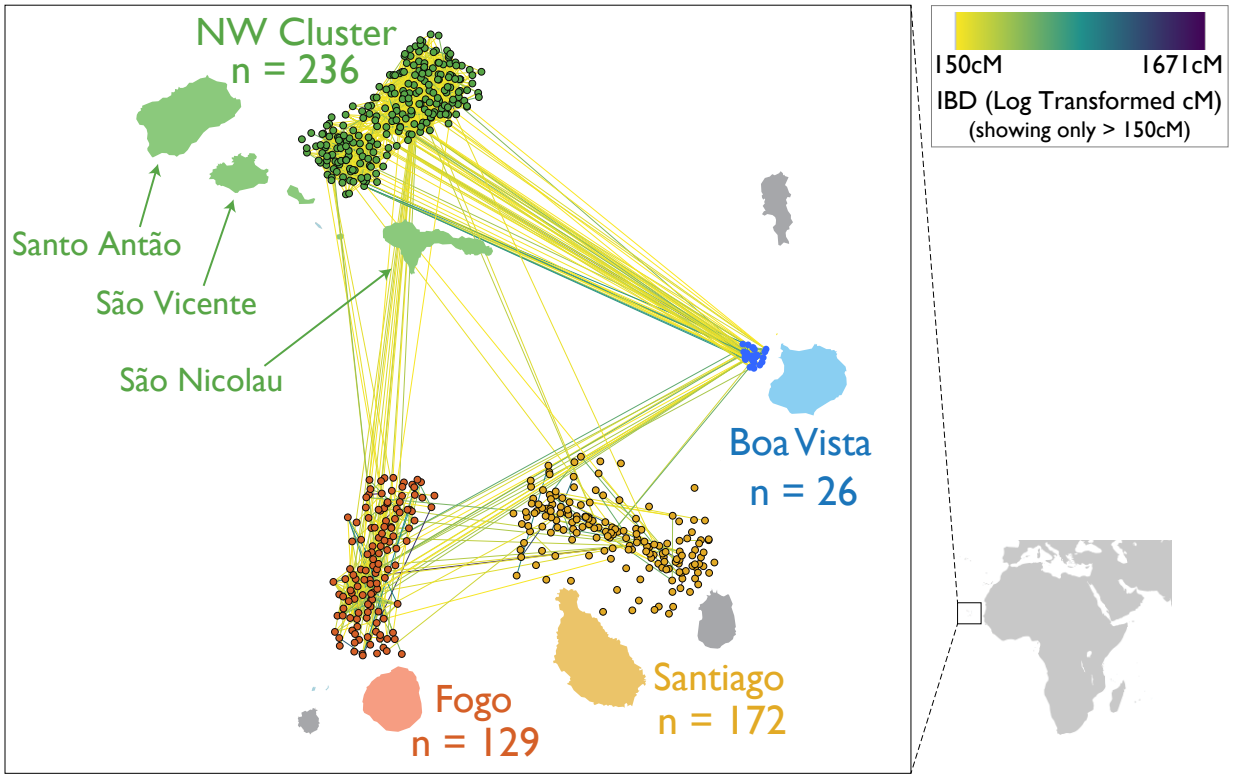

Fig 1 | Shared ancestry between and within the islands in the context of geography. Each island has a corresponding cluster of nodes representing all sampled individuals, with the individuals localized to be adjacent to the island where they were sampled. Node placement within islands is determined by a force-directed algorithm using pairwise shared IBD, meaning that the spread of each cluster reflects the level of relatedness in each population. Edges between the nodes represent total IBD tract length for pairs of individuals sharing more than 150 cM of total IBD, with edges colored using the log transformed total IBD length.

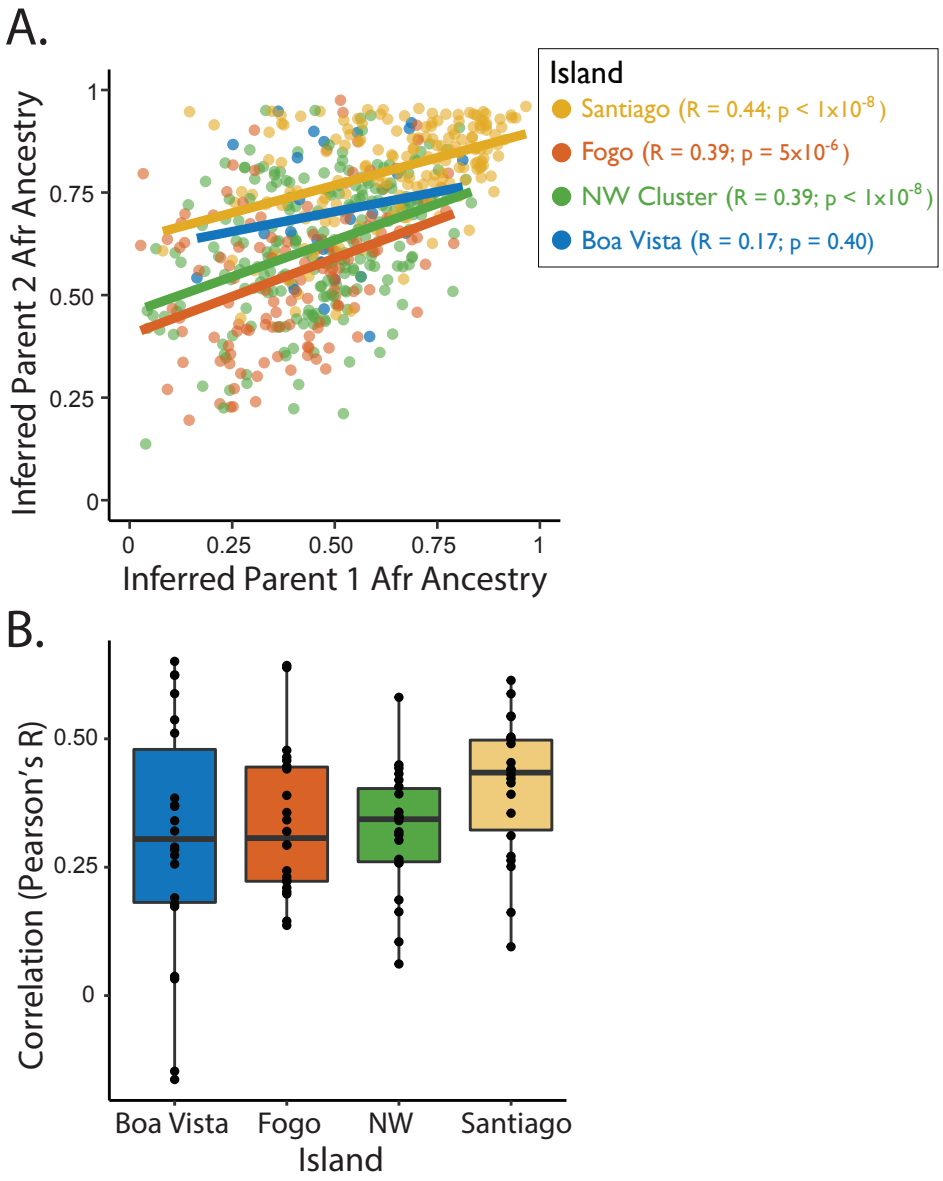

Fig 2 | Correlation between inferred parental ancestries. (A) For each Cabo Verdean individual, the inferred Parent 1 vs Parent 2 ancestry proportions are shown for an example chromosome (Chromosome 7), colored by island. (B) shows the set of correlation coefficients between inferred parental ancestries for each autosomal chromosome.

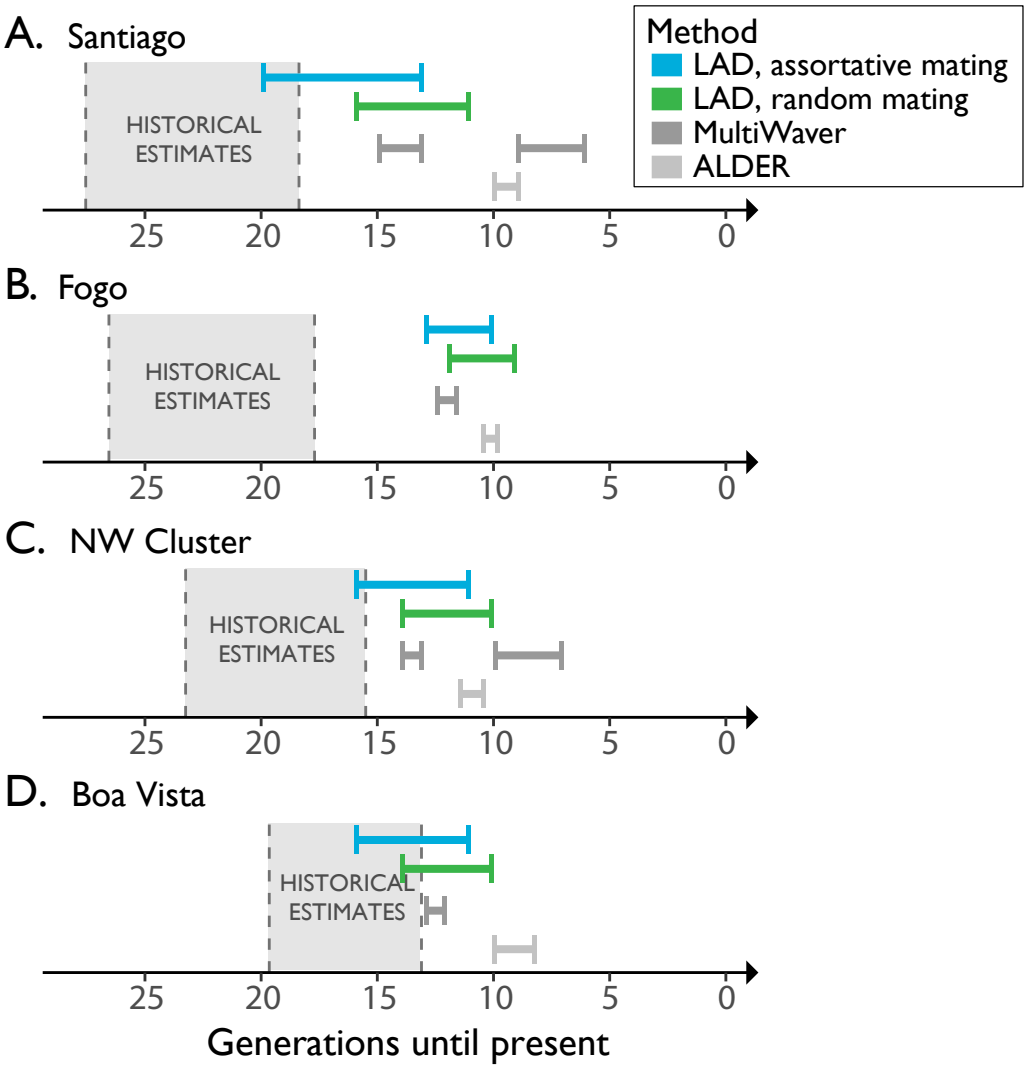

Fig 3 | Timelines of inferred generations of admixture for each island of Cabo Verde. The island regions are listed in order of known settlement timing beginning with Santiago (A), followed by Fogo (B), and then the later migrations to the Northwest Cluster (C) and Boa Vista (D). For each island, admixture timing inferred with population genetic methods are shown in comparison to historical records. Historical estimates of when each island was first settled are shown in generations, based on a range of generation times (20-30 year generation time). Within each LAD-based estimate of admixture timing, the presented interval reflects the range of estimates generated under the assumption of constant migration rate ($m = 0.01$; yielding older estimates) to the assumption of no migration ($m = 0$; yielding more recent estimates). MultiWaver results show the admixture generations inferred under the admixture model selected by MultiWaver, with 0.95 CI.

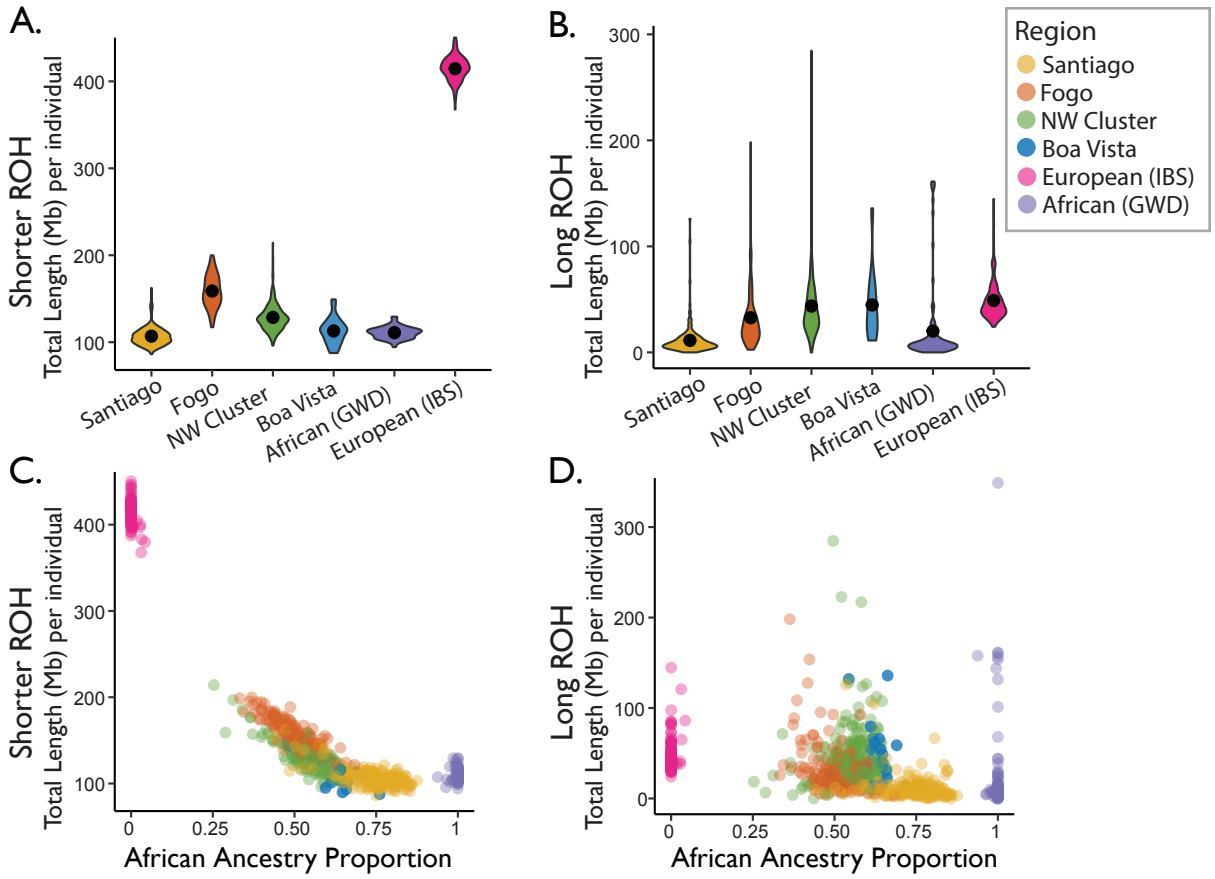

Fig 4 | ROH content by island and in the context of ancestry. (A-B) Violin plots show the population-specific distributions of the total (summed over each genome) length of autosomal ROH per individual. Solid black dots represent the within-population means. (C-D) Total length of autosomal ROH per individual plotted against West African ancestry proportions and colored by population.

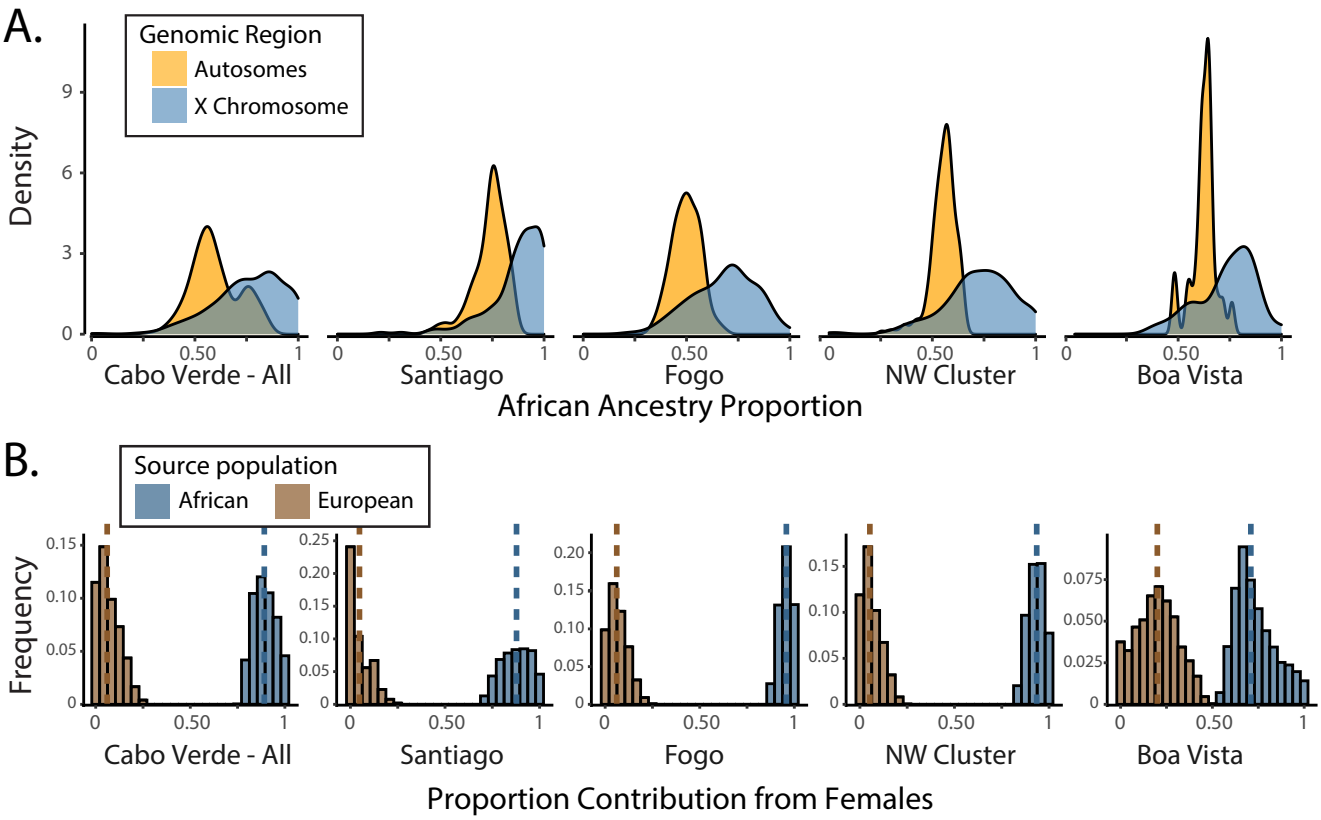

Fig 5 | Sex-biased admixture in Cabo Verde. (A) The distribution of West African ancestry proportion on the autosomes and X chromosome for each of the island regions, estimated with ADMIXTURE. (B) Under a model of constant admixture over time, the fraction of the total contribution of genetic material originating from females for West African and European source populations. Here, we show the distribution of parameter sets for the smallest 0.1% of Euclidean distances between the model-predicted and observed X and autosomal ancestry from a grid of possible parameter values. The range of sex-specific contributions from West African and European source populations that produce ancestry estimates closest to those observed in Cabo Verde are shown for Cabo Verde as a whole (left), and then broken down by region, with medians (dashed lines).

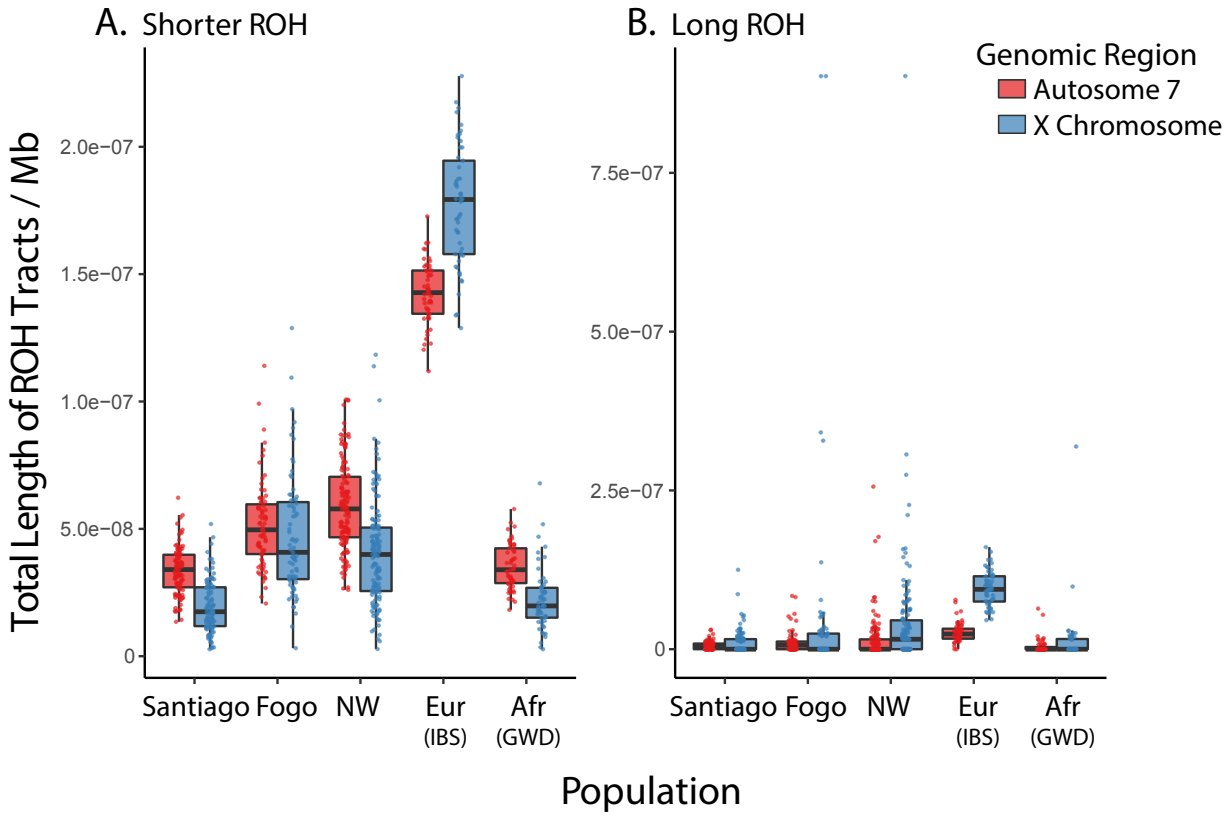

Fig 6 | Autosomal (chromosome 7) vs X chromosome distributions of ROH by class. (A) shows the population-specific distributions of the total length of shorter ROH per individual on the X chromosome compared to an autosome. Chromosome 7 was chosen as the autosomal point of comparison, given that it is the autosome most similar in size to the X chromosome. Totals are shown for females only, so that the same samples are being compared across the X chromosome and chromosome 7. Totals are plotted per Mb to account for slight differences in chromosome lengths. (B) shows the same information for long ROH.