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

ORCIDs
Katharine L Korunes: 0000-0002-2648-4707
Giordano Bruno Soares-Souza: 0000-0001-6590-5589
Katherine Bobrek: 0000-0003-0936-8779
Hua Tang: 0000-0002-0177-8864
Isabel Inês Araújo: 0000-0001-7766-1059
Amy Goldberg: 0000-0001-9306-1539
Sandra Beleza: 0000-0002-7004-3431

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.

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[8,11,21–25].

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]. Recent work has begun to assess the implications of these patterns, including effects on common demographic inference strategies and on genetic/phenotypic variation[7,14,24]. One genetic consequence of assortative mating is an increase in long runs of homozygosity (ROH), which arise when long tracts of homozygous genotypes occur due to inheritance of identical by descent (IBD) haplotypes from recent common ancestors. Across modern human populations, ROH are common and typically range in length from kilobases to megabases[26–29]. In the field of population genetics, we are just beginning to learn about ROH dynamics and expectations in admixed populations, which do not necessarily fit expectations built from homogenous populations[29,30]. For example, we know that distinct length classes of ROH reflect demographic factors from different timescales, but as we explore here, it is unclear what these distributions of ROH mean in populations that have undergone major demographic transitions such as recent admixture. Patterns of genetic variation can also be shaped by differences in male and female demographic processes. In many human populations, population sizes and migration rates differ between males and females, leading to differences in ancestry patterns on the X chromosome vs the autosomes[16–20].

63 Here, we use the admixed population of Cabo Verde as a case study to understand the patterns
64 of genetic variation left by multifaceted social and demographic processes in humans. Admixed
65 populations of African ancestry, such as the population of Cabo Verde, are often excluded in the
66 context of medical genetics and human evolutionary genetics, despite their importance and
67 widespread global presence[31–33]. We consider patterns of genetic variation alongside historical
68 records that document many aspects of Cabo Verdean history, including settlement patterns,
69 timing, and sociocultural dynamics. Specifically, we use distributions of genetic ancestry to test
70 for patterns of nonrandom mating and sex-specific gene flow. We examine the consequences of
71 these processes for genetic variation, such as patterns of homozygosity, and for common
72 demographic inference methods. In turn, by elucidating the patterns of genetic variation social
73 processes leave in human genomes for a population with a well-studied historical and
74 ethnographic record, we can better use population-genomic methods to explore the history of
75 populations without extensive records.

76 Cabo Verdean history

77 The recently admixed human population of Cabo Verde presents a particularly relevant
78 opportunity to examine how sex-biased admixture and nonrandom mating shape genetic
79 patterns and influence demographic inference. Cabo Verde is an archipelago off the coast of
80 Senegal, inhabited today by admixed descendants of Portuguese colonizers and enslaved West
81 African people who settled the unpopulated islands beginning in the mid-1400s[34–37]. Since the
82 archipelago was unoccupied prior to the 1460s, we have historical knowledge of the start of
83 admixture and the contributing source populations, with written records clearly documenting

arrival times by island. Additionally, the island geography imposes population structure within Cabo Verde, and historical records document differences in population sizes, mating patterns, and social customs by island population (see Methods for historical data sources).

The settlement of the islands was influenced by island geography and ecology, and is often divided into three different temporal stages, which are associated with changes in the economy[38]. The initial settlement stage (beginning in the 1460s) corresponds to the peopling of the Southern island Santiago, followed by the nearby island of Fogo 20 years later (Fig 1). During this time, the economy was centered around large cotton plantations and expanding trade with the Senegambian coast. This trade included the exportation of great numbers of enslaved Africans to Cabo Verde, the majority of which became part of the commercial trade to the Americas[39,40]. In the 17th century, following the decline of the slavery-based economy of Santiago and Fogo, many free farmers seeking better farming conditions migrated to the northwestern islands of Santo Antão and São Nicolau. These migrations comprised the second settlement stage. In contrast to Santiago and Fogo, these northwestern islands did not become populous economic centers[38]. During the third and final settlement stage, São Vicente in the northwestern group of islands was colonized by settlers from the neighboring islands in the mid 19th century. With the opening of Mindelo's harbor at that time, São Vicente prospered from the advent of commercial Atlantic shipping, becoming the second most important island of the archipelago after Santiago in terms of population size. In the 19th century, the eastern islands Sal, Boa Vista, and Maio were also open to English and North American ships, but these islands never became densely populated[38,40].

Previous investigations of genetic ancestry in Cabo Verde[34,35,37] have shown extensive admixture in the archipelago and differences in the mean African genetic ancestry across islands, which may reflect the differences in the settlement history and suggest restricted gene flow between islands. Using genome-wide SNP data from 563 individuals spanning the three settlement stages of Cabo Verde, we integrate several population genomic approaches to infer the sex-specific demographic history of the last ~20 generations since founding, and consider the potential biases of nonrandom mating on common inference methods. Overall, we provide insights into the population history of Cabo Verde and demonstrate how admixed populations can provide powerful test cases for understanding evolutionary processes on short timescales.

Methods

Study population and ancestry reference panels

We used genotype data from Beleza et al. (2013)[35], 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). In accordance with historical records and previous work showing that Iberian and Senegambian populations are suitable proxies for the ancestry sources of Cabo Verde[37], 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[41,42]. 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)[38,43], Cabral (2001; 2012)[44,45] and Baleno (2001)[39]. 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)[45] and Barbosa (1997)[46] for evidence of the structure of the Fogo's society and for evidence of first-cousin marriages within Fogo[46].

Characterization of ancestry

To produce estimates of admixture proportions, we first performed unsupervised clustering of the samples using ADMIXTURE[47]. ADMIXTURE was run separately for the autosomal and X chromosome datasets after pruning based on linkage disequilibrium (LD) using the indep-

145 pairwise option of PLINK v1.9 with a 50-SNP sliding window incremented by 10 SNPs, and an
146 LD threshold of r2 = 0.5[48]. This pruned dataset included 514,551 autosomal and 12,706 X
147 chromosome SNPs. We estimated individual ancestries by averaging over ten independent
148 unsupervised ADMIXTURE runs using K = 2, given the historical and genetic support for two
149 source populations[35,37]. Using the LD-pruned dataset, we also visualized the data using PLINK's
150 principal component analysis (--pca) function, confirming that the samples cluster
151 approximately based on their geographic memberships, and that West African vs European
152 ancestry clearly separates on the first principal component (Supp Fig 1). We phased all samples
153 using SHAPEIT2[49]. After running SHAPEIT -check to exclude sites not contained within the
154 reference map, we ran SHAPEIT to yield phased genotypes at 881,279 autosomal SNPs and
155 20,793 X chromosome SNPs. We then called local ancestry with RFMix v1.5.4[50] PopPhased
156 under a two-way admixture model using the West African and European reference genotypes
157 described above. These local ancestry calls are publicly available via Zenodo[51]. Ancestry
158 proportions estimated with ADMIXTURE and RFMix here are highly correlated with results
159 from Beleza et al. (2013)[35], using the same genotype data but with different software (*frappe*[52]
160 and SABER[53]) and an older reference dataset (Supp Fig 2).

161 Inference of admixture timing

162 We applied three distinct strategies for estimating the timing of the onset of admixture in Cabo
163 Verde: ALDER[4], MultiWaver 2.0[5], and a method based on patterns of linkage disequilibrium
164 between local ancestry tracts [7]. We first converted the genotypes to EIGENSTRAT format using
165 *convertf*[54,55]. We then used ALDER to date admixture timing on each island using default

166 parameters with mindis = 0.005 and source populations GWD and IBS. We ran MultiWaver
167 using the ancestry tracts inferred by RFMix, default parameters, and 100 bootstraps. Finally, we
168 followed the pipeline of Zaitlen et al. (2017)[7] and its accompanying scripts to use the RFMix
169 local ancestry calls to measure local ancestry disequilibrium (LAD) decay in 10 Mb windows,
170 overlapping by 1 Mb. Specifically, we started with the first SNP on each autosome, used the 10
171 Mb window end point to identify the SNP closest to the inside of this boundary, and then used
172 local ancestry calls at these positions to determine LAD. We repeated this process along each
173 chromosome to obtain LAD in 279 autosomal windows. Using possible values of admixture
174 generations ranging from 5-25, we determined the best fit using island-specific mean LAD
175 decay over the 279 autosomal windows, assortative mating parameters estimated with
176 ANCESTOR (see below), a starting autosomal admixture proportion of 0.65, and either no
177 migration or with migration (migration rate = 0.01).

178 Tests for assortative mating

179 We tested for assortative mating over the last generation using ANCESTOR[15,56]. ANCESTOR
180 uses phased local ancestry tracts in the current generation to estimate the ancestral proportions
181 of the two parents of each individual. We used the inferred parental ancestries to test for
182 assortative mating seen as a positive correlation in ancestry between inferred mating pairs, and
183 we used the strength of the observed positive correlation (Pearson's R for each population,
184 using the median value from all chromosomes) as the assortative mating parameter in our
185 application of the LAD-based method for inferring admixture timing. We also examined

186 evidence of nonrandom mating by estimating genetic relatedness within Cabo Verde using both
187 IBD (see below) and kinship coefficients[57].

188 Identification of IBD tracts and ROH

189 Following the analysis pipeline of S. R. Browning et al. (2018)[2], we inferred segments of IBD
190 using the haplotype-based IBD detection method Refined IBD[58], and we estimated ancestry-
191 specific population size using IBDNe[1]. After running Refined IBD, we used its gap-filling utility
192 to remove gaps between segments less than 0.6 cM that had at most one discordant
193 homozygote. We then filtered out IBD segments smaller than 5 cM, as short segments below
194 this threshold are difficult to accurately detect[59]. To visualize IBD sharing in Fig 1 and Supp Fig
195 3, the sum of IBD shared in each pairwise comparison of individuals was plotted by using
196 Cytoscape 3.8 to group nodes by island, position nodes within each island according to a
197 prefuse force-directed algorithm, and scale the color of edges based on log-transformed total
198 IBD length[60].
199
200 We called ROH using GARLIC v.1.1.6[61], which implements the ROH calling pipeline of
201 Pemberton et al. (2012)[29]. We performed this analysis separately for the autosomes and X
202 chromosome. Based on the pipeline described in Szpiech et al. (2017), we used a single constant
203 genotyping error rate of 0.001, we allowed GARLIC to automatically choose a window size for
204 each population (--auto-winsize), and we used the resample flag to mitigate biases in allele
205 frequency estimates caused by differing sample sizes. This resulted in GARLIC selecting a
206 window size of 50 SNPs in each of the Cabo Verde regions and in GWD, and a window size of

207 40 for IBS. Using a three-component Gaussian mixture model, GARLIC classified ROH into
208 three length groups: small/class A, medium/class B, and long/class C ROH. Across all
209 populations, Class A/B and class B/C size boundaries were inferred as approximately 300 kb
210 and 1 Mb, respectively (see Supp Table 1 for population-specific parameters, including LOD
211 cutoffs and size boundaries). Using only the females, we then classified ROH for the X
212 chromosome.

213 Estimation of sex bias

214 To examine sex bias, we applied a mechanistic model of sex-biased admixture[3] to infer
215 admixture parameters under a model of constant migration. We performed this analysis for all
216 563 Cabo Verdean individuals pooled together, and then repeated the analysis for each island
217 independently. We provided the model with the following parameters: grid increment size
218 (0.02), the observed autosomal and X admixture proportions (based on ADMIXTURE estimates
219 described above), the proportion of females in the sample, and the percentile cutoff for
220 parameters to keep (0.1%).

221 Results

222 Patterns of shared ancestry reflect the colonization history of the islands

223 Using genome-wide SNP data from 563 individuals, we examined four island regions of Cabo
224 Verde (Fig 1), which differ in their settlement histories. The partitioning of the considered island

225 regions was also supported by genetic patterns. Namely, these regions showed quantitatively
226 distinct distributions of IBD tracts, as we explore below, and were supported by clustering
227 patterns in PCA (Supp Fig 1), in which the three northwestern islands grouped together (and
228 with the eastern island Boa Vista) despite the different settlement time of São Vicente Island.
229
230 Global ancestry proportions reflect extensive European and West African admixture and concur
231 with previous results[35] (Supp Fig 2), with the median West African autosomal ancestry in Cabo
232 Verdeans varying by island from 50% in Fogo, 56% in the northwestern islands in aggregate
233 (the Northwest Cluster), 63% in Boa Vista, and 75% in Santiago. These inter-island differences
234 alongside historical knowledge provide a comparative dataset for understanding how ancestry
235 patterns can be used to infer demographic processes.
236
237 The process of admixture can influence measures such as IBD and ROH that are often used to
238 inform inference of population history, and theoretical expectations for these measures are less
239 clear for admixed populations compared to homogeneous populations. Thus, we use multiple
240 methods to examine relatedness in Cabo Verde, and we use this case study to underscore the
241 need for further empirical and theoretical work to understand the dynamics of IBD and ROH in
242 admixed populations. Using IBD to examine relatedness between and within regions of Cabo
243 Verde (Fig 1), we found that patterns of shared ancestry reflect the successive settlement history
244 of the islands. Patterns of IBD within and between populations provide opportunities to
245 examine common ancestry based on the number and sizes of segments of IBD. Notably,
246 Santiago has the lowest mean number and total length of IBD segments between individuals

247 (Supp Fig 4-5). In contrast, we observe the highest levels of IBD within and between the
248 Northwest Cluster and Boa Vista.
249
250 Focusing on pairs of individuals with relatively high levels of IBD, we observe that recent
251 relatives both within and across islands are common (Fig 1; see Supp Fig 3 for lower levels of
252 IBD and within-island networks). Notably, relatedness parameters estimated under simple
253 models of structure likely do not hold in many human populations, such as recently admixed
254 populations. Kinship estimated under a framework[57] specifically designed for arbitrary
255 population structures (Supp Fig 6A) is consistent with both the overall trends in IBD tracts
256 (Supp Fig 4-5), and supports the presence of a subset of very closely related individuals within
257 the data. We observed that the ordering of the global ancestry proportions according to a
258 kinship matrix ordered using seriation, which places low kinship values away from the
259 diagonal[57] (Supp Fig 6), shows that Cabo Verdeans with greater European ancestry tend to
260 share greater kinship compared to individuals with greater West African ancestry.
261
262 The lower levels of IBD (Fig 1, Supp Fig 3-5) and lower kinship estimates (Supp Fig 6) observed
263 in Santiago are consistent with expectations based on the island's history, as Santiago was the
264 first Cabo Verdean island to be founded and has the largest population size. Santiago is also
265 likely to have relatively high genetic diversity as a result of having the highest proportion of
266 West African ancestry, based on the expectation that kinship in human populations generally
267 reflects serial bottlenecks due to dispersal from Africa[62,63]. In contrast, the high levels of IBD
268 within and between the Northwest Cluster and Boa Vista reflect well-documented serial

269 founding and migrations during the settlement of the islands. In general, kinship estimates also
270 reflect the colonization history of the islands, with Cabo Verdeans of greater European ancestry
271 sharing greater kinship compared to individuals with greater West African ancestry (Supp Fig
272 6). Exceptions to this trend were most noticeable in individuals from Fogo and the Northwest
273 Cluster, where it was most apparent that some observed relatedness patterns do not result in
274 smooth ordering of global ancestry proportions (Supp Fig 6B). This observation led us to
275 hypothesize that other sociocultural processes beyond proportions of ancestry, such as
276 nonrandom mating, may drive patterns of relatedness in Cabo Verde.

277 Nonrandom mating shapes patterns of ancestry and influences
278 demographic inference

279 To test for ancestry assortative mating within Cabo Verde, we examined whether the genomic
280 ancestries of individuals in mating pairs correlate with each other. To this end, we applied
281 ANCESTOR to computationally infer the parental ancestry proportions that likely preceded the
282 ancestry haplotypes we observe today. We found that the inferred ancestries of mating pairs in
283 the previous generation are significantly positively correlated (Fig 2).
284
285 We next considered how the observed genetic evidence of assortative mating can be leveraged
286 in the context of inferring population history. To estimate the timing of the onset of admixture,
287 we first applied two strategies, ALDER[4] and MultiWaver 2.0[5], that are not intended to account
288 for assortative mating. ALDER uses the extent of LD decay among neighboring loci to infer

289 mixture proportions and dates. In contrast, MultiWaver uses the length distribution of ancestral
290 tracks. Assuming a generation time of 25 years, decay rates estimated with ALDER suggest
291 admixture timing in the mid to late 1700s (Fig 3). MultiWaver chose a multi-wave admixture
292 model for Santiago (Fig 3A) and the Northwest Cluster (Fig 3C), but selected an isolation model
293 for Fogo (Fig 3B) and Boa Vista (Fig 3D). To incorporate nonrandom mating, we applied the
294 method of Zaitlen et al. (2017)[7], which describes the decay of local ancestry disequilibrium
295 (LAD) as a function of assortative mating strength, migration rate, recombination rate, and the
296 number of generations since admixture began. Results from the LAD-based method described
297 in Zaitlen et al. (2017)[7] estimated more generations of admixture than ALDER or MultiWaver,
298 with older estimates of admixture timing under a model including both assortative mating and
299 constant migration.
300
301 The estimates of admixture timing from ALDER and MultiWaver are noticeably more recent
302 than the known founding of Cabo Verde in late 1400s. Results from the LAD-based method
303 described in Zaitlen et al. (2017)[7] align more closely with historical records (Fig 3), particularly
304 when we allowed for both assortative mating and constant migration. These results suggest that
305 explicitly accounting for nonrandom mating improves estimates of admixture timing, while the
306 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[29,61]. This method classifies ROH into three categories (short, medium, and long) based on the length distributions in each population. 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 ROH within Cabo Verde, partitioning the distributions into shorter ROH (short and medium, as called with GARLIC) and long ROH (Fig 4A-B). While ROH in homogeneous populations are often modeled as three length classes, we observe that pre- and post-admixture dynamics in Cabo Verde are best captured by focusing on long ROH vs all shorter ROH lengths. Long ROH primarily reflects post-admixture dynamics such as recent parental relatedness in the admixed population, whereas shorter ROH reflects primarily pre-admixture homozygosity of older haplotypes and background relatedness from the source populations.

Since methods for detecting and interpreting ROH have been developed primarily in homogeneous populations[29,64], 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 7). Ancestry-specific population sizes in Cabo Verde suggest bottlenecks of both source

328 populations, followed by population expansion within the past 10 generations[1,2]. If the length of
329 ROH in an individual's genome is driven by distance from Africa, as expected based on
330 previous work[28,29], we predict the total length of ROH per genome to negatively correlate with
331 West African admixture proportion. Consistent with this expectation, we observe a negative
332 correlation between the total ROH and West African ancestry proportions in Santiago, Fogo,
333 and the Northwest Cluster (Supp Fig 8A). Breaking total ROH into shorter and long ROH, we
334 see that this negative correlation is driven primarily by shorter ROH (Fig 4; Supp Fig 8B-C).
335
336 Despite recent colonization bottlenecks, we observed that some admixed individuals in Cabo
337 Verde present lower levels of ROH than either of the source population proxies. Comparing the
338 Cabo Verdean islands to each other, we observed the lowest amounts of all length classes of
339 ROH in Santiago (pairwise Mann-Whitney U tests $p < 1 \times 10^{-8}$). The low levels of all length
340 classes of ROH in Santiago are consistent with several historical attributes of the islands,
341 including serial founding beginning with the settlement of Santiago. Serial founding resulting
342 in bottlenecks is expected to increase shorter ROH by reducing the number of ancestral
343 lineages. Because shorter ROH reflects background relatedness from the source populations,
344 shorter ROH are expected to be more sensitive to the contributions of the source populations.
345 Our classification of ROH into shorter and long ROH (Fig 4) revealed that Santiago has
346 significantly lower levels of shorter ROH compared to even the West African reference
347 population (Mann-Whitney U $p = 5.87 \times 10^{-7}$), while long ROH are not depleted (Mann-Whitney
348 U $p = 0.841$). This observation that Santiago has significantly lower levels of shorter ROH
349 compared to even the West African reference is perhaps surprising given the canonical view of

African source populations as generically having the lowest ROH in worldwide samples. This is consistent with the idea that shorter ROH are contributed to the population from older shared ancestors in the source populations, and these tracts are often interrupted with local ancestry from another source population upon admixture. In contrast, long ROH are likely driven by post-admixture dynamics such as small population size and mating preferences, and long ROH tracts may span multiple local ancestries.

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 homogeneous 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. 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.

367 **Ancestry patterns on the X chromosome vs the autosomes reflect sex-**

368 **biased demographic processes**

369 On all islands, autosomal versus X chromosome ancestry patterns suggest that male and female
370 contributions differ significantly by source population (Fig 5A; Wilcoxon Signed-Rank Test $p <$
371 1×10^{-8} for Santiago, Fogo, and the Northwest Cluster; $p = 0.0013$ for Boa Vista). To
372 quantitatively explore these differences in male vs female demographic history, we used a
373 model-based approach to estimate differences between male and female contributions[3]. A
374 model of constant admixture is supported over a model of instantaneous admixture by
375 historical records and by our comparison of admixture timing estimates (above). Using the
376 method of Goldberg and Rosenberg for a model of constant admixture, we compared observed
377 and expected X-chromosomal and autosomal ancestry over a grid of possible parameter values.
378 Following the implementation from Goldberg et al. (2017)[65], Fig 5B presents results from the
379 0.1% of parameter sets closest to observed data based on a Euclidean distance between observed
380 population mean ancestries on the X and autosomes and expected ancestries under the model
381 (Supp Fig 9 shows the same data represented by plotting the male contributions). These
382 estimated sex-specific contributions from West African and European source populations under
383 a model of constant admixture suggest greater contributions from West African women and
384 European men.

385

386 Given our observation that source populations make distinct contributions to ROH in Cabo
387 Verde, we next investigate differences in autosomal vs X chromosome ROH content. Differences

388 in autosomal vs X chromosome ROH may arise in part due to the source populations
389 contributing in a sex-biased manner. Notably, this effect would be seen most in shorter ROH,
390 since shorter ROH reflect the homozygosity of older haplotypes and background relatedness
391 from the source populations. However, it is challenging to disentangle processes shaping X vs
392 autosomal ROH, due to the smaller effective population size of the X chromosome. Our results
393 suggest that sex-biased admixture processes in Cabo Verde are reflected in ROH, with lower
394 levels of shorter ROH on the X chromosome than on autosomes (Fig 6A). European individuals
395 have higher levels of ROH than West African individuals, and the higher contributions of
396 African X chromosomes (vs European X chromosomes) may drive the lower levels of short
397 ROH in Cabo Verde on the X chromosome vs the autosomes (Fig 6A). Long ROH reflects
398 different dynamics, potentially including the reduced post-admixture population size and other
399 sex-specific processes.

400 Discussion

401 In this study, we leveraged patterns of genetic variation in Cabo Verde to infer the demographic
402 history of the past ~20 generations. We found that distinct genetic patterns of four island
403 regions within the archipelago reflect the colonization history of the islands, including island-
404 specific settlement timing, admixture dynamics, mating patterns, and sex-biased demography.
405 Together these results demonstrate how patterns of ancestry and genetic variation are shaped
406 by social and demographic forces on short timescales. By better understanding how complex

population histories generate genetic variation, we can improve interpretation of inference from populations without historical records.

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

The observed island-specific genetic patterns reflect the complex ecological and social factors that shaped the settlement of Cabo Verde. Historical evidence suggests that the islands were founded in a stepping-stone pattern, with settlement stages that were accompanied by important economic and sociocultural shifts[38]. Early written records of the population (described below) suggest that admixture began early during the settlement of Cabo Verde, despite the highly racially stratified, slavery-based system that characterized the first settlement stage[44]. The second and third stages were carried out by mostly already-admixed individuals who had become a significant group within Cabo Verdean society and who migrated from the southern to the northern islands[43]. We found that the staggered settlement history and the island-specific population dynamics shaped patterns of ancestry and genetic variation within Cabo Verde.

Admixture in Cabo Verde is consistent with a model of continuous gene flow from Europe and 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[66,67]. PCA (Supp Fig 1), IBD (Fig 1, Supp Fig 3-5), and kinship (Supp Fig 6) are

427 consistent with genetic drift occurring during the archipelago's settlement history of
428 consecutive founder effects and subsequent relative isolation of the islands. Specifically,
429 individuals are genetically more related within an island than among islands, and considerably
430 more individuals from the northern islands and Fogo are connected to each other by high IBD
431 pairwise connections (Supp Fig 4-5). We observed differences in IBD that reflect the
432 contributions of the source populations, such as Santiago having both the highest West African
433 ancestry proportions and the lowest overall levels of IBD sharing. The higher African genomic
434 ancestry on Santiago is consistent with previous results[34,35], and indicates differences in
435 settlement patterns despite the fact that colonization of the other islands involved migrations
436 from Santiago and was based on the same slavery-based system.
437
438 Using a mechanistic model of sex-biased admixture[3], we used differences in ancestry on the X
439 chromosome vs the autosomes to examine sex-biased migration in Cabo Verde. We found that
440 admixture occurred primarily through the mating of European males and African females,
441 consistent with historical work that documents sex bias during the settlement of Cabo Verde.
442 First records of the demographic distribution in Cabo Verde are found in a letter by a judicial
443 official Pero de Guimarães to the Portuguese King in 1513, which stated the presence of 118
444 European individuals in Santiago, of which only four were (single) women[68]. The marginal
445 presence of European females throughout the founding of Cabo Verde contributed to the
446 extensive genetic admixture currently found in the archipelago. The distribution of ROH in
447 autosomes versus X chromosome is influenced by initial male versus female contributions to the
448 admixed population's gene pool. Throughout the islands, the higher contributions of African X

449 chromosomes (vs European X chromosomes) may drive the lower levels of short ROH in Cabo
450 Verde on the X chromosome vs the autosomes (Fig 6A). These results underscore the long-
451 lasting genetic impacts of the trans-Atlantic slave trade, as has been recently shown in admixed
452 populations across the Americas[67].
453
454 Despite only a few hundred years of unique population histories among the islands, we also
455 found that patterns of genetic variation reflect island-specific patterns in post-admixture
456 dynamics. Observed patterns of IBD, kinship, and long ROH are all consistent with overall high
457 levels of relatedness within Cabo Verde, particularly in Fogo and the northwestern islands.
458 These patterns of relatedness may be shaped by both founder effects and by post-admixture
459 dynamics such as nonrandom mating keeping similar haplotypes together rather than
460 distributing them randomly throughout mating pairs in the population. Indeed, we inferred a
461 positive correlation between mating pairs of the previous generation, suggesting assortative
462 mating on the short timescale since the founding of the islands. Mating patterns and sex-biased
463 admixture are distinct—but interacting—processes in the history of Cabo Verde. Sex-biased
464 migration determined the gene pool within Cabo Verde, while patterns of nonrandom mating
465 within the islands shaped subsequent generations of admixture. These mating patterns were
466 tied to social and economic structure. European enslavers sexually coerced enslaved African
467 women during slavery. After the end of slavery, racially unbalanced control of landownership
468 and higher social status of Europeans resulted in continued socioeconomic pressures for African
469 women to have children with European men. Such unions often resulted in improved
470 socioeconomic status for both the women and their children. For example, the first admixed

individual to assume an important position within the administration of Santiago was recorded in 1546[69]. Over generations, Cabo Verde became a population of extensively admixed ancestry, with mating patterns shaped by the interconnected effects of social position and skin color. This type of nonrandom mating may have been more prominent in the second half of the 17th century, due to the large exodus of the white population following the decay of the slave trade. With this population shift, historical sources say that the socioeconomic elite of the islands became predominately the "white children of the islands"[44], leading to further racial segregation. This historical evidence of nonrandom mating patterns—together with our genetic observations—demonstrates how quickly changes in social structure can impact population homozygosity and differentiation.

The island of Fogo stands out as having unique social and historical processes, even though it was settled shortly after Santiago. Fogo's society was, since its origins, a conservative rural society whose main economic activity was to produce goods to trade in the African coast. The socioeconomic elite was patriarchal and aristocratic[45], and composed primarily of related families, which promoted first-cousin marriages[46]. The sustenance of this class depended on land ownership, which allowed the slavery-based system to be perpetuated longer in Fogo than in Santiago[38,45]. 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

493 greater levels of local ancestry disequilibrium than the other islands. Higher local ancestry
494 disequilibrium may be driven by assortative mating. Additionally, it may be that Fogo has
495 experienced stronger founder effects, which would decrease the number of ancestral lineages.
496 Indeed, the higher levels of shorter ROH within Fogo (Fig 4) are consistent with founder effects
497 increasing background relatedness and thus increasing shorter ROH, though we emphasize the
498 need for further work to understand the dynamics of ROH in admixed populations.

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

501 The admixture dynamics, founder effects, and mating patterns within Cabo Verde shape
502 summaries of genetic variation that are often used to inform demographic inference, such as
503 local ancestry disequilibrium, IBD, and ROH. To investigate the effects of these population
504 dynamics on demographic inference, we estimated admixture timing using multiple population
505 genetic tools. Estimates of admixture timing from genetic data were most concordant with
506 historical records when using inference based on local ancestry disequilibrium[7], allowing for
507 assortative mating and migration. Under this method, we tested a range of assortative mating
508 strengths (0 vs values inferred using ANCESTOR) and migration rates (0 vs 1% each
509 generation). The migration levels are used to demonstrate the trends of timing estimates using
510 constant migration and assortative mating, rather than to obtain a specific estimate of the
511 migration parameters. Under random mating, haplotypes are distributed randomly in mating
512 pairs in a population, allowing recombination to shuffle haplotypes over generations. In

contrast, assortative mating can keep ancestral haplotypes more intact, leading to underestimation of the number of generations of admixture when estimating admixture timing. Our evidence of 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[7,13–15,70], but remains an under-recognized force in human population genetics and demographic inference.

Notably, assortative mating alone was not enough to obtain time of admixture estimates concordant with historical records. Admixture timing estimates were most concordant with the historical records when allowing both assortative mating and constant admixture. An admixture scenario that allows multi-wave or constant admixture model is compatible with historical work suggesting constant migration from both Europe and West African throughout the settlement of the islands. Importantly, the authors of ALDER and MultiWaver point out that methods for inferring admixture timing have varying sensitivities to different admixture scenarios[4,5]. MultiWaver may not be able to accurately infer demographic histories that deviate from its pre-defined admixture models. ALDER notes that multi-wave or continuous admixture can lead to more recent estimates of admixture timing, as the method inherently assumes single-point admixture. Despite this caveat, ALDER is frequently applied in cases where admixture is not strictly instantaneous.

534 All of the timing methods we used placed admixture timing for the different islands closer
535 together than historical dates of settlement, consistent with historical expectations that the
536 initial admixture in the southern islands was significant, and that many individuals that
537 occupied the northern and eastern islands during the second and third settlement stages of
538 Cabo Verde were already admixed. This observation, together with IBD and kinship patterns,
539 support the serial founding of the groups of islands as the main model of settlement of Cabo
540 Verde, as opposed to their independent settling as some historical data suggest[38]. This type of
541 serial founding scenario is common throughout recent human migration, underscoring that
542 settlement patterns, in addition to settlement timing, are critical components of accurately
543 inferring human demographic history.
544
545 We found that Cabo Verde's island-specific demographic history and admixture dynamics have
546 important genomic consequences, as observed with ROH. Despite the relatively recent
547 colonization of the islands, some individuals presented even lower overall levels of ROH than
548 African reference populations. We found that low overall levels of ROH in Cabo Verde are
549 driven by shorter length ROH. This observation is consistent with the idea that shorter ROH
550 can be attributed to older shared ancestors from the source populations, and these tracts can be
551 interrupted with admixed ancestry upon admixture. However, many admixed individuals still
552 present excess long ROH, likely reflecting post-admixture processes such as serial founding and
553 assortative mating. The observed island-specific distributions of ROH are consistent with the
554 colonization history of the islands. For example, the cultural dynamics and higher rates of first-
555 cousin marriages in Fogo[46], 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.

Supplemental Information Description

This article has 9 associated supplementary figures and 1 supplementary table.

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. Finally, we would like to thank the Cabo Verdean participants for their invaluable contributions, and the University of Cabo Verde administration for their support.

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, S.R., and Browning, B.L. (2015). Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. Am. J. Hum. Genet. *97*, 404–418.

2. Browning, S.R., Browning, B.L., Daviglus, M.L., Durazo-Arvizu, R.A., Schneiderman, N., Kaplan, R.C., and Laurie, C.C. (2018). Ancestry-specific recent effective population size in the Americas. PLOS Genet. *14*, e1007385.

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

4. Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. Genetics *193*, 1233–54.

5. Ni, X., Yuan, K., Liu, C., Feng, Q., Tian, L., Ma, Z., and Xu, S. (2019). MultiWaver 2.0 : modeling discrete and continuous gene flow to reconstruct complex population admixtures. Eur. J. Hum. Genet. *27*, 133–139.

6. Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. Nat. Genet. *46*, 919–925.

7. Zaitlen, N., Huntsman, S., Hu, D., Spear, M., Eng, C., Oh, S.S., White, M.J., Mak, A., Davis, A., Meade, K., et al. (2017). The Effects of Migration and Assortative Mating on Admixture Linkage Disequilibrium. Genetics *205*, 375–383.

8. Busby, G.B.J., Hellenthal, G., Montinaro, F., Tofanelli, S., Bulayeva, K., Rudan, I., Zemunik, T., Hayward, C., Toncheva, D., Karachanak-Yankova, S., et al. (2015). The Role of Recent Admixture in Forming the Contemporary West Eurasian Genomic Landscape. Curr. Biol. *25*, 2518–2526.

9. Fernandes, V., Brucato, N., Ferreira, J.C., Pedro, N., Cavadas, B., Ricaut, F.-X., Alshamali, F., and Pereira, L. (2019). Genome-Wide Characterization of Arabian Peninsula Populations: Shedding Light on the History of a Fundamental Bridge between Continents. Mol. Biol. Evol. *36*, 575–586.

10. Font-Porterias, N., Arauna, L.R., Poveda, A., Bianco, E., Rebato, E., Prata, M.J., Calafell, F., and Comas, D. (2019). European Roma groups show complex West Eurasian admixture footprints and a common South Asian genetic origin. PLOS Genet. *15*, e1008417.

624 11. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H.,
625 Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three
626 ancestral populations for present-day Europeans. Nature *513*, 409–413.

627 12. Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M.T.P.,
628 Götherström, A., and Jakobsson, M. (2012). Origins and genetic legacy of Neolithic farmers and
629 hunter-gatherers in Europe. Science *336*, 466–469.

630 13. Risch, N., Choudhry, S., Via, M., Basu, A., Sebro, R., Eng, C., Beckman, K., Thyne, S.,
631 Chapela, R., Rodriguez-Santana, J.R., et al. (2009). Ancestry-related assortative mating in Latino
632 populations. Genome Biol. *10*, R132.

633 14. Sebro, R., Peloso, G.M., Dupuis, J., and Risch, N.J. (2017). Structured mating: Patterns and
634 implications. PLOS Genet. *13*, e1006655.

635 15. Zou, J.Y., Park, D.S., Burchard, E.G., Torgerson, D.G., Pino-Yanes, M., Song, Y.S.,
636 Sankararaman, S., Halperin, E., and Zaitlen, N. (2015). Genetic and socioeconomic study of mate
637 choice in Latinos reveals novel assortment patterns. Proc. Natl. Acad. Sci. U. S. A. *112*, 13621–6.

638 16. Arbiza, L., Gottipati, S., Siepel, A., and Keinan, A. (2014). Contrasting X-Linked and
639 Autosomal Diversity across 14 Human Populations. Am. J. Hum. Genet. *94*, 827–844.

640 17. Bustamante, C.D., and Ramachandran, S. (2009). Evaluating signatures of sex-specific
641 processes in the human genome. Nat. Genet. *41*, 8–10.

642 18. Keinan, A., Mullikin, J.C., Patterson, N., and Reich, D. (2009). Accelerated genetic drift on
643 chromosome X during the human dispersal out of Africa. Nat. Genet. *41*, 66–70.

644 19. Ramachandran, S., Rosenberg, N.A., Feldman, M.W., and Wakeley, J. (2008). Population
645 differentiation and migration: Coalescence times in a two-sex island model for autosomal and
646 X-linked loci. Theor. Popul. Biol. *74*, 291–301.

647 20. Wilkins, J.F., and Marlowe, F.W. (2006). Sex-biased migration in humans: what should we
648 expect from genetic data? BioEssays *28*, 290–300.

649 21. Busby, G.B., Band, G., Si Le, Q., Jallow, M., Bougama, E., Mangano, V.D., Amenga-Etego,
650 L.N., Enimil, A., Apinjoh, T., Ndila, C.M., et al. (2016). Admixture into and within sub-Saharan
651 Africa. ELife *5*, e15266.

652 22. Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., and Myers, S.
653 (2014). A Genetic Atlas of Human Admixture History. Science *343*, 747–751.

654 23. Laso-Jadart, R., Harmant, C., Quach, H., Zidane, N., Tyler-Smith, C., Mehdi, Q., Ayub, Q.,
655 Quintana-Murci, L., and Patin, E. (2017). The Genetic Legacy of the Indian Ocean Slave Trade:

656 Recent Admixture and Post-admixture Selection in the Makranis of Pakistan. Am. J. Hum.
657 Genet. *101*, 977–984.

658 24. Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias,
659 W., Fuentes, M., Pizarro, M., Everardo, P., Avila, F. de, et al. (2014). Admixture in Latin
660 America: Geographic Structure, Phenotypic Diversity and Self-Perception of Ancestry Based on
661 7,342 Individuals. PLOS Genet. *10*, e1004572.

662 25. Triska, P., Soares, P., Patin, E., Fernandes, V., Cerny, V., and Pereira, L. (2015). Extensive
663 Admixture and Selective Pressure Across the Sahel Belt. Genome Biol. Evol. *7*, 3484–3495.

664 26. Blant, A., Kwong, M., Szpiech, Z.A., and Pemberton, T.J. (2017). Weighted likelihood
665 inference of genomic autozygosity patterns in dense genotype data. BMC Genomics *18*, 928.

666 27. Kirin, M., McQuillan, R., Franklin, C.S., Campbell, H., McKeigue, P.M., and Wilson, J.F.
667 (2010). Genomic Runs of Homozygosity Record Population History and Consanguinity. PLOS
668 ONE *5*, e13996.

669 28. McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc,
670 L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., et al. (2008). Runs of
671 Homozygosity in European Populations. Am. J. Hum. Genet. *83*, 359–372.

672 29. Pemberton, T.J., Absher, D., Feldman, M.W., Myers, R.M., Rosenberg, N.A., and Li, J.Z.
673 (2012). Genomic Patterns of Homozygosity in Worldwide Human Populations. Am. J. Hum.
674 Genet. *91*, 275–292.

675 30. Mooney, J.A., Huber, C.D., Service, S., Sul, J.H., Marsden, C.D., Zhang, Z., Sabatti, C., Ruiz-
676 Linares, A., Bedoya, G., Fears, S.C., et al. (2018). Understanding the Hidden Complexity of Latin
677 American Population Isolates. Am. J. Hum. Genet. *103*, 707–726.

678 31. Popejoy, A.B., and Fullerton, S.M. (2016). Genomics is failing on diversity. Nature *538*, 161–
679 164.

680 32. Popejoy, A.B., Ritter, D.I., Crooks, K., Currey, E., Fullerton, S.M., Hindorff, L.A., Koenig, B.,
681 Ramos, E.M., Sorokin, E.P., Wand, H., et al. (2018). The clinical imperative for inclusivity: Race,
682 ethnicity, and ancestry (REA) in genomics. Hum. Mutat. *39*, 1713–1720.

683 33. Landry, L.G., Ali, N., Williams, D.R., Rehm, H.L., and Bonham, V.L. (2018). Lack Of
684 Diversity In Genomic Databases Is A Barrier To Translating Precision Medicine Research Into
685 Practice. Health Aff. Proj. Hope *37*, 780–785.

686 34. Beleza, S., Campos, J., Lopes, J., Araújo, I.I., Hoppfer Almada, A., e Silva, A.C., Parra, E.J.,
687 and Rocha, J. (2012). The Admixture Structure and Genetic Variation of the Archipelago of Cape
688 Verde and Its Implications for Admixture Mapping Studies. PLoS ONE *7*, e51103.

35. Beleza, S., Johnson, N.A., Candille, S.I., Absher, D.M., Coram, M.A., Lopes, J., Campos, J., Araújo, I.I., Anderson, T.M., Vilhjálmsson, B.J., et al. (2013). Genetic Architecture of Skin and Eye Color in an African-European Admixed Population. PLoS Genet. *9*, e1003372.

36. Fernandes, A.T., Velosa, R., Jesus, J., Carracedo, A., and Brehm, A. (2003). Genetic differentiation of the Cabo Verde archipelago population analysed by STR polymorphisms. Ann. Hum. Genet. *67*, 340–347.

37. Verdu, P., Jewett, E.M., Pemberton, T.J., Rosenberg, N.A., and Baptista, M. (2017). Parallel Trajectories of Genetic and Linguistic Admixture in a Genetically Admixed Creole Population. Curr. Biol. *27*, 2529–2535.e3.

38. Correia e Silva, A.L. (2002). Dinâmicas de decomposição e recomposição de espaços e sociedades. In História Geral de Cabo Verde., Santos, MEM, ed. (Lisbon, Portugal; Praia, Cabo Verde: Instituto de Investigação Científica Tropical, Instituto Nacional de Investigação, Promoção e Património Culturais de Cabo Verde), pp. 1–66.

39. Baleno, I.C. (2001). Povoamento e Formação da Sociedade. In História Geral de Cabo Verde., Albuquerque, L and Santos, MEM, ed. (Lisbon, Portugal; Praia, Cabo Verde: Instituto de Investigação Científica Tropical, Instituto Nacional de Investigação, Promoção e Património Culturais de Cabo Verde), pp. 125–177.

40. Russell-Wood, A. (1998). The Portuguese Empire, 1415-1808: A World on the Move. (Baltimore, Maryland: Johns Hopkins University Press).

41. Clarke, L., Fairley, S., Zheng-Bradley, X., Streeter, I., Perry, E., Lowy, E., Tassé, A.-M., and Flicek, P. (2017). The international Genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data. Nucleic Acids Res. *45*, D854–D859.

42. Fairley, S., Lowy-Gallego, E., Perry, E., and Flicek, P. (2020). The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. Nucleic Acids Res. *48*, D941–D947.

43. Correia e Silva, A.L. (2001). Espaço, ecologia e economia interna. In História Geral de Cabo Verde., de Albuquerque L, Santos, MEM, ed. (Lisbon, Portugal; Praia, Cabo Verde: Instituto de Investigação Científica Tropical, Instituto Nacional de Investigação, Promoção e Património Culturais de Cabo Verde), p.

44. Cabral, I. (2001). Ribeira Grande: vida urbana, gente, mercancia, estagnação. In História Geral de Cabo Verde., Santos, MEM, ed. (Lisbon, Portugal; Praia, Cabo Verde: Instituto de Investigação Científica Tropical, Instituto Nacional de Investigação, Promoção e Património Culturais de Cabo Verde), pp. 225–273.

723 45. Cabral, I. (2012). A evolução da sociedade foguense: Através de um estudo prosopográfico
724 da elite da ilha (sec. XVI-XVIII).

725 46. Barbosa, Gilda (1997). Casamento entre primos. J. Terra Nova No 246.

726 47. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of
727 ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.

728 48. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J.,
729 Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome
730 association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–75.

731 49. Delaneau, O., Zagury, J.-F., and Marchini, J. (2013). Improved whole-chromosome phasing
732 for disease and population genetic studies. Nat. Methods *10*, 5–6.

733 50. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative
734 modeling approach for rapid and robust local-ancestry inference. Am. J. Hum. Genet. *93*, 278–
735 88.

736 51. Hamid, I., Korunes, K., Beleza, S., and Goldberg, A. (2020). Rapid adaptation to malaria
737 facilitated by admixture in the human population of Cabo Verde (Zenodo).

738 52. Tang, H., Peng, J., Wang, P., and Risch, N.J. (2005). Estimation of individual admixture:
739 analytical and study design considerations. Genet. Epidemiol. *28*, 289–301.

740 53. Tang, H., Coram, M., Wang, P., Zhu, X., and Risch, N. (2006). Reconstructing Genetic
741 Ancestry Blocks in Admixed Individuals. Am. J. Hum. Genet. *79*, 1–12.

742 54. Patterson, N., Price, A.L., and Reich, D. (2006). Population Structure and Eigenanalysis.
743 PLOS Genet. *2*, e190.

744 55. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D.
745 (2006). Principal components analysis corrects for stratification in genome-wide association
746 studies. Nat. Genet. *38*, 904–909.

747 56. Zou, J.Y., Halperin, E., Burchard, E., and Sankararaman, S. (2015). Inferring parental
748 genomic ancestries using pooled semi-Markov processes. Bioinformatics *31*, i190–i196.

749 57. Ochoa, A., and Storey, J.D. (2019). New kinship and FST estimates reveal higher levels of
750 differentiation in the global human population. BioRxiv 653279.

751 58. Browning, B.L., and Browning, S.R. (2013). Improving the accuracy and efficiency of
752 identity-by-descent detection in population data. Genetics *194*, 459–471.

59. Nelson, D., Kelleher, J., Ragsdale, A.P., Moreau, C., McVean, G., and Gravel, S. (2020). Accounting for long-range correlations in genome-wide simulations of large cohorts. PLOS Genet. *16*, e1008619.

60. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. *13*, 2498–2504.

61. Szpiech, Z.A., Blant, A., and Pemberton, T.J. (2017). GARLIC: Genomic Autozygosity Regions Likelihood-based Inference and Classification. Bioinformatics *33*, 2059–2062.

62. Reed, F.A., and Tishkoff, S.A. (2006). African human diversity, origins and migrations. Curr. Opin. Genet. Dev. *16*, 597–605.

63. Tishkoff, S.A., and Williams, S.M. (2002). Genetic analysis of African populations: human evolution and complex disease. Nat. Rev. Genet. *3*, 611–621.

64. Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M., and Wilson, J.F. (2018). Runs of homozygosity: windows into population history and trait architecture. Nat. Rev. Genet. *19*, 220–234.

65. Goldberg, A., Günther, T., Rosenberg, N.A., and Jakobsson, M. (2017). Ancient X chromosomes reveal contrasting sex bias in Neolithic and Bronze Age Eurasian migrations. Proc. Natl. Acad. Sci. *114*, 2657–2662.

66. Eltis, D., Richardson, D., and Blight, D.W. (2010). Atlas of the Transatlantic Slave Trade (Yale University Press).

67. Micheletti, S.J., Bryc, K., Esselmann, S.G.A., Freyman, W.A., Moreno, M.E., Poznik, G.D., Shastri, A.J., Agee, M., Aslibekyan, S., Auton, A., et al. (2020). Genetic Consequences of the Transatlantic Slave Trade in the Americas. Am. J. Hum. Genet. *107*, 265–277.

68. Pero de Guimarães (1991). Corregedor de Cabo Verde, carta de 22 de Maio de 1513. In História Geral de Cabo Verde., de Albuquerque L, Santos, MEM, ed. (Lisbon, Portugal; Praia, Cabo Verde: Instituto de Investigação Científica Tropical, Instituto Nacional de Investigação, Promoção e Património Culturais de Cabo Verde), pp. 219–223.

69. Seibert, G. (2014). Crioulização em Cabo Verde e São Tomé e Príncipe: divergências históricas e identitárias. Afro-Ásia 41–70.

70. Robinson, M.R., Kleinman, A., Graff, M., Vinkhuyzen, A.A.E., Couper, D., Miller, M.B., Peyrot, W.J., Abdellaoui, A., Zietsch, B.P., Nolte, I.M., et al. (2017). Genetic evidence of assortative mating in humans. Nat. Hum. Behav. *1*, 1–13.

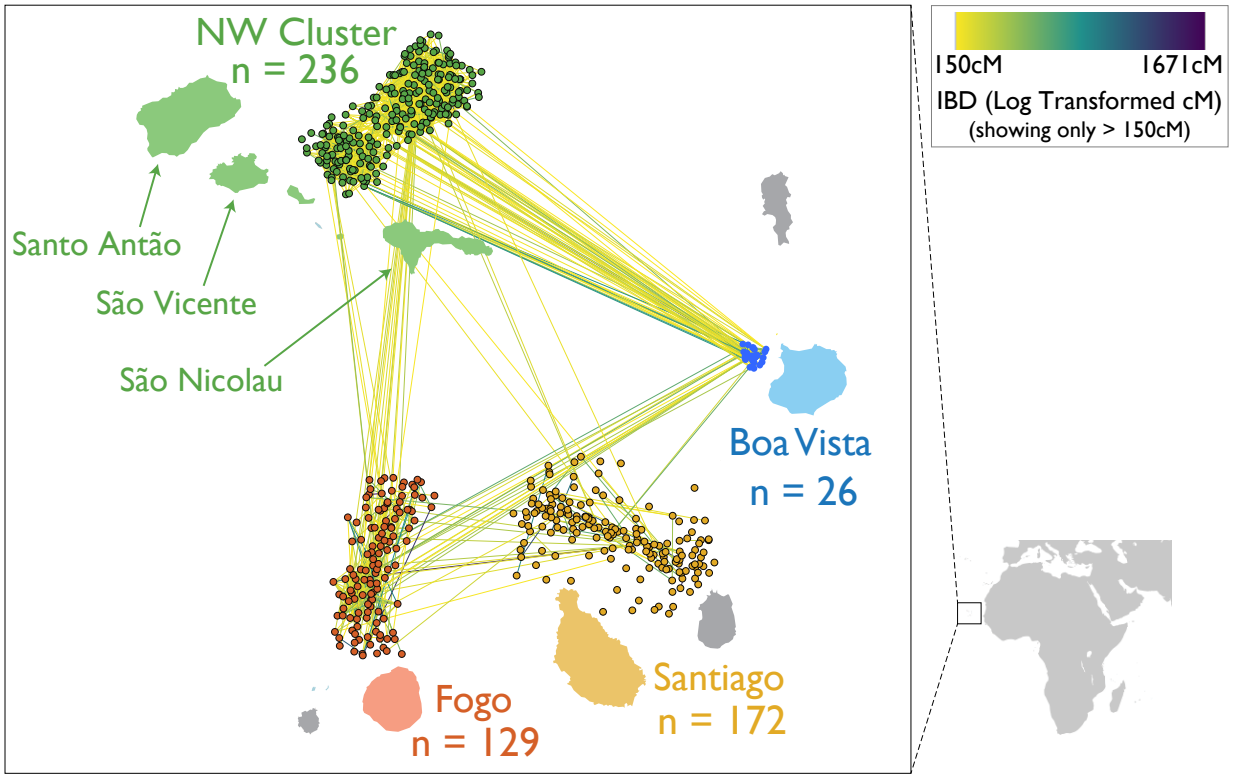

Fig 1 | Shared ancestry between and within the islands in the context of geography. Each island has a corresponding cluster of nodes representing all sampled individuals, with the individuals localized to be adjacent to the island where they were sampled. Node placement within islands is determined by 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.

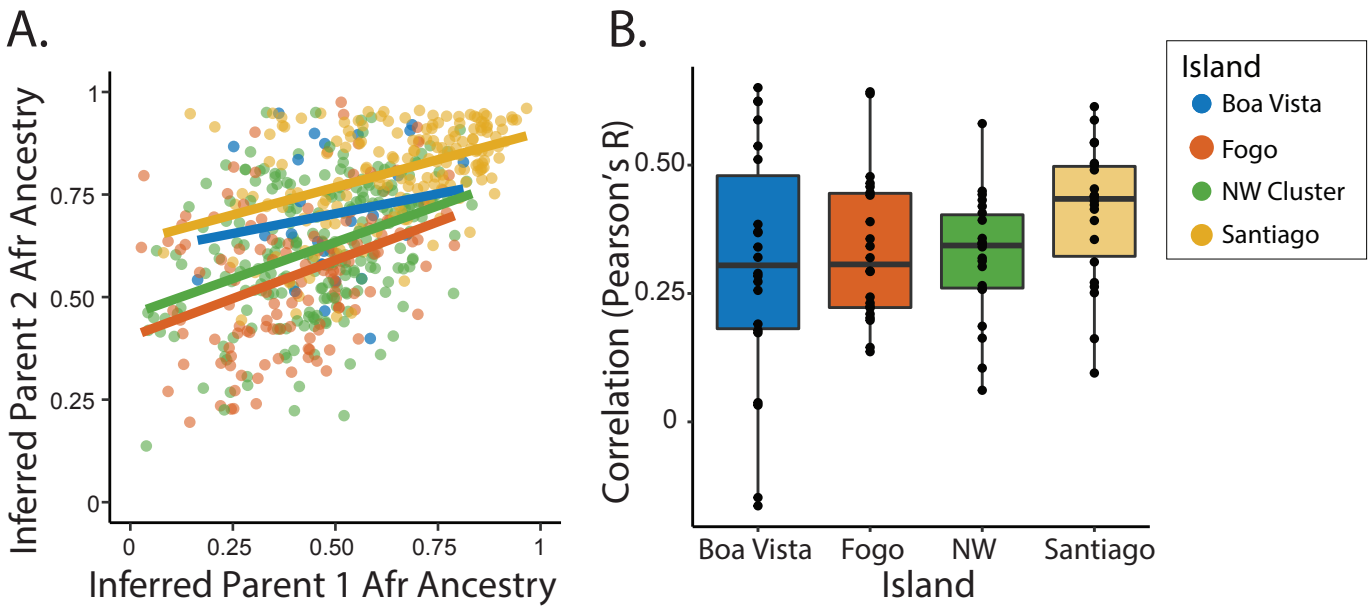

Fig 2 | Correlation between inferred parental ancestries. (A) For each Cabo Verdean individual, the inferred Parent 1 vs Parent 2 ancestry proportions shown for an example chromosome (Chromosome 7), colored by island. (B) shows the set of correlation coefficients between inferred parental ancestries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

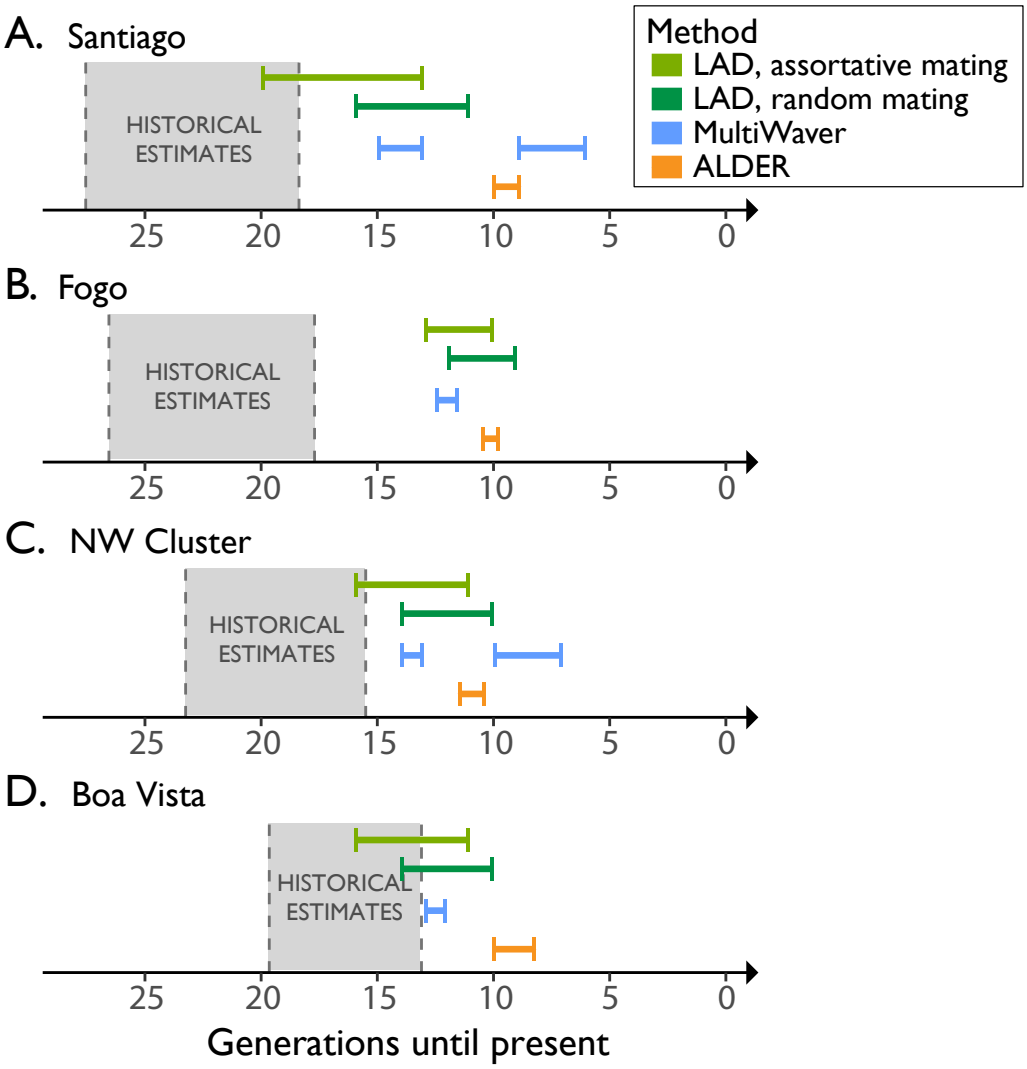

Fig 3 | Timelines of inferred generations of admixture for each island of Cabo Verde. The island regions are listed in order of known settlement timing beginning with Santiago (A), followed by Fogo (B), and then the later migrations to the Northwest Cluster (C) and Boa Vista (D). For each island, admixture 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). LAD-based estimates of admixture timing reflect 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.

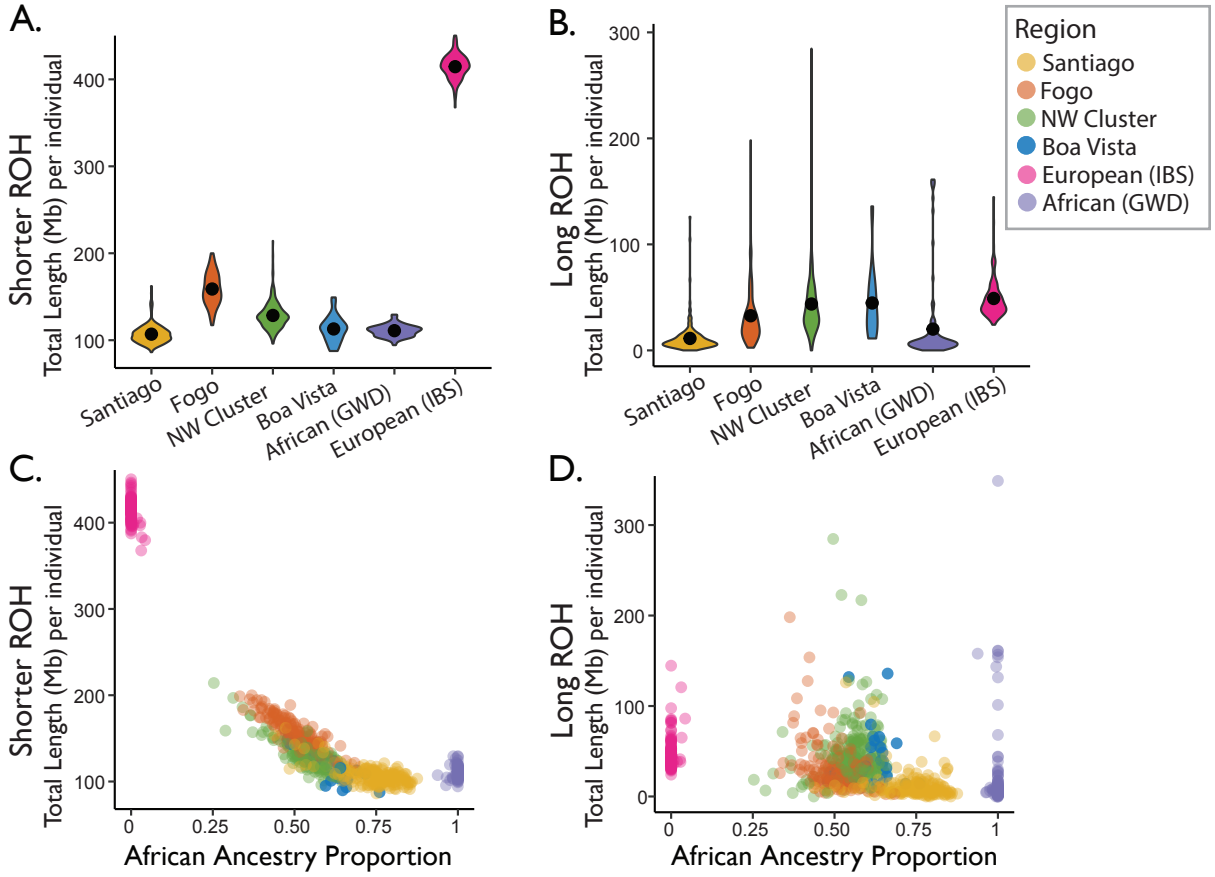

Fig 4 | ROH content by island and in the context of ancestry. (A-B) Violin plots show the population-specific distributions of the total (summed over each genome) length of autosomal ROH per individual. Solid black dots represent the within-population means. (C-D) Total length of autosomal ROH per individual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

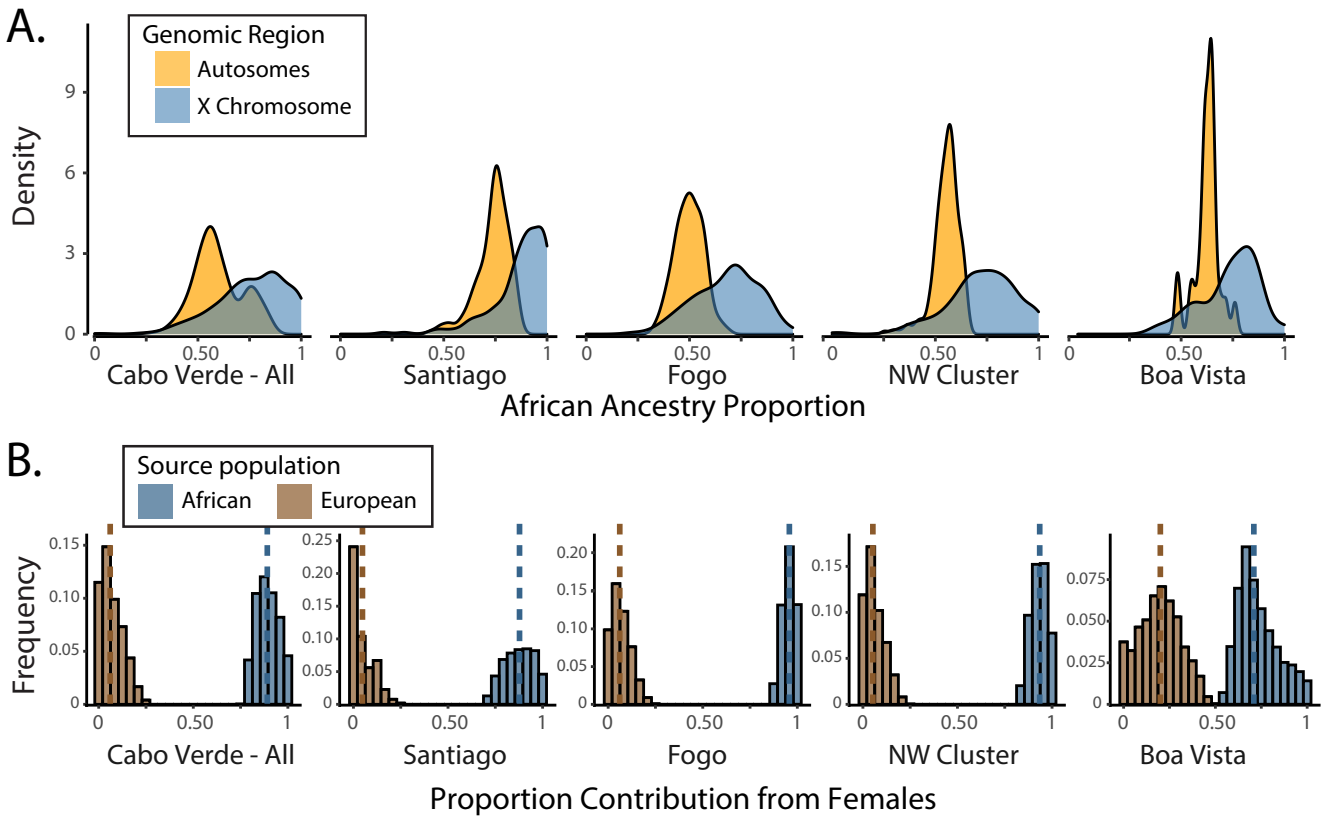

Fig 5 | Sex-biased admixture in Cabo Verde. (A) The distribution of West African ancestry proportion on the autosomes and X chromosome for each of the island regions, estimated with ADMIXTURE. (B) Under a model of constant admixture over time, the fraction of the total contribution of genetic material 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).

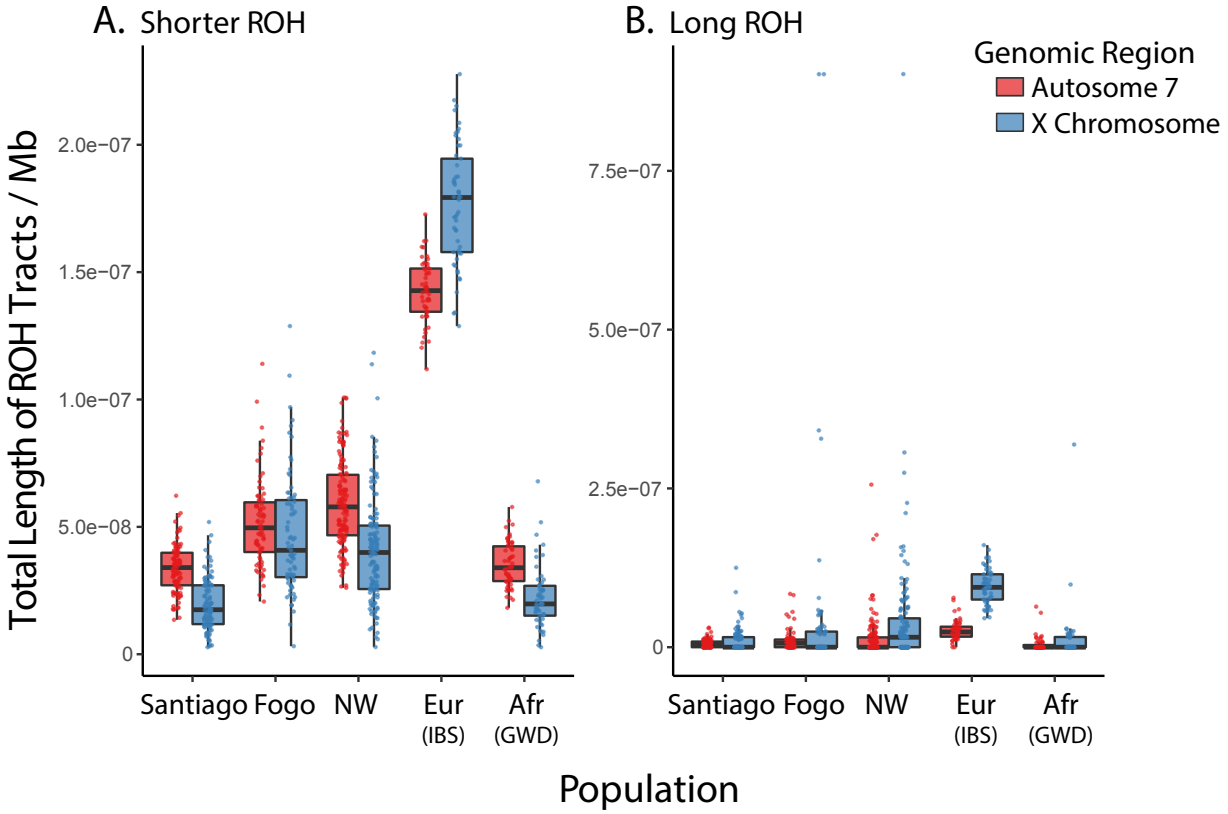

Fig 6 | Autosomal (chromosome 7) vs X chromosome distributions of ROH by class. (A) shows the population-specific distributions of the total length of shorter ROH per individual on the X chromosome compared to an autosome. Chromosome 7 was chosen as the autosomal point of comparison, given that it is 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.